DECIPHERING PROTEIN GLYCOSYLATION THROUGH NOVEL MASS SPECTROMETRY-BASED PROTEOMIC STRATEGIES

A Dissertation
Presented to
The Academic Faculty

by

Haopeng Xiao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry & Biochemistry

Georgia Institute of Technology
May 2018

DECIPHERING PROTEIN GLYCOSYLATION THROUGH NOVEL MASS SPECTROMETRY-BASED PROTEOMIC STRATEGIES

Approved by:

Dr. Ronghu Wu, Advisor
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Bridgette A. Barry
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Julia M. Kubanek
School of Biological Sciences
Georgia Institute of Technology

Dr. Thomas Orlando
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Matthew P. Torres
School of Biological Sciences
Georgia Institute of Technology

Date Approved: March 28, 2018

ACKNOWLEDGEMENTS

Graduate school is a once-in-a-lifetime experience, and the path toward this dissertation has been meandering. Without the guidance, support, and challenges I received along the way, I certainly would have been lost. Foremost, I would like to thank our God for designing such complex yet perfect biological systems and bestowing us the wisdom and courage to reveal the hidden secrets and be amazed at the findings each and every time.

My parents have been extremely supportive throughout all these years. They always loved me unconditionally and provided me all I could ever ask for. I have been away from home for nearly ten years, but their love and support have always accompanied me every step of the way. I am deeply thankful for the wonderful parents I have.

I would like to express my sincere gratitude to my advisor, Dr. Ronghu Wu. I am beyond fortunate to have joined this group and set foot in the field of mass spectrometry-based proteomics. Without your help and guidance throughout every stage of this journey, I would not have been able to accomplish one-tenth of the work in this thesis. Working with you has always been a great pleasure. Your aspirations for research and your work ethics have inspired me over these years in graduate school, and will keep motivating me throughout my entire career. The various life advice you provided have also been most invaluable, helping me make the right choices outside of research.

I also would like to thank other thesis committee members, Dr. Julia M. Kubanek, Dr. Bridgette A. Barry, Dr. Matthew P. Torres, and Dr. Thomas Orlando, for their advisement and assistance in pursuing my Ph.D. and for the countless recommendation letters they have written for me. A special thanks goes to Dr. Mostafa A. El-Sayed and Dr. John F. McDonald for their support on all of my applications. I would also like to thank the chemical education team in freshmen chemistry program, Dr. Carrie G. Shepler, Dr. Michael Evans, Dr.

Kimberly D. Schurmeier, and Ms. Geneva C. Bernoudy, for making teaching general chemistry enjoyable. I also greatly appreciate the National Science Foundation and the National Institute of Health for funding our research over these years.

I am especially grateful for all the support I received from Dr. Weixuan Chen, Dr. Johanna Smeekens, Dr. Jiangnan Zheng, Chendi Jiang, Suttipong (Jay) Suttapitugsakul, Ming Tong, Fangxu Sun, Senhan Xu, Zhenyu Zhou, and all other former and current members of the Wu lab. I am really thankful that I have met you, and our friendship will be a life-long treasure for me. A special thanks goes to George X. Tang, Alexis Pérez, and Alexander A. Choi. Thank you for the proofreading and your support during all the down moments in my past few years. I would also like to thank Dr. Moustafa Ali and Yue Wu for the joy we could share during the productive collaborations. I am extremely thankful for the reunion with Zhiying (Zoey) Hu, my friend who went to the same high school, university, and now graduate school with me. You and your roommate, Liuyi Meng, have brought so much laugher, warmth, and support to my life. I would also like to thank Tianrui Luo, Yalong Cai, Biao Leng, and Xi Zheng, who have been my best friends since high school and are now pursuing dreams across the oceans along with me. Thank you for always being my friends throughout all these years and helping me out whenever I needed. Last but not least, I am deeply grateful for all the mental and spiritual support I received from Ju Eun (Esther) Hwang, her family, and the Atlanta Bethel Church. Thank you for always bringing peace to me at my struggling moments. I truly do not know what I would have done without the help I received.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1 D PEP	One dimensional posterior error probability
ACN	Acetonitrile
AD	Alzheimer's disease
AGC	Automatic gain control
APP	Amyloid precursor protein
Arg^6	$^{13}C_6$ L-arginine
ASCs	Antibody-secreting cells
ATP	Adenosine triphosphate
AuNPs	Gold nanoparticles
AuNRs	Gold nanorods
BA	Boronic acid
BCA	Bicinchoninic acid
BEMAD	Mild beta-elimination followed by Michael addition with dithiothreitol
Boc-Lys(Boc)-OSu	(S)-2,5-dioxopyrrolidin-1-yl 2,6-bis((tert-butoxycarbonyl) amino) hexanoate
CD	Cluster of differentiation
CID	Collision-induced dissociation
CM	Conditioned medium
Con A	Concanavalin A
CuAAC	Copper-catalyzed azide-alkyne [3+2] cycloaddition
CVDs	Heart and cardiovascular diseases
DAVID	The Database for Annotation, Visualization and Integrated Discovery
DBA	Dendrimer-conjugated boronic acid derivative
DBCO	Dibenzocyclooctyne
DCM	Dichloromethane
DDE	Bis-N-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]
diFBS	Dialyzed fetal bovine serum
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulfoxide
Dol-P	Dolichyl phosphate
DTT	Dithiothreitol

ECD	Electron-capture dissociation
ECM	Extracellular matrix
EDC	N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
Endo H	Endoglycosidase H
ER	Endoplasmic reticulum
ERLIC	Electrostatic repulsion hydrophilic interaction chromatography
ETD	Electron-transfer dissociation
Exo-Depl CM	Exosome-Depleted conditioned medium
FA	Formic acid
Farnesyl-PP	Farnesyl-pyrophosphate
FBS	Fetal bovine serum
FDR	False discovery rate
Fmoc-L-Lys(Boc)-OSu	(2,5-dioxopyrrolidin-1-yl) (2S)-2-(9H-fluoren-9-ylmethoxycarbonylamino)-6-[(2-methylpropan-2-yl)oxycarbonylamino] hexanoate
GalNAc	N-acetylgalactosamine
GalNAz	N-azidoacetylgalactosamine
Geranylgeranyl-PP	Geranylgeranyl pyrophosphate
GlcNAc	N-acetylglucosamine
GlcNAz	N-azidoacetylglucosamine
HCD	Higher-energy collisional dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HexNAC	N-acetylhexoseamine
HILIC	Hydrophilic interaction liquid chromatography
HMGCR	Hydroxy-3-methyl-glutaryl-coenzyme A reductase
HOAc	Acetic acid
HOAt	1-hydroxy-7-azabenzotriazole
HPLC	High-performance liquid chromatography
IAA	Iodoacetamide
IsoTaG	Isotope-targeted glycoproteomics

LC	Liquid chromatography
LDA	Linear discriminant analysis
LTQ	Linear ion trap
LWAC	Lectin weak-affinity chromatography
Lys^8	$^{13}C_6$, $^{15}N_2$ L-lysine
MALDI	Matrix-assisted laser desorption/ionization
ManNAc	N-acetylmannosamine
ManNAz	N-azidoacetylmannosamine
Maximum IT	Maximum ion accumulation times
MS	Mass spectrometry
MSCs	Marrow-derived mesenchymal stromal cells
NCE	Normalized collision energy
NETs	Neutrophil extracellular traps
NGAG	Solid phase extraction of N-linked glycans and glycosite-containing peptides
NHS	N-hydroxysuccinimide
NIR	Near-infrared
OC	Ovarian cancer
OD	Optical density
PBS	Phosphate buffered saline
PC	Photocleavable
PKA	Protein kinase A
PKC	Protein kinase C
PNGase F	Peptide-N-glycosidase F
PPTT	Plasmonic photothermal therapy
PTK	Protein tyrosine kinase
PTM	Post translational modification
PUGNAc	O-(2-Acetamido-2-deoxy-D-glucopyranosylidenamino)N-phenylcarbamate
RCA	Ricinus communis agglutinin
RF	Rifampicin
RIPA	Radioimmunoprecipitation assay
S/N	Signal-to-noise ratio

SDC	Sodium deoxycholate
SERS	Surface-enhanced Raman spectroscopy
SILAC	Stable isotope labelling with amino acid in cell culture
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAP	Tandem affinity purification
TEA	Trimethylamine
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
TFMS	Trifluoromethanesulfonic acid
THPTA	Tris(3-hydroxypropyltriazolylmethyl) amine
TM	Tunicamycin
TM	Transmembrane
TMT	Tandem mass tag
TOF	Time-of-flight
VVA	Vicia villosa agglutinin
WGA	Wheat germ agglutinin
XCorr	Cross-correlation
YPD	Yeast extract peptone dextrose
ZFN	Zinc-finger nuclease
ZIC-HILIC	Zwitter-ionic hydrophilic interaction chromatography

SUMMARY

Protein glycosylation is essential for cell survival and proliferation. Comprehensive analysis of protein glycosylation can aid in a better understanding of protein functions, cellular activities, and the molecular mechanisms of diseases. Emerging mass spectrometry (MS)-based proteomics enables comprehensive analysis of protein glycosylation and many other types of modifications. However, due to the heterogeneity of glycans and the low abundance of many glycoproteins in complex biological samples, it is extraordinarily challenging to globally and site-specifically analyze glycoproteins. This thesis focuses on the development of new methods for global analysis of glycoproteins, and the applications of the newly developed methods for biomedical research.

This thesis is constituted of six chapters. Chapter 1 is an overview of MS-based glycoproteomics analysis, with an emphasis on the endeavors in the literature to solve the two major problems for global analysis of glycoproteins mentioned above. This chapter retraces the developments of important chemical and enzymatic methods in this field, and includes the discussion regarding how these methods have enabled qualitative and quantitative analyses of glycoproteins in a variety of biological systems. Chapter 2 focuses on the development of a strategy that utilizes the universal recognition between boronic acid and sugars, in order to enrich glycopeptides for LC-MS/MS analysis. Chapter 3 shows the approach of achieving quantitative analysis of protein glycosylation through the combination of boronic acid enrichment and quantitative proteomics. Chapter 4 describes a strategy for cell-surface N-glycoproteome analysis. Metabolic labeling, click chemistry, and MS-based proteomics were combined to specifically map the glycoproteins located only on cell surface. The labeling efficiencies of different sugar analogs were compared, and this method was combined with either stable isotope labeling in cell culture (SILAC) or tandem mass tag (TMT)-labeling to quantitatively study the surface N-glycoproteins. Chapter 5 explains how protein S-

GlcNAcylation was unexpectedly found in human cells. Starting with an attempt to profile protein O-GlcNAc, hundreds of S-GlcNAcylation sites were surprisingly identified on cysteine residues. This modification was demonstrated not to be caused by chemical reactions with the cleavable linker during sample preparation nor due to false site assignment. Furthermore, protein S-GlcNAcylation events were investigated with different sugar analog labeling in three cell lines. Chapter 6 features an application of MS-based proteomics in biomedical research. In this chapter, the cellular responses and pleiotropic effects in statin-treated cells on the proteome, glycoproteome, and phosphoproteome levels were analyzed.

In addition to the independent projects discussed above, the collaborative projects about that investigation of the cellular mechanisms of gold-nanorod assisted cancer photothermal therapy, and the discordance between mRNA and proteome in ovarian cancer tissues were also conducted. The abstracts of the publications resulted from the collaborations are shown in the appendix.

In conclusion, the work presented in this thesis majorly combines chemical biology and modern MS-based proteomics to study protein modifications, especially glycosylation. This thesis strives to advance the techniques of glycoproteomics and apply the state-of-the-art methods to investigate biological and biomedical problems.

CHAPTER 1. INTRODUCTION TO MASS SPECTROMETRIC ANALYSIS OF GLYCOPROTEINS IN COMPLEX BIOLOGICAL SYSTEMS

1.1 Background Introduction

Mass spectrometry (MS)-based proteomics has become an increasingly powerful tool to study diverse topics in complex biological systems. Protein post-translational modifications (PTMs) are extremely important in biological systems and regulate nearly every cellular activity including gene expression, signal transduction and cellular response to environmental cue. Comprehensive and site-specific analysis of protein modifications is beyond reach of conventional biochemistry methods. Modern MS technology provides a unique opportunity to globally and site-specifically characterize protein PTMs. However, it is extraordinarily challenging because of the low abundance of many modified proteins, sub-stoichiometry of protein modification, and the complexity of biological samples. In addition, the modified groups are different, and therefore no common method can be used for all types of protein modifications. Innovative and effective methods are crucial to achieve the global analysis of protein modification.

Among hundreds of known PTMs, protein glycosylation is one of the most important modifications and is essential for cell survival because it determines protein folding, trafficking, and stability and regulates many cellular activities, especially extracellular activities. Many glycoproteins are of extremely low abundance compared to non-modified proteins; meanwhile, glycosylation is very complex due to the heterogeneous glycan structures and a variety of amino acid residues being modified with glycans. To overcome the challenges brought by the complexity of glycosylation and the low abundance of glycosylated proteins, many elegant

methods were developed to qualitatively and quantitatively study protein glycosylation events in various kinds of samples.

Here some major MS-based methods for global analysis of protein glycosylation are discussed. In the first part, we summarize the chemical and enzymatic methods for MS-based glycoproteomics, including different enrichment methods and the methods to generate a common tag for global analysis of protein N- and O-glycosylation with MS. Second, because reversible protein glycosylation makes glycoproteins highly dynamic, MS-based methods for glycoprotein dynamics study are included. The last part includes selected applications of MS-based glycoproteomics in a variety of biology systems.

1.2 Global Analysis of Glycoproteins

1.2.1 Glycopeptide enrichment

In order to globally analyze protein glycosylation, glycoprotein enrichment is imperative prior to MS analysis due to the low abundance of many glycoproteins, the dynamic nature of protein modifications and the complexity of biological samples [1-4]. Enrichment can allow us to minimize the interference from highly abundant non-glycoproteins on the analysis of protein glycosylation, and to reach low-abundance glycoproteins. In the literature, a variety of enrichment methods were reported, and each method has its own advantages and limitations.

1.2.1.1 Lectin

Commercially available lectins were mostly originated from plants, with some also from bacteria and animal species. Each lectin can bind one or several types of glycans, and therefore lectins were extensively used to perform glycoprotein/peptide enrichment, although to an extent the specific binding also limits lectins from enriching all types of glycans for more comprehensive analysis of glycoproteins in complex biological samples. Several lectins, such

as concanavalin A (Con A, mainly specific for internal and nonreducing terminal α-D-mannosyl and α-D-glucosyl groups), wheat germ agglutinin (WGA, specific for N-acetyl-D-glucosamine and sialic acids), and ricinus communis agglutinin I (RCA I/ RCA 120, binds galactose or N-acetylgalactosamine residues) were widely used for glycoprotein/peptide enrichment [5]. Typically, lectins were immobilized onto solid support and served for solid-phase extraction of glycopeptides. The lectin-functionalized beads can also be packed into separation columns, allowing for enrichment and elution of glycopeptides coupling with liquid chromatography (LC). Zielinska et al. combined lectin enrichment with filter-aided sample preparation (FASP) to map the N-glycosylation sites in four mouse tissues and blood plasma [6]. They identified 6,367 sites on 2,352 proteins from extracellular space, organelle lumens, and other cellular locations. In addition to the widely known NX[S/T] (X stands for any amino acid residue other than proline) motif for protein N-glycosylation, they also found other rare motifs, such as the NXC motif. The same group further employed the FASP method coupled with lectins to profile the N-glycosylation sites across seven evolutionarily distant species, and found the distant species have common characteristics including sequence recognition patterns, structural constraints, and subcellular localization although the N-glycoproteome from those species are largely divergent [7].

Besides protein N-glycosylation, lectin-based enrichment methods were also applied for O-glycosylation analysis although the strategies are not as mature as those for N-glycosylation analysis. For instance, Darula and Medzihradszky used the jacalin lectin (specific for recognizing GalNAcα1-O- extension of the core 1 structure) to enrich O-glycopeptides and identified O-glycosylation sites from bovine serum through mass spectrometric analysis [8]. The same lectin was also employed by Durham and Regnier, and they immobilized jacalin onto silica beads and further packed an LC column for serial lectin affinity chromatography analysis of O-glycopeptides after removing N-glycopeptides with Con A [9]. Steentoft et al. performed

O-glycosylation analysis combining vicia villosa agglutinin (VVA) lectin chromatography with the SimpleCell technology [10], which will be discussed in more details in a section below.

As an important and special type of O-glycosylation, O-GlcNAcylation has been well-studied in the past three decades, and lectin-based methods have also been developed for comprehensive mapping of protein O-GlcNAcylation. In this context, WGA was exploited for the enrichment of O-GlcNAcylated proteins or peptides due to its substrate specificity. For example, Vosseller et al. developed lectin weak-affinity chromatography (LWAC), and by combining it with β-elimination/Michael addition with DTT (BEMAD) and ECD mass spectrometry, they analyzed 145 unique O-GlcNAcylated peptides from a postsynaptic density preparation [11]. Overall, lectin-based strategies have greatly expanded the pool of N- and O-glycosylation sites identified.

1.2.1.2 Hydrazide chemistry

In 2003, Zhang et al. developed an elegant strategy based on hydrazine chemistry for glycopeptide enrichment, followed by protein N-glycosylation analysis with MS [12]. They oxidized the glycans to generate aldehyde groups, and then conjugated the glycoproteins to a solid support using hydrazide chemistry. After on-beads digestion, the non-glycopeptides were removed and the enriched glycopeptides were recovered by using Peptide-N-Glycosidase F (PNGase F) to cleave off the glycans. The peptides were then identified and quantified by LC-MS/MS. This method was further modified by many groups for better performance and has been extensively used to study protein N-glycosylation in a variety of samples and species, from cells to clinical samples [13-15]. During the oxidation, the glycans are damaged, and thus this method cannot be employed to analyze intact glycopeptides with glycan structural information.

Although initially this method was designed for protein N-glycosylation analysis as PNGase F can only recognize the N-glycans, researchers developed variants and used them for O-glycosylation analysis. Nilsson et al. combined hydrazide chemistry with acid cleavage to analyze sialylated glycoproteins, and identified 36 N-linked and 44 O-linked glycosylation sites from human cerebrospinal fluid with site and glycan structural information [16]. Later they applied a similar strategy to study human urinary glycoproteins containing glycan information, and 58 N- and 63 O-linked "intact" glycopeptides were identified from 53 glycoproteins [17]. This method was further improved by pretreating the samples with PNGase F to remove N-glycans, and used CID-MS^2/MS^3 and ECD/ETD to comprehensively analyze the samples [18]. Taga et al. combined hydrazide chemistry with galactose oxidase oxidation and formic acid-induced cleaving of the hydrazone bond to study the O-glycosylations unique to collagen [19]. Klement et al. also applied hydrazide chemistry to O-GlcNAcylation studies where they elevated the concentration of sodium periodate and reaction temperatures, and enriched the modified proteins through hydrazide chemistry [20]. The O-GlcNAcylated peptides were released by hydroxylamine treatment and analyzed by tandem MS. A total of 12 O-GlcNAcylated peptides from 5 proteins were identified in that study.

1.2.1.3 HILIC

Hydrophilic interaction liquid chromatography (HILIC) is a separation technique that has been widely used in glycoproteomics to separate/enrich glycopeptides from non-glycopeptides. Contrary to reverse-phase chromatography, the stationary phase of HILIC is very hydrophilic, allowing binding of hydrophilic analytes. Silica particles, amino or hydroxyl groups, zwitter ions are common materials for the HILIC stationary phase [21]. The gradient of the mobile phase usually starts with high percentage of relatively nonpolar organic solvent,

then the percentage of polar component (usually aqueous solutions) is increased through the gradient timeline, resulting in higher hydrophilicity of the solution and stronger elution power.

Because glycans are extremely hydrophilic molecules, the retention time of glycopeptides are generally longer than non-glycopeptides. In the literature, Hägglund et al. combined zwitter-ionic hydrophilic interaction chromatography (ZIC-HILIC) enrichment and partial deglycosylation to study protein N-glycosylation in human plasma, and identified 62 glycosylation sites from 37 glycoproteins [22]. The same group later employed a similar strategy to establish an enzymatic deglycosylation scheme to study core fucosylated N- and O-glycosylation among human plasma proteins [23]. HILIC enrichment was extensively applied for glycosylation analysis, including site mapping, intact glycopeptide analysis, and also contributed in O-GlcNAcylation studies [24]. However, if a non-glycopeptide contains multiple hydrophilic amino acid residues, its retention time may be comparable to glycopeptides, rendering the specificity of HILIC lower than many other enrichment methods. Nevertheless, HILIC can be coupled with other enrichment methods to perform two-dimensional separation and fractionation. There are some reports in the literature to improve the performance of HILIC, such as introducing ion-pairing reagents in the mobile phase and further functionalizing the stationary phase [25, 26], and have considerably advanced the use of HILIC in glycoproteomics.

1.2.1.4 IsoTag

Recently Woo et al. developed a highly innovative isotope-targeted glycoproteomics (IsoTaG) method [27], which combined metabolic labeling, isotopic recording, and MS-based proteomics to analyze intact N- and O-glycopeptides. They synthesized an isotopic affinity probe, which has four critical parts: the aizde for tagging the glycans through copper-catalyzed azide-alkyne [3+2] cycloaddition (CuAAC), the biotin group for enriching the tagged glycopeptides through strong biotin-avidine interactions, the silane scaffold being readily

cleaved to release the glycopeptides after enrichment, the dibromide motif for MS detection of glycopeptides. Due to the natural abundance of the stable isotopes of Br (Br^{79}:Br^{80}=1:1), glycopeptides tagged with the probe display a special pattern in MS analysis (termed IsoStamp) [28].

Four major steps are involved in the IsoTaG strategy: (1) metabolically label the glycans with a functional sugar analog (i.e. N-azidoacetylgalactosamine (Ac4GalNAz)); (2) tag the labeled glycoproteins with the probe, and then capture the glycoproteins on a solid phase through biotin-avidin interactions. The enriched glycoproteins were digested using trypsin and the glycopeptides were released by cleaving the silane scaffold; (3) analyze the glycopeptides by tandem MS. Here they used a pattern-searching algorithm, which recognizes the 2:5:1 distribution of the dibromide motif to selectively sequence the glycopeptides; (4) assign glycosylation sites and glycan structures using Byonic software.

IsoTaG increases the selection and detection speed of glycopeptides owning to the pattern-searching algorithm, and greatly aids in the analysis of intact glycopeptides. This strategy can be widely used for not only glycoproteomics, but also a variety of PTMs and targeted protein analyses. The metabolic labelling may limit its application for glycoprotein analysis of clinical samples. In addition, it requires many steps to perform the analysis (from probe synthesis to data-independent MS analysis), which may hinder its wide usage.

1.2.1.5 NGAG

Zhang and co-workers recently reported an innovative chemoenzymatic method, named solid phase extraction of N-linked glycans and glycosite-containing peptides (NGAG), to comprehensively analyze N-glycoproteins and glycans in complex samples [29]. This method utilizes enzymatic and chemical reactions to analyze the N-glycans and their parent deglycosylated peptides. Briefly, proteins were firstly digested, and the resulting peptides were

guanidinized to block the ε-amino groups on the lysine side chain. Peptides were then conjugated to the solid phase via their N-termini, and the carboxyl groups of aspartic acid (D), glutamic acid (E), peptide C termini, and sialic acids were reacted with aniline to facilitate the mass spectrometric detections. N-glycans were released from the solid support by PNGase F treatment and subjected to MS analysis while their corresponding asparagine residues were converted to aspartic acids. This provides the opportunity for Asp-N-induced cleavage of the peptides, and thus releasing them from the solid phase while the aspartic acids that are not generated from PNGase F cleavage of N-glycans will not be affected because they were modified by aniline in the prior step. The released deglycosylated peptides were also identified by MS.

This strategy led to the identification of 2,044 unique N-glycopeptides, and in an experiment comparing NGAG and hydrazide chemistry methods, they analyzed a total of 3,083 unique N-glycosite-containing peptides from 1,473 glycoproteins in OVCAR-3 cells. Quantitative analysis of glycopeptides based on NGAG was also performed, and proved treating cells with tunicamycin mainly caused glycan occupancy reduction on the glycosylation sites. Further glycan dynamic experiments also showed differential alteration of glycans by the tunicamycin treatment. Overall, NGAG is an excellent example of rationale usage of chemical and enzymatic reactions to advance MS-based glycoproteomics, although the complex steps could potentially put limitations to its applications.

1.2.1.6 Boronic acid-based enrichment methods

Boronic acids have great potential for enriching glycopeptides/glycoproteins because one common feature of all glycans is that they all contain multiple hydroxyl groups. The covalent interactions between boronic acids and cis-diols on glycans have been extensively studied in the literature, and has been applied for glycoprotein analysis. Yang and co-workers

designed a boronic acid-functionalized core-satellite-structured composite material to capture glycopeptides/proteins, and analyzed 194 unique glycosylation sites from 155 different glycoproteins [30]. Later the same group synthesized a boronic acid-functionalized detonation nanodiamond to enrich glycopeptides and analyze glycoproteins, which has led to the identification of 40 unique N-glycospeptides from 34 unique glycoproteins in mouse liver [31]. Zeng et al. designed a surface patterned sample support with a hydrophobic outer layer and an internal boronic acid-modified gold microspot, to selectively enrich glycopeptides. The enriched glycopeptides were then directly subjected to MALDI MS analysis [32]. Metz and co-workers combined boronate affinity chromatography with ETD MS to analyze non-enzymatically glycated peptides [33].

Wu and co-workers have conjugated boronic acids onto magnetic beads, and then used the functionalized beads to enrich glycopeptides from protein digestions for MS-based proteomic analysis [34]. Due to the nature of the pH-dependent reversible interactions between boronic acids and hydroxyl groups, the enrichment was performed under basic condition (pH=10) to capture glycopeptides on the beads. After several washes to remove non-glycopeptides, the elution was performed under acidic condition to release glycopeptides. The enriched glycopeptides were then treated with PNGase F in heavy oxygen water ($H_2{}^{18}O$) for only three hours to generate a common tag for N-glycosylation site identifications, which distinguishes the bona fide glycosylation sites from spontaneous asparagine deamidation sites. They applied this strategy to study the yeast glycoproteome, and identified 816 N-glycosylation sites on 332 proteins.

The reactions between boronic acid and hydroxyl groups make boronic acid-based chemical enrichment universal for nearly all types of glycopeptides/glycoproteins. The reversible nature of this interaction and the mild reaction conditions ensures the glycans undamaged after enrichment, therefore this method is able to be used for intact glycopeptide

analysis with glycan structural information. This method can be further improved by strengthening the binding between boronic acids and glycans, allowing the capturing of more glycopeptides and minimizing glycopeptide loss. Overall, the combination of boronic acid-based enrichment methods and MS-based proteomics has the potential to universally analyze protein glycosylation with structural information on a large scale.

1.2.1.7 Click chemistry-based methods

Metabolic labeling of glycans with unnatural sugar analogs has been proven to be powerful to study glycoproteins[35]. In the recent two decades, the Bertozzi group has been a pioneer in using unnatural sugar analogs to label glycoproteins. They firstly developed a cell surface engineering strategy by combining metabolic labeling with a modified Staudinger ligation reaction [36]. They found that acetylated azide-containing sugar analogs have much greater labeling efficiency than the non-acetylated versions. Azide-containing sugar analogs are commonly used in these methods because the azide group is small, which does not create an unacceptable steric hindrance that renders the analog unable to be recognized by the enzymes, and it is relatively stable and biologically inert.

To date, a variety of sugar analogs have been used for metabolic labeling, such as GalNAc, GlcNAc, fucose, and ManNAc analogs [37]. After incorporation of the azido sugar analogs, a click reaction is performed to introduce another chemical handle for affinity enrichment. The glycoproteins/ glycopeptides can be analyzed by a variety of methods including proteomics techniques.

1.2.1.8 Other chemical and enzymatic enrichment methods

Chemical and/or enzymatic methods facilitated the identification and quantification of not only protein N-glycosylation, but also O-glycosylation, especially O-GlcNAcylation. For

instance, Khidekel et al. established a method, termed quantitative isotopic and chemoenzymatic tagging (QUIC-Tag), which enzymatically incorporates the ketone group onto O-GlcNAcylated proteins, and then tags it with biotin for affinity enrichment of O-GlcNAcylated peptides/proteins for rapid and sensitive identification and quantification [38]. Through combining this strategy with quantitative isotopic dimethyl labeling, they studied O-GlcNAcylation dynamics in cultured neurons and rat brain samples.

Wang et al. developed a click chemistry-based strategy by combining chemical/enzymatic tagging, photochemical cleavage, and electron-transfer dissociation (ETD) mass spectrometry to enrich O-GlcNAc modified peptides and map O-GlcNAcylation sites [39]. They used enzyme GalT1 to transfer an azide-containing sugar analog (UDP-GalNAz) onto the O-GlcNAc moieties on the modified peptides, and then incorporated a biotin group on it through CuAAC for the enrichment. The enriched O-GlcNAcylated peptides were then released by photochemical cleavage, followed by LC-MS analysis. The same group has also made significant contributions to study the biological importance of protein O-GlcNAcylation and investigate the cross-talk between O-GlcNAcylation and phosphorylation through proteomics-based strategies [40-45].

Chemical and enzymatic strategies significantly broaden the toolbox of MS-based proteomics, and have facilitated various qualitative and quantitative studies of protein PTMs. We envision that further development of chemoenzymatic methods will tremendously advance our understanding of protein glycosylation.

1.2.2 Generating a common tag for MS analysis

Glycans are highly heterogeneous, and the diverse structures contain a wealtj of biological information. At the same time, diversity of glycans results in difficulty for global analysis of protein glycosylation with MS. Unlike other modified groups with a fixed structure,

such as protein phosphorylation or methylation with a universal mass shift for all modified peptides and proteins [46], protein glycosylation does not have a common mass tag for MS analysis. In order to globally identify glycosylation sites, methods that generate a tag for glycosylation sites will provide convenience for spectra matching. This section covers the methods generating a common mass tag for MS-based glycoproteomics analysis.

1.2.2.1 A common tag for protein N-glycosylation

The simplest strategy is to remove the glycans while the residual mass can serve as a tag to localize the glycosylation site. To achieve this goal, enzymatic methods have been widely used to deglycosylate N-glycans. The enzyme peptide-N4-(N-acetyl-β-glucosaminyl)asparagine amidase F (PNGase F) is the most commonly used to remove protein N-glycans. PNGase F was initially isolated from *Flavobacterium meningoseptic* in 1984 [47], and has been used for N-glycan removal since then. It was reported to recognize a broad spectrum of substrates and can hydrolyze the glycosylamine linkage between the glycan and the amino acid, which generates a deglycosylated peptide and an intact oligosaccharide with the di-N-acetylchitobiose unit at the reducing end [48]. In the process, the asparagine residue will be converted to aspartic acid, introducing a mass shift which can serve as a universal mass tag for glycosylation site localization. The reaction is shown in Figure 1.1.

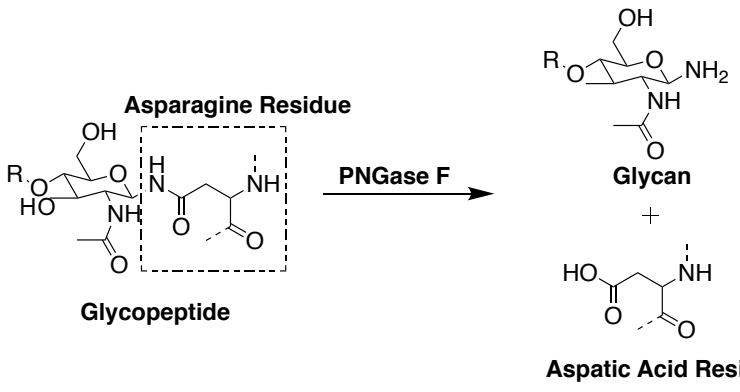


Figure 1.1 Deglycosylation using PNGase F. After glycoprotein/glycopeptide enrichment, the glycan can be removed by PNGase F, converting Asn to Asp at the same time.

Although PNGase F has been widely used, there are still a few drawbacks of using this method. First, spontaneous deamidation of asparagine residues occurs at all time *in vitro* and *in vivo* which generates the same mass shift as PNGase F cleaving reaction, rendering it hard to control the false positive glycosylation site identification. To solve this problem, heavy oxygen water ($H_2{}^{18}O$) has been employed as the solvent to carry out the enzymatic removal of N-glycans by PNGase F [49, 50]. In this case, the mass tag of accumulated spontaneous deamidation is different from the one from authentic deglycosylation with PNGase F. The reaction can be performed for a shorter period of time to limit the impact of deamidation during PNGase F treatment. In our previous experiment, under the PNGase F treatment at neutral conditions, the spontaneous deamination effect can be negligible. The other limit of this method is that PNGase F is not able to remove glycans with fucose attached α1 → 3 to the asparagine-linked N-acetylglucosamine residue [51]. This problem has not been solved yet, although combining several deglycosylation enzymes may help release the glycans more efficiently.

Aside from PNGase F, the endoglycosidases are a family of important deglycosylation enzymes that have also been reported in the literature. Endoglycosidase H (Endo H) is the most widely used one among the endoglycosidases. It cleaves within the chitobiose core of high mannose and some hybrid oligosaccharides from N-linked glycoproteins, although it does not work efficiently against complex glycans. Endo H was firstly isolated from *Streptomyces plicatus* and its structure was described by Robbins et al. in 1984 [52]. The residual mass of Endo H cleavage is different from that of PNGase F removal because the innermost GlcNAc residue remains on the peptide after the glycan cleavage. The cleaving process of Endo H is shown in Figure 1.2.

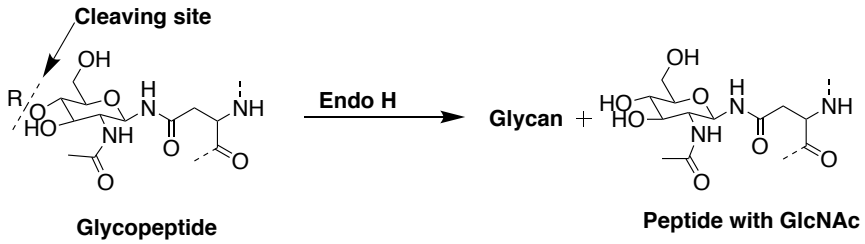


Figure 1.2 Deglycosylation using Endo H. After cleaving the glycan, the peptide is left with one residual GlcNAc.

The remaining GlcNAc can serve as the mass tag for glycosylation site localization. However, this tag is considerably larger than the one from PNGase F treatment, which may have a negative effect on the quality of tandem MS, especially when collision-induced dissociation (CID) is used to fragment glycopeptides. In addition, Endo H has higher substrate specificity comparing to PNGase F, rendering the enzymatic glycan release not as complete. To improve the glycoproteome coverage, researchers have combined several endoglycosidases to remove glycans. For instance, Hägglund et al. combined Endo H and Endo D to deglycosylate glycopeptides enriched by hydrophilic interaction liquid chromatography, which led to the identification of 62 glycosylation sites on 37 glycoproteins from human plasma samples (Hägglund et al., 2004).

Although enzymatic methods have their advantages, each enzyme has its own specificity besides that they are not cost-effective. Therefore, it is very difficult to universally remove all glycans using a single enzyme or a combination of several enzymes. Chemical methods were also developed to remove glycans for glycosylation mapping. More than three decades ago, Edge et al. developed a method using trifluoromethanesulfonic acid (TFMS) to remove glycans from fetuin, and found that this treatment at 0 or 25°C results in rapid cleavage of peripheral sugars, slow loss of serine- and threonine-linked GalNAc, and retention of N-linked GlcNAc [53]. Other methods such as ammonium hydroxide/carbonate-based chemical

deglycosylation were also reported in the literature [54]. Recently, Chen et al. developed a method combining removal of glycans using TFMS and MS-based proteomics to perform large-scale analysis of protein N-glycosylation from complex biological samples [55]. This method takes the advantage of the difference between the amide bond of the innermost N-linked GlcNAc and the glycosidic bond among the rest of the sugars, where TFMS can cleave the glycosidic bonds but not the amide bond. Therefore, after cleavage, the innermost GlcNAc remains on the peptide to serve as a tag for N-glycosylation site mapping. Although the tag is the same as that from Endo H treatment, this chemical method has the advantage of not being affected by the compositional and structural variation of the glycans, which can lead to a much broader glycoproteome coverage.

Combining this method with lectin enrichment of glycopeptides, the authors identified 555 N-glycosylation sites from 219 glycoproteins without further glycopeptide fractionation. The authors also compared this method to the Endo H method, and demonstrated that this chemical method outperformed the other by a large margin. Following this study, Ma et al. developed a strategy used TFA to deglycosylate glycopeptides with the assistance of microwave heating, which shortened the treatment time to merely ten minutes [56]. Combining the new strategy with ZIC-HILIC enrichment and higher-energy collisional dissociation (HCD) fragmentation, they identified a total of 257 N-glycosylation sites and 144 N-glycoproteins from healthy human serum. Although chemical deglycosylation methods are generally more universal than enzymatic methods, the harsh deglycosylation conditions sometimes would damage the peptide backbone and thus sabotage the glycosylation site identification.

1.2.2.2 A common tag for protein O-glycosylation

Although it is well known that protein O-glycosylation plays crucial roles in biological systems [57-59], the tools to study protein O-glycosylation is relatively under-represented

compared to N-glycosylation. As one of the methods that can deglycosylate O-glycans and leave a mass tag on the peptides, β-elimination has been well studied in the literature. The general principle of this method is to perform alkaline-induced release of glycans and leaves an alkene group on the deglycosylated site. The carbon-carbon double bond is reactive and susceptible to nucleophilic attack, and therefore, reduction was then performed on the alkene to stabilize the structure and create a mass tag for proteomic studies.

More than two decades ago, Greis et al. designed a β-elimination-based strategy to analyze O-GlcNAc-modified glycopeptides using MS. They demonstrated that β-elimination can create a mass shift for O-GlcNAcylated peptides, converting the previously glycosylated serine and threonine residues to alanine and 2-aminobutyric acid, respectively, and thus can be used for detection and site mapping of glycopeptides in complex samples [60]. The same research group further developed this method by coupling β-elimination with Michael addition of dithiothreitol, termed BEMAD [45]. This method creates a 136.2 Dalton mass shift through the loss of the glycans and addition of the DTT molecule, which can serve as a common mass tag for O-glycosylation (especially O-GlcNAcylation) site analysis. In the literature, there are also many other reports on β-elimination-based methods for maping protein O-glycosylation [9, 44, 61-69]. For instance, Rademaker et al. used NH_4OH to initiate β-elimination, and after completion, NH_3 was incorporated onto the amino acid residue from which the glycan was released. This method yielded a unique mass tag for database searching and was proved to be effective for as low as 1 pmol of starting material [70]. Despite β-elimination-based methods hold the potential to generate common mass tags for O-glycosylation sites, the major drawback is that the reaction conditions are relatively harsh, which induces significant degradation of peptides.

Unlike N-glycans that can usually be removed by Endo H or PNGase F prior to LC-MS/MS analysis, O-glycans are more daunting to be released by enzymatic methods. Despite this fact, attempts were still made on using enzymes to deglycosylate O-glycans. Hägglund et

al. combined two enzymatic deglycosylation strategies to identify both core fucosylated N-glycans and O-glycosylation sites from human plasma proteins. They carried out PNGase F removal of N-glycans in $H_2{}^{18}O$ first, then Endo D and Endo H, along with several exoglycosidases (β-galactosidase, neuraminidase and N-acetyl- -glucosaminidase), were used to cleave the glycosidic bond between the two GlcNAc residues in N-glycans, leaving only one GlcNAc residue with potential fucosyl side chain on the peptide. Although initially this strategy was devised for N-glycosylation analysis, several O-glycosylated peptides were also found with a single GalNAc attached to the modification site, which was attributed to partial de-O-glycosylation by the combination of endo- and exoglycosidases [23].

In addition to the enzymatic and chemical methods mentioned above, another very elegant method was developed by Steentoft et al., which employs zinc-finger nuclease (ZFN) to genetically engineer cells, simplifying the O-glycan structures to create a common mass tag [10]. The modified cell lines are named SimpleCell lines. They applied ZFN targeting to modify the O-glycan elongation pathway in human cells, and thus truncate the human glycans. SimpleCell lines with homogenous O-glycosylation were generated. These cell lines solely express GalNAcα (Tn) or NeuAcα2-6GAlNAcα (STn) O-glycans, allowing O-glycopeptides to be easily enriched by lectins. The glycopeptide sequencing process is also greatly simplified due to the common tags. A total of >100 O-glycoproteins with >350 O-glycosylation sites were identified by combining this method with nano-flow liquid chromatography-mass spectrometry (nLC-MS/MS) with electron transfer dissociation (ETD) fragmentation. This method has opened up a new avenue to analyze the O-glycoproteome. With the development in gene editing techniques in recently years, similar strategies should have broader applications in protein modification studies.

Following their first publication, Steentoft et al. further optimized the experimental conditions and mapped nearly 3,000 glycosylation sites in over 600 O-glycoproteins from 12

human cell lines [71]. These cells were from different organ origins and the glycoproteomes are considerably different across these cell lines. Meanwhile, they also improved NetOGlyc4.0 as a tool for O-glycosylation prediction. In summary, this study contains very comprehensive O-glycosylation information.

1.3 Glycoprotein Dynamics

Glycosylation is a reversible protein modification and glycoproteins are dynamic in and outside of cells. However, investigating glycoprotein dynamics can be quite challenging. The rapid advancement in glycoproteomics and multiplexed proteomics has provided the exciting possibility. The following subsection review several studies on MS-based proteomic investigation of glycoprotein dynamics.

1.3.1 O-GlcNAcylated protein dynamics

There are several methods reported in the literature to study glycoprotein dynamics and most of these studies were conducted in recent years. For instance, Wang et al. metabolically labeled O-GlcNAc by feeding cells with $^{13}C_6$-glucose. The isotopically labeled glucose metabolized into ^{13}C-labeled UDP-GlcNAc through the hexosamine biosynthetic pathway, and eventually labeled O-GlcNAcylated proteins. They then employed the boronic acid-based glycoprotein enrichment method to enrichment O-GlcNAcylated peptides for quantitative proteomics analysis. Through this strategy, protein O-GlcNAcylation turnover rates were determined. They identified 105 O-GlcNAcylated peptides from 42 proteins, and determined the turnover rates of 20 O-GlcNAcylated peptides from 14 proteins in the HeLa cells [72].

1.3.2 N-glycoprotein dynamics

Recently Xiao et al. developed a method that integrated isotopic labeling, chemical enrichment, and multiplexed proteomics to perform glycoprotein degradation and synthesis rates simultaneous [73]. In this study, cells were cultured in media containing heavy lysine and arginine, then chased in media with all light amino acids for different duration before being harvested. Then the proteins were extracted, digested, and the resulting peptides were subjected to boronic acid-based glycopeptide enrichment. The enriched glycopeptides from cells harvested at different time points were labeled by tandem mass tags (TMT) reagents, and then analyzed by LC-MS/MS. Due to the fact that after chasing cells with light media, the abundance of existing heavy amino acid-labeled proteins decrease and the light proteins increase, heavy glycoproteins were used for the determination of degradation rates and light glycoproteins for synthesis rate investigations. The synthesis rates of 847 N-glycoproteins and degradation rates of 704 N-glycoproteins were calculated in this study.

1.4 Applications of MS-Based Glycoproteomics

Protein glycosylation is one of the most complex modifications in all types of organisms. Characterizing glycosylation events in various samples will certainly lead to a deeper understanding of many cellular processes and better solutions of biomedical problems. The strategies developed in the field of glycoproteomics have been widely applied to investigate a great variety of experimental subjects, from prokaryotic cells to eukaryotic cells, from plants to animals, and from cultured cells to clinical samples. In this section, we review the applications of MS-based glycoproteomics.

1.4.1 Yeast

Yeast (*Saccharomyces cerevisiae*) is most commonly used in biology laboratories. Although it serves as an excellent model system for eukaryotic cells, the global analysis of glycoproteins in yeast is still challenging. Unlike human glycan structures, the yeast glycans are mostly the high-mannose type [74], and the molecular weight of these N-glycans can sometimes be very large. In addition, yeast also has cell walls, and many cell wall proteins are heavily mannosylated. Investigating protein glycosylation in yeast has long been intriguing to researchers.

In 2009, Schulz and Aebi designed a novel strategy to quantify glycosylation site occupancy in yeast. They enriched glycoproteins bound to the yeast polysaccharide cell wall, and released the glycans using Endo H, which also creates a mass tag at the same time. The peptides and glycopeptides were analyzed by LC-MS/MS. Their experimental results also revealed that the paralogues Ost3p and Ost6p have crucial roles in efficient glycosylation of distinct defined glycosylation sites [75].

Bailey and Schulz demonstrated that adding a protein deglycosylation step prior to enzymatic protein digestion can systematically improve N-glycoprotein identification in yeast lacking Alg3p. By treating the proteins with PNGase F before AspN or trypsin digestion, the quality of yeast cell wall proteome identification was improved [76].

As discussed above, Chen et al. developed a chemical deglycosylation method to study lectin-enriched yeast glycoproteome, and 555 protein N-glycosylation sites were identified on 250 glycoproteins in yeast cells. They later devised a boronic acid-based enrichment strategy to universally analyze glycoproteins in yeast, and identified 816 N-glycosylation sites from 332 glycoproteins.

Xiao et al. performed quantification of the proteome and glycoproteome changes in yeast cells with or without the tunicamycin treatment. A total of 4,259 proteins and 135

glycoprorteins were quantified. More than 5% of the proteins were found to be decreased by at least 2-fold and 168 out of 465 glycopeptides were down-regulated due to the protein N-glycosylation inhibition effect of tunicamycin [77]. Smeekens et al. identified and quantified secreted yeast proteins (including glycoproteins) from tunicamycin-treated cells. The secreted yeast glycoproteins were separated from cells through mild washing and centrifugation that avoided cell death, limiting the impact of the intracellular proteins on the secretome analysis. A total of 27 glycoproteins were quantified and 26 of them were down-regulated, testifying that the secretion of some proteins is regulated by glycosylation [78].

1.4.2 Plant

Since glycosylation is an essential protein modification in all eukaryotic cells that regulates a variety of cellular processes, plant glycosylation is also of importance to investigate. Similar to mammalian cells, most proteins of the extracellular and endomembrane systems are glycosylated by N-linked oligosaccharides in plants. Protein N-glycosylation impacts not only their physicochemical properties, but also their biological functions [79]. Despite the fact that protein glycosylation is relatively conserved across all eukaryotic species, glycosylation in plant cells does have its uniqueness compared to mammalian cells. However, compared to lots of endeavors been put into glycosylation studies in mammalian cells, plant glycosylation is still relatively not as well-studied. While there is no universal protocol or procedure for plant proteomics, in general, a plant proteomics experiment typically involves the following steps: cell/tissue preparation, protein extraction and digestion, peptide separation, and MS-based identification [80]. For modification studies, there usually is an additional enrichment step. The protein extraction procedure for plant can be very different from mammalian protein extraction because (1) plants have a great variety of tissues and the extraction procedure can be different for different tissues; (2) protein concentrations in plant samples are usually low and yet plant

tissues have high amount of proteases; (3) many plant-originated chemicals, such as the polysaccharides in cell wall, lipids, pigments, and metabolites, can greatly interfere with several steps of the typical proteomics workflow.

Proteomic studies of plant glycoproteins have emerged since early 2000s. In 2003, Andon et al. performed a proteomic study on the mannose-binding proteins in rice (*Oryza sativa*). Instead of directly analyzing glycoproteins, they studied the proteins that are involved rice sugar metabolism, including several rice lectins [81]. The method they used to enrich these proteins is column affinity chromatography, and α-D-mannose was used as the ligand to pack the column and bind the desired glycoproteins. Saravanan and Rose evaluated several extraction techniques to analyze proteins in recalcitrant plant tissues, and found that compared to acetone-based protein precipitation methods, phenol-based methods gave higher numbers of protein and glycoprotein identifications as shown in their results [82].

Wimmer et al. designed an method to isolate "membrane-associated, boron-interacting proteins", such as glycoproteins, glycosylphosphatidylinositol (GPI)-anchored proteins, using boronate affinity chromatography. Resin-immobilized phenylboronic acid was employed to capture glycoproteins from root microsomal preparations of arabidopsis (*Arabidopsis thaliana*) and maize (*Zea mays*). These proteins were then analyzed by 2D-gel electrophroresis and matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) MS [83].

Albenne et al. systematically studied plant cell wall proteins including glycoproteins. They established a workflow to prepare cell wall peptides for MS analysis, and developed a bioinformatics tool to interpret the data. In that study, N-glycosylation was found on peroxidases, such as PER32. Their MS data also provided insights into N-glycan structures and facilitated protein glycosylation prediction [84].

In later studies, especially those published in recent years, lectin-based enrichment strategies were the most widely used to enrich plant glycoproteins. Typically, one or several

lectins are immobilized on the resin to pack a column, and then lectin affinity chromatography is performed to separate and purify plant glycoproteins. Through this enrichment, researchers were able to analyze both N- and O-glycosylation. For instance, Rose and co-workers carried out a comparative study to analyze N-glycoproteins in tomato fruit. In one experiment, by combining three lectins, namely Con A, snowdrop lectin, and lentil lectin to enrich glycoproteins, they identified 448 putative N-glycoproteins. In the other parallel experiment using lectin affinity chromatography plus hydrophilic interaction chromatography as the enrichment method, 318 putative N-glycosylation sites on 230 N-glycoproteins were identified. Of note, 17 N-glycan structures were also studied [85]. In another study on tomato fruit glycoproteins, Thannhauser and co-workers combined ConA lectin affinity chromatography and LC-MALDI-MS/MS, and analyzed glycoproteins that involved in biological processes such as carbohydrate metabolism, proteolytic activity, oxidative catabolism, phosphatase activity, nucleic acid catabolism/transcriptional regulation [86].

1.4.3 Mouse

Mouse has been an excellent model for researchers to study biology in complex and dynamic physiological systems. It is frequently used in biomedical research as a substitute for human subjects. To overcome its several limitations in modeling human disease, many clinically-relevant mouse models were generated to mimic the cellular processes of human diseases [87]. Through these models, molecular mechanisms and clinical responses of diseases were investigated. The data collected in these studies contribute immensely to developing new therapies to cure diseases. Due to the biomedical importance of mouse models, analyzing mouse glycoproteins have been the focus of many reports in the literature.

Besides the large-scale mouse N-glycoproteome study performed by Mann and co-workers that is described in the first section of this review, there are also many other impactful

glycosylation studies on mouse samples. Cima et al. used a glycoproteomics approach to aid in the discovery of serum biomarkers for prostate cancer. Hydrazide chemistry coupling with solid phase extraction was employed for the enrichment, and through this, they identified 775 N-glycoproteins from sera and prostate tissue of wild-type and *Pten*-null mice [88]. Label-free quantification was then performed, and the results demonstrated that *Pten* deletion led to changes in prostate and serum glycoproteomes. Based on these results, further targeted-proteomics and bioinformatics studies were carried out to screen out potential biomarkers. This study is an excellent example of how rational design of proteomic analysis in mouse models can lead to the discovery of biomarkers.

Goldberg et al. developed a program named Peptoonist that enabled automated N-glycopeptide identification, which can identify glycopeptides and also annotate glycan composition. To validate this strategy, they prepared mouse models and then isolated proteins from these mouse ovaries and tested the program [89]. Zhang et al. optimized a protocol for the enrichment of both glycopeptides and phosphopeptides through electrostatic repulsion hydrophilic interaction chromatography (ERLIC), and then analyzed 922 glycosylation sites on 544 unique glycoproteins, and 915 phosphorylation sites on 383 phosphoproteins from mouse brain membrane [90].

Many mouse O-glycosylation studies were also reported in the literature. Alfaro et al. devised a strategy combining metabolic labeling and chemical/enzymatic photochemical cleavage to study mouse brain O-GlcNAcylated proteins as discussed above [41]. From 100 μg tryptic peptides, they were able to identify 458 O-GlcNAc sites and 195 glycoproteins. Palmisano et al. combined titanium dioxide enrichment with HILIC to investigate protein modifications during mouse brain development, and were able to identify 3246 unique formerly sialoglycopeptides. More than 10% of these peptides were found to be differentially regulated in the development process [91]. Although O-glycosylation studies in mouse are still quite

challenging now, with the development in enrichment strategies and MS fragmentation methods, we expect more applications of O-glycosylation analyses in mouse models will come up and greatly expand our knowledge of the functions of O-glycoprotein in cells with healthy or diseased states.

1.4.4 Human cell lines and intact glycopeptide analyses

The commonly used human cell lines along with stem cells [92] are the most widely used models in the glycoproteomics field. Using human cell lines, glycoproteome analysis has advanced from glycoprotein identification-centric studies to glycosylation site localization and site-specific structural characterization of glycan-containing peptides and proteins. Although bottom-up and top-down glycoproteomics are both developing rapidly, glycopeptide-based glycoproteomics is more commonly used to study the glycosylation sites and glycan structures on glycopeptide. In addition, stoichiometry and glycan occupancy information can also be obtained from this working route. To study glycan-containing glycopeptides in human cells, different fragmentation methods have their advantages and disadvantages. CID can generate abundant B- and Y- ions that are useful to identify the glycan composition and structure. However, CID, especially resonance-type CID technique cannot produce adequate and reliable fragments for glycosylation site and peptide sequence determination. Beam-type higher-energy collisional dissociation (HCD) usually can generate enough b- and y- ions for peptide identification, but the frequent detachment of glycans renders the site localization hardly reliable. ECD (electron-capture dissociation) and ETD (electron-transfer dissociation) type fragmentation methods yield c- and z- ions to determine the peptide backbone sequence and glycosylation site, but not enough information can be gathered to study the actual glycan side chain substructure.

Due to the drawbacks of each technology, attempts were made to perform MS^3 or combine different fragmentation methods to analyze intact glycopeptides [93, 94]. These attempts can improve peptide backbone fragmentation efficiency or allow pre-selection of glycopeptides, thus filtering out non-glycopeptides and allocating more analysis time for glycopeptides. Parker et al. combined glycomics and glycoproteomics to study the N-glycoproteome [95]. They analyzed the glycosylation sites and glycan structures separately. The glycopeptides were deglycosylated with PNGase F and identified by LC-MS/MS first to accurately localize the N-glycosylation site, then glycan-containing form of the same peptide and glycans were analyzed to find out the glycan composition. Eventually the data collected from both analyses were combined to reconstruct the intact-glycopeptides. With both glycome and glycoproteome information, a total of 863 unique N-glycopeptides from 161 glycoproteins were studied. Other attempts were focused on developing computational algorithms for site-specific assignments of intact glycopeptides. Zhang and co-workers developed software named GPQuest to analyze intact glycopeptides using the data collected from HCD-LC-MS/MS. This software firstly generates a spectral library of glycosite-containing peptides from MS analysis using HCD as the fragmentation method. Intact glycopeptides are then selected based on the oxonium ions, and the spectra are compared with the library generated from glycosite-containing peptides. This step assigns MS/MS spectra of intact glycopeptides to specific glycosite-containing peptides. The glycans were then determined by calculating the mass shift between the precursor ion of intact glycopeptide and the glycosite-containing peptide, and match the mass difference to a glycan database [96, 97].

In 2012, Frese et al. developed a novel method that combined HCD with ETD (termed EThcD) to improve peptide backbone fragmentation. After the initial electron-transfer dissociation, all ions were fragmented by collision induced dissociation. Therefore, in the end, b-, y-, c-, and z-ions can be observed in the same spectrum [98]. They later applied it for

phosphorylation analysis [99]. This strategy was recently adopted and optimized by Yu et al. to analyze intact glycopeptides. After lectin and HILIC enrichment, the glycopeptides were analyzed by EThcD MS [100]. Since both glycosidic and amide bonds were cleaved, rich information on glycan structure and peptide sequence were obtained. The authors also compared the number of glycoforms identified from EThcD or HCD alone, and demonstrated that a greater number of glycoforms were observed using EThcD.

1.4.5 Clinical samples

Glycosylation can changes chemical and physical properties of proteins, and regulate their binding and interactions with ligands or extracellular matrix, which is important in many biological processes. Aberrant glycosylation patterns reflect abnormal cellular processes and can be used to monitor disease status [101]. Glycoproteomics has been applied in the research of a great variety of diseases, such as hepatitis, cancer, and infectious diseases. For instance, an increase on sialylation is a common feature of cancer cells [102]. With the increased levels of sialic acid residues in cells, abnormal glycosylation patterns such as STn antigens start to present on cells [103]. Also, as described by the Warburg effect, cancer cells predominantly produce energy through glycolysis instead of mitochondrial oxidative phosphorylation. This leads to a great increase in glucose uptake, which elevates the level of UDP-GlcNAc, the end product of hexosamine biosynthetic pathway. Correspondingly protein O-GlcNAcylation may increase dramatically. Due to this reason, hyper O-GlcNAcylation is also one of the hallmarks of cancer progression in cells [104].

Currently, one of the most important glycoproteomic applications in clinical samples is to find new disease biomarkers [105]. Unlike traditional strategies that screen a single or a small number of potential biomarker(s), monitoring a larger group of candidates using glycoproteomic approaches can often result in higher sensitivity and specificity. For example,

Ahn et al. attempted to screen biomarkers for small cell lung cancer through a glycoproteomic approach. They enriched glycoproteins using a lectin column, and then analyzed the glycoproteome changes through both label-free quantification and multiplex proteomics. The results demonstrated that the expression and glycosylation changes in fucosylated proteins, such as paraxonase 1, might serve as serological markers for small cell lung cancer [106]. Qiu et al. profiled plasma glycoproteins in order to find biomarkers for colorectal cancer. They combined lectin glycoarray with LC-MS to determine the glycan patterns from the plasma samples from 9 normal, 5 adenoma, and 6 colorectal cancer patients, and found several proteins with elevated sialylation and fucosylation as potential biomarkers of colorectal cancer. These markers were then validated by lectin blotting of plasma samples from thirty patients [107].

Halim et al. devised several applicable strategies to analyze glycoproteins with glycan structural information in clinical samples. In 2012, they performed a study to analyze N- and O-linked glycoproteins in urine samples. After eliminating the interferences by dialysis, they performed the enrichment with hydrazide chemistry, and characterized intact glycopeptides by CID and ECD. Later they combined PNGase F pretreatment and automated CID-MS2/MS3 fragmentation for glycopeptide identification, and ECD/ETD for glycosylation site localization, to analyze intact O-glycopeptides in glycoproteins from human cerebrospinal fluid samples. This study provided guidance for future research of finding O-glycoproteins as potential biomarkers. In 2014, by using synthetic glycopeptides, urine samples, and peptides from human cerebrospinal fluid samples, they found that oxonium fragmentation patterns can be used to differentiate O-GlcNAc from O-GalNAc, which is a important discovery for intact O-glycopeptide analysis [108].

The applications of glycoproteomics in clinical samples are far more than what we described in this section, there are many comprehensive reviews that discuss the importance of glycoproteomic technologies in biomedical research [109-114]. Due to the irreplaceable roles of

glycoproteins play in various diseases, high-throughput glycoproteomic technologies will continue to be developed to enable large-scale and sensitive analysis of glycopeptides and glycan structures. Foreseeably, glycoproteomics will help shape the directions of future glycoscience research.

1.5 Conclusions

The tremendous development in chemical biology and MS technologies have allowed for rapid advancements in the glycoproteomics field. Here, we reviewed the chemical and enzymatic methods for glycoprotein analysis. We first included widely used methods for the enrichment of glycopeptides/proteins, such as lectin, hydrazide chemistry, HILIC, click chemistry-based and boronic acid-based enrichment methods. Following this, we discussed several enzymatic and chemical methods for the generation of common mass tags. This section focused on how researchers have attempted to solve the two challenging problems in glycoproteme analysis (low abundance of many glycoproteins and heterogeneity of glycan structures). With the methods described above and multiplexed proteomics, large-scale investigation of glycoprotein dynamics has come to realization. Therefore, The methods recently reported for studying the dynamics of the whole glycoproteome were also included. In the last section, we discussed many applications of MS-based glycoproteomics in different species, from yeast, plant, mouse models, to clinical samples, and intact glycopeptide analysis in human cells.

With these methods and many other strategies underway, much valuable information about protein glycosylation have been and will be obtained, including glycoprotein identification, glycosylation site localization, and glycan structure elucidation, glycosylation stoichiometry investigation, and glycoprotein analysis with spatial and temporal information. Global analysis of protein glycosylation will undoubtedly provide important data that can have

tremendous impact on biochemical and biomedical research. In addition, as we are entering an era where computational power advances immensely, we expect to see further developments in mass spectrometry. Next generation mass spectrometers will be able to enable more sensitive peptide sequencing, faster analyzing speed, and more suitable fragmentation techniques.

Higher computational power also will allow for new bioinformatics tools to aid in glycoproteomics. New and innovative software will be developed to help us quickly and accurately identify glycopeptides, especially intact glycopeptides, and to provide us more valuable information regarding protein glycosylation sites and glycan structures. With the development of hardware and software, and effective chemical and enzymatic methods, it is expected that the field of glycoproteomics will grow exponentially in the next decade.

1.6 References

1. Ramachandran, P. et al. Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* **5**, 1493-1503 (2006).
2. Wei, X., Dulberger, C. & Li, L.J. Characterization of murine brain membrane glycoproteins by detergent assisted lectin affinity chromatography. *Anal. Chem.* **82**, 6329-6333 (2010).
3. Zheng, J.N., Xiao, H.P. & Wu, R.H. Specific identification of glycoproteins bearing the Tn antigen in human cells. *Angew. Chem.-Int. Edit.* **56**, 7107-7111 (2017).
4. Wohlgemuth, J., Karas, M., Eichhorn, T., Hendriks, R. & Andrecht, S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Anal. Biochem.* **395**, 178-188 (2009).
5. Amari, F. et al. Lectin electron-microscopic histochemistry of the pseudoexfoliative material in the skin. *Invest Ophth Vis Sci* **35**, 3962-3966 (1994).
6. Zielinska, D.F., Gnad, F., Wisniewski, J.R. & Mann, M. Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907 (2010).
7. Zielinska, D.F., Gnad, F., Schropp, K., Wisniewski, J.R. & Mann, M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol Cell* **46**, 542-548 (2012).
8. Darula, Z. & Medzihradszky, K.F. Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol Cell Proteomics* **8**, 2515-2526 (2009).
9. Durham, M. & Regnier, F.E. Targeted glycoproteomics: Serial lectin affinity chromatography in the selection of O-glycosylation sites on proteins from the human blood proteome. *J Chromatogr A* **1132**, 165-173 (2006).
10. Steentoft, C. et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* **8**, 977-982 (2011).
11. Vosseller, K. et al. Quantitative analysis of both expression and serine/threonine post-translational modifications through use of beta-elimination/Michael addition with DTT (BEMAD). *Faseb J* **18**, C148-C148 (2004).
12. 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).
13. Liu, T. et al. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **4**, 2070-2080 (2005).
14. Wang, L. et al. Mapping N-Linked glycosylation sites in the secretome and whole cells of aspergillus niger using hydrazide chemistry and mass spectrometry. *J. Proteome Res.* **11**, 143-156 (2012).
15. Yu, L.Y., He, H.B., Hu, Z.F. & Ma, Z.J. Comprehensive quantification of N-glycoproteome in Fusarium graminearum reveals intensive glycosylation changes against fungicide. *J Proteomics* **142**, 82-90 (2016).
16. Nilsson, J. et al. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods* **6**, 809-U826 (2009).
17. Halim, A., Nilsson, J., Ruetschi, U., Hesse, C. & Larson, G. Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol Cell Proteomics* **11** (2012).

18. Halim, A., Ruetschi, U., Larson, G. & Nilsson, J. LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* **12**, 573-584 (2013).

19. Taga, Y., Kusubata, M., Ogawa-Goto, K. & Hattori, S. Development of a novel method for analyzing collagen O-glycosylations by hydrazide chemistry. *Mol Cell Proteomics* **11** (2012).

20. Klement, E., Lipinszki, Z., Kupihar, Z., Udvardy, A. & Medzihradszky, K.F. Enrichment of O-GlcNAc modified proteins by the periodate oxidation-hydrazide resin capture approach. *J. Proteome Res.* **9**, 2200-2206 (2010).

21. Zauner, G., Deelder, A.M. & Wuhrer, M. Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **32**, 3456-3466 (2011).

22. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O.N. & Roepstorff, P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **3**, 556-566 (2004).

23. Hagglund, P. et al. An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J. Proteome Res.* **6**, 3021-3031 (2007).

24. Nettleship, J.E. Hydrophilic interaction liquid chromatography in the characterization of glycoproteins. *Chromatogr Sci Ser* **103**, 523-550 (2011).

25. Mysling, S., Palmisano, G., Hojrup, P. & Thaysen-Andersen, M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* **82**, 5598-5609 (2010).

26. Ding, W., Nothaft, H., Szymanski, C.M. & Kelly, J. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Mol Cell Proteomics* **8**, 2170-2185 (2009).

27. Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* **12**, 561-+ (2015).

28. Palaniappan, K.K. et al. Isotopic signature transfer and mass pattern prediction (IsoStamp): an enabling technique for chemically-directed proteomics. *Acs Chem Biol* **6**, 829-836 (2011).

29. Sun, S.S. et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat Biotechnol* **34**, 84-88 (2016).

30. Zhang, L.J. et al. Boronic acid functionalized core-satellite composite nanoparticles for advanced enrichment of glycopeptides and glycoproteins. *Chem-Eur J* **15**, 10158-10166 (2009).

31. Xu, G.B., Zhang, W., Wei, L.M., Lu, H.J. & Yang, P.Y. Boronic acid-functionalized detonation nanodiamond for specific enrichment of glycopeptides in glycoproteome analysis. *Analyst* **138**, 1876-1885 (2013).

32. Zeng, Z.F., Wang, Y.D., Guo, X.H., Wang, L. & Lu, N. On-plate glycoproteins/glycopeptides selective enrichment and purification based on surface pattern for direct MALDI MS analysis. *Analyst* **138**, 3032-3037 (2013).

33. Zhang, Q.B. et al. Enrichment and analysis of nonenzymatically glycated peptides: Boronate affinity chromatography coupled with electron-transfer dissociation mass spectrometry. *J. Proteome Res.* **6**, 2323-2330 (2007).

34. Chen, W.X., Smeekens, J.M. & Wu, R.H. A universal chemical enrichment method for mapping the yeast n-glycoproteome by mass spectrometry (MS). *Mol Cell Proteomics* **13**, 1563-1572 (2014).

35. Kayser, H. et al. Biosynthesis of a nonphysiological sialic-acid in different rat organs, using N-propanoyl-D-hexosamines as precursors. *J Biol Chem* **267**, 16934-16938 (1992).

36. Saxon, E. & Bertozzi, C.R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007-2010 (2000).

37. Palaniappan, K.K. & Bertozzi, C.R. Chemical glycoproteomics. *Chem Rev* **116**, 14277-14306 (2016).

38. Khidekel, N. et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* **3**, 339-348 (2007).

39. Wang, Z.H. et al. 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* **9**, 153-160 (2010).

40. Ma, J.F. et al. Comparative proteomics reveals dysregulated mitochondrial O-GlcNAcylation in diabetic hearts. *J. Proteome Res.* **15**, 2254-2264 (2016).

41. Alfaro, J.F. et al. Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *P Natl Acad Sci USA* **109**, 7280-7285 (2012).

42. Hu, P., Shimoji, S. & Hart, G.W. Site-specific interplay between O-GlcNAcylation and phosphorylation in cellular regulation. *Febs Lett* **584**, 2526-2538 (2010).

43. Ma, J.F. & Hart, G.W. O-GlcNAc profiling: from proteins to proteomes. *Clin Proteom* **11** (2014).

44. Vosseller, K. et al. Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* **5**, 388-398 (2005).

45. Wells, L. et al. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **1**, 791-804 (2002).

46. Wu, R.H. et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods* **8**, 677-U111 (2011).

47. Plummer, T.H., Elder, J.H., Alexander, S., Phelan, A.W. & Tarentino, A.L. demonstration of peptide-N-glycosidase-F activity in endo-beta-N-acetylglucosaminidase F Preparations. *J Biol Chem* **259**, 700-704 (1984).

48. Tarentino, A.L., Gomez, C.M. & Plummer, T.H. Deglycosylation of asparagine-linked glycans by peptide - N-glycosidase-F. *Biochemistry-Us* **24**, 4665-4671 (1985).

49. Kuster, B. & Mann, M. O-18-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal. Chem.* **71**, 1431-1440 (1999).

50. Kaji, H. et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* **21**, 667-672 (2003).

51. Tretter, V., Altmann, F. & Marz, L. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase-F cannot release glycans with fucose attached alpha-1-]3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem* **199**, 647-652 (1991).

52. Robbins, P.W. et al. Primary structure of the streptomyces enzyme "endo-beta-N-acetylglucosaminidase-H. *J Biol Chem* **259**, 7577-7583 (1984).

53. Edge, A.S.B., Faltynek, C.R., Hof, L., Reichert, L.E. & Weber, P. Deglycosylation of glycoproteins by trifluoromethanesulfonic acid. *Anal. Biochem.* **118**, 131-137 (1981).

54. Triguero, A. et al. Chemical and enzymatic N-glycan release comparison for N-glycan profiling of monoclonal antibodies expressed in plants. *Anal. Biochem.* **400**, 173-183 (2010).

55. 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).

56. Ma, C. et al. Convenient and precise strategy for mapping N-Glycosylation sites using microwave-assisted acid hydrolysis and characteristic ions recognition. *Anal. Chem.* **87**, 7833-7839 (2015).

57. Tian, E. & Ten Hagen, K.G. Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconjugate J* **26**, 325-334 (2009).

58. Jentoft, N. Why are proteins O-glycosylated. *Trends Biochem Sci* **15**, 291-294 (1990).

59. Moremen, K.W., Tiemeyer, M. & Nairn, A.V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Bio* **13**, 448-462 (2012).

60. Greis, K.D. et al. Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry. *Anal. Biochem.* **234**, 38-49 (1996).

61. Boysen, A. et al. A novel mass spectrometric strategy "BEMAP" reveals extensive O-linked protein glycosylation in enterotoxigenic escherichia coli. *Sci Rep-Uk* **6** (2016).

62. Hahne, H. et al. Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J. Proteome Res.* **12**, 927-936 (2013).

63. Hanisch, F.G., Teitz, S., Schwientek, T. & Muller, S. Chemical de-O-glycosylation of glycoproteins for application in LC-based proteomics. *Proteomics* **9**, 710-719 (2009).

64. Lee, Y. et al. Glycosylation and sialylation of macrophage-derived human apolipoprotein E analyzed by SDS-PAGE and mass spectrometry. *Mol Cell Proteomics* **9**, 1968-1981 (2010).

65. Nakano, M., Saldanha, R., Gobel, A., Kavallaris, M. & Packer, N.H. Identification of glycan structure alterations on cell membrane proteins in desoxyepothilone B resistant leukemia cells. *Mol Cell Proteomics* **10** (2011).

66. Overath, T. et al. Mapping of O-GlcNAc sites of 20 S proteasome subunits and Hsp90 by a novel biotin-cystamine tag. *Mol Cell Proteomics* **11**, 467-477 (2012).

67. Taylor, A.M., Holst, O. & Thomas-Oates, J. Mass spectrometric profiling of O-linked glycans released directly from glycoproteins in gels using in-gel reductive beta-elimination. *Proteomics* **6**, 2936-2946 (2006).

68. Vosseller, K. et al. O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* **5**, 923-934 (2006).

69. Whelan, S.A. & Hart, G.W. Proteomic approaches to analyze the dynamic relationships between nucleocytoplasmic protein glycosylation and phosphorylation. *Circ Res* **93**, 1047-1058 (2003).

70. Rademaker, G.J. et al. Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal. Biochem.* **257**, 149-160 (1998).

71. Steentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *Embo J* **32**, 1478-1488 (2013).

72. Wang, X.S. et al. A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol Cell Proteomics* **15**, 2462-2475 (2016).

73. Xiao, H.P. & Wu, R.H. Simultaneous quantitation of glycoprotein degradation and synthesis rates by integrating isotope labeling, chemical enrichment, and multiplexed proteomics. *Anal. Chem.* **89**, 10361-10367 (2017).

74. Kukuruzinska, M.A., Bergh, M.L.E. & Jackson, B.J. Protein glycosylation in yeast. *Annu Rev Biochem* **56**, 915-944 (1987).

75. Schulz, B.L. & Aebi, M. Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific N-glycosylation efficiency. *Mol Cell Proteomics* **8**, 357-364 (2009).

76. Bailey, U.M. & Schulz, B.L. Deglycosylation systematically improves N-glycoprotein identification in liquid chromatography-tandem mass spectrometry proteomics for analysis of cell wall stress responses in Saccharomyces cerevisiae lacking Alg3p. *J Chromatogr B* **923**, 16-21 (2013).

77. Xiao, H.P., Smeekens, J.M. & Wu, R.H. Quantification of tunicamycin-induced protein expression and N-glycosylation changes in yeast. *Analyst* **141**, 3737-3745 (2016).

78. Smeekens, J.M., Xiao, H.P. & Wu, R.H. Global Analysis of Secreted Proteins and Glycoproteins in Saccharomyces cerevisiae. *J. Proteome Res.* **16**, 1039-1049 (2017).

79. Rayon, C., Lerouge, P. & Faye, L. The protein N-glycosylation in plants. *J Exp Bot* **49**, 1463-1472 (1998).

80. Jorrin-Novo, J.V. Plant proteomics methods and protocols. *Methods Mol Biol* **1072**, 3-13 (2014).

81. Andon, N.L., Eckert, D., Yates, J.R. & Haynes, P.A. High-throughput functional affinity purification of mannose binding proteins from oryza sativa. *Glycobiology* **13**, 843-843 (2003).

82. Saravanan, R.S. & Rose, J.K.C. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics* **4**, 2522-2532 (2004).

83. Wimmer, M.A., Lochnit, G., Bassil, E., Muhling, K.H. & Goldbach, H.E. Membrane-associated, boron-interacting proteins isolated by boronate affinity chromatography. *Plant Cell Physiol* **50**, 1292-1304 (2009).

84. Albenne, C. et al. Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. *Mol Plant* **2**, 977-989 (2009).

85. Ruiz-May, E. et al. A comparative study of lectin affinity based plant N-glycoproteome profiling using tomato fruit as a model. *Mol Cell Proteomics* **13**, 566-579 (2014).

86. Catala, C., Howe, K.J., Hucko, S., Rose, J.K.C. & Thannhauser, T.W. Towards characterization of the glycoproteome of tomato (Solanum lycopersicum) fruit using Concanavalin A lectin affinity chromatography and LC-MALDI-MS/MS analysis. *Proteomics* **11**, 1530-1544 (2011).

87. Saxena, M. & Christofori, G. Rebuilding cancer metastasis in the mouse. *Mol Oncol* **7**, 283-296 (2013).

88. Cima, I. et al. Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *P Natl Acad Sci USA* **108**, 3342-3347 (2011).

89. Goldberg, D. et al. Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J. Proteome Res.* **6**, 3995-4005 (2007).

90. Zhang, H.M. et al. Simultaneous characterization of glyco- and phosphoproteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography. *Mol Cell Proteomics* **9**, 635-647 (2010).

91. Palmisano, G. et al. A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Mol Cell Proteomics* **11**, 1191-1202 (2012).

92. Stadlmann, J. et al. Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **549**, 538-+ (2017).

93. Zeng, W.F. et al. pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD-and CID-MS/MS and MS3. *Sci Rep-Uk* **6** (2016).

94. Wu, S.W., Liang, S.Y., Pu, T.H., Chang, F.Y. & Khoo, K.H. Sweet-Heart - An integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *J Proteomics* **84**, 1-16 (2013).

95. Parker, B.L. et al. Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J Proteome Res* **12**, 5791-5800 (2013).

96. Eshghi, S.T., Shah, P., Yang, W.M., Li, X.D. & Zhang, H. GPQuest: a spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact N-glycopeptides. *Anal. Chem.* **87**, 5181-5188 (2015).

97. Hu, Y.W. et al. GPQuest 3: A tool for large-scale and comprehensive glycosylation analysis on MS data. *Glycobiology* **27**, 1218-1218 (2017).

98. Frese, C.K. et al. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**, 9668-9673 (2012).

99. Frese, C.K. et al. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **12**, 1520-1525 (2013).

100. Yu, Q. et al. Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J Am Soc Mass Spectr* **28**, 1751-1764 (2017).

101. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3-49 (2017).

102. Vajaria, B.N., Patel, K.R., Begum, R. & Patel, P.S. Sialylation: an avenue to target cancer cells. *Pathol Oncol Res* **22**, 443-447 (2016).

103. Ju, T.Z., Otto, V.I. & Cummings, R.D. The Tn antigen-structural simplicity and biological complexity. *Angew. Chem.-Int. Edit.* **50**, 1770-1791 (2011).

104. Ma, Z.Y. & Vosseller, K. Cancer metabolism and elevated O-GlcNAc in oncogenic signaling. *J Biol Chem* **289**, 34457-34465 (2014).

105. Zhang, Y., Jiao, J., Yang, P.Y. & Lu, H.J. Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clin Proteom* **11** (2014).

106. Ahn, J.M. et al. Integrated glycoproteomics demonstrates fucosylated serum Paraoxonase 1 alterations in small cell lung cancer. *Mol Cell Proteomics* **13**, 30-48 (2014).

107. Qiu, Y.H. et al. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J. Proteome Res.* **7**, 1693-1703 (2008).

108. Halim, A. et al. Assignment of Saccharide Identities through Analysis of oxonium ion fragmentation profiles in LC MS/MS of glycopeptides. *J. Proteome Res.* **13**, 6024-6032 (2014).

109. Jankovic, M. Glycans as biomarkers: status and perspectives. *J Med Biochem* **30**, 213-223 (2011).

110. Kim, Y.J. & Varki, A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate J* **14**, 569-576 (1997).

111. Plomp, R., Bondt, A., de Haan, N., Rombouts, Y. & Wuhrer, M. Recent advances in clinical glycoproteomics of immunoglobulins (Igs). *Mol Cell Proteomics* **15**, 2217-2228 (2016).

112. Tian, Y. & Zhang, H. Glycoproteomics and clinical applications. *Proteom Clin Appl* **4**, 124-132 (2010).

113. Tousi, F., Hancock, W.S. & Hincapie, M. Technologies and strategies for glycoproteomics and glycomics and their application to clinical biomarker research. *Anal Methods-Uk* **3**, 20-32 (2011).

114. Ueda, K. Glycoproteomic strategies: From discovery to clinical application of cancer carbohydrate biomarkers. *Proteom Clin Appl* **7**, 607-617 (2013).

CHAPTER 2. A CHEMICAL METHOD BASED ON SYNERGISTIC AND REVERSIBLE COVALENT INTERACTIONS FOR LARGE-SCALE ANALYSIS OF GLYCOPROTEINS

Xiao, H. P., Chen, W. X., Smeekens, J. M., Wu, R. H., An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins, *Nature Communications*, accepted.

2.1 Introduction

Glycosylation is one of the most common and essential protein modifications in cells. It often determines protein folding, trafficking and stability, and regulates many cellular events, especially cell-cell communication, cell-matrix interactions, and cellular response to environmental cues[1-4]. Glycoproteins contain a wealth of information related to cellular developmental and diseased statuses[5, 6], and aberrant protein glycosylation is directly related to human disease, including cancer and infectious diseases[7-10]. Global analysis of protein glycosylation is critical in understanding glycoprotein functions and identifying glycoproteins as biomarkers and drug targets[10-12]. However, due to the low abundance of many glycoproteins and heterogeneity of glycans, it is extraordinarily challenging to comprehensively analyze glycoproteins in complex biological samples.

Currently mass spectrometry (MS)-based proteomics provides a unique opportunity to globally analyze protein modifications[13-22], including glycosylation[23-31]. However, effective enrichment prior to MS analysis is imperative for each type of protein modification. For example, with the maturity of phosphoprotein enrichment methods, the global analysis of

protein phosphorylation has advanced tremendously, from the identification of several hundred phosphorylation sites a decade ago to over ten thousand sites in recent studies[32-34].

In order to comprehensively analyze protein glycosylation in complex biological samples, several glycoprotein/peptide enrichment methods have been reported, including lectin-based[35, 36] and hydrazide chemistry-based methods[37, 38], and hydrophilic interaction liquid chromatography (HILIC)[39, 40]. Currently lectin-based methods are most commonly used to enrich glycopeptides prior to MS analysis. Due to the inherent specificity of lectins, each type of lectin can only recognize a specific glycan structure, and thus, no single lectin or a combination of several lectins can universally enrich all glycosylated peptides or proteins. HILIC has also been extensively used to enrich glycoproteins or glycopeptides based on the increased hydrophilicity of glycopeptides. However, this method lacks specificity because it cannot distinguish glycopeptides from many hydrophilic non-glycopeptides. Recently, two elegant methods, i.e. isotope-targeted glycoproteomics (IsoTaG)[41] and solid phase extraction of N-linked glycans and glycosite-containing peptides (NGAG)[42], have been reported. By using IsoTaG, 32 N-glycopeptides and over 500 intact and fully elaborated O-glycopeptides from 250 proteins across three human cell lines were identified[41]. NGAG was beautifully designed for N-glycopeptide enrichment, and 2,044 unique N-glycopeptides were identified in mammalian cells[42]. According to prediction and computational results, protein glycosylation is the most common modification[43, 44]. Despite the considerable progress that has been made in the past decade[35, 37, 41, 42, 45-50], there is still a substantial gap between the number of glycoproteins reported in the literature and those existing in complex biological samples. Effective enrichment of glycopeptides/glycoproteins will profoundly advance the global analysis of protein glycosylation through MS-based proteomics.

Previously, boronic acid (BA) was demonstrated to have great potential in universally enriching glycopeptides for the global analysis of protein glycosylation because of its

reversible covalent interactions with glycans[51, 52]. However, the method suffers from relatively weak interactions; therefore, low-abundance glycoproteins are not effectively enriched. In this work, we develop a new method to more effectively enrich glycopeptides, especially those of low-abundance, by greatly enhancing the interactions between boronic acid and glycopeptides. First, different boronic acid derivatives are tested, and benzoboroxole is found to be highly effective to enrich glycopeptides due to dramatically strengthened interactions. Second, based on the common features of a glycan containing multiple monosaccharides and one sugar bearing several hydroxyl groups, benzoboroxole conjugated dendrimer beads can synergistically interact with glycopeptides. The experimental results demonstrate that conjugating benzoboroxole to a dendrimer significantly increases the enrichment efficiency, even for glycopeptides only containing O-GlcNAc (N-acetyl glucosamine).

The novel method is applied for the global analysis of glycoproteins in yeast (*S. cerevisiae*), mouse brain tissue, and human cells (MCF7, HEK 293T and Jurkat). Over 1,000 N-glycosylation sites in yeast and 4,691 sites on 1,906 glycoproteins in human cells are identified, including many proteins with low abundance. The reversible nature of the interactions allows us to analyze intact O-glycopeptides with glycan structure information. We identify 234 O-mannosylated proteins in yeast and many glycoproteins with O-GlcNAc in human cells. These results demonstrate that the new method is universal and highly effective in enriching glycopeptides, especially from low-abundance glycoproteins that are normally of great biological importance. The current results also provide valuable information regarding glycoproteins in yeast and human cells to biological and biomedical research communities. Without sample restrictions, the current method can be applied to many other samples for glycoprotein analysis.

2.2 Experimental section

2.2.1 Materials

Complete protease inhibitors were purchased from Roche Applied Sciences and sequencing grade trypsin was from Promega. Dulbecco's Modified Eagle's Medium (DMEM), phosphate buffered saline (PBS), N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), 4-carboxy-2-nitrophenylboronic acid, (2-aminomethyl-5-fluoro) phenylboronic acid hydrochloride, 2-aminomethyl-4-fluorophenylboronic acid hydrochloride, trifluoroacetic acid (TFA), formic acid (FA), trimethylamine (TEA), piperidine, methanol, chloroform, dichloromethane (DCM), acetonitrile (ACN), and dimethylsulfoxide (DMSO) were from Sigma-Aldrich. 3-aminomethylphenylboronic acid hydrochloride was from Frontier Scientific Inc. 5-carboxybenzoboroxole and 1-hydroxy-7-azabenzotriazole (HOAt) were purchased from AK Scientific, Inc. (2,5-dioxopyrrolidin-1-yl) (2S)-2-(9H-fluoren-9-ylmethoxycarbonylamino)-6-[(2-methylpropan-2-yl)oxycarbonylamino] hexanoate (Fmoc-L-Lys(Boc)-OSu) and (S)-2,5-dioxopyrrolidin-1-yl 2,6-bis((tert-butoxycarbonyl) amino) hexanoate (Boc-Lys(Boc)-OSu) were from Ark Pharm, Inc. and Sigma-Aldrich. MagnaBind™ amine derivatized beads, MagnaBind™ carboxyl derivatized beads, and fetal bovine serum (FBS) were bought from Thermo Fisher Scientific.

2.2.2 Magnetic beads derivatization

MagnaBind^{TM} carboxyl (or amine) derivatized beads were washed with DMSO three times. EDC was added to the beads slurry and incubated end-over-end for 10 min; HOAt was subsequently added, and the reaction mixture was further incubated for one hour. HOAt-activated beads were washed with DMSO twice and incubated overnight with different amino boronic acids in DMSO containing 3.0% triethylamine (TEA). The boronic acid functionalized

beads were washed with DMSO twice and 20% ACN three times and stored in 20% ACN for further use.

For dendrimer boronic acid derivatization, the solvent containing the MagnaBind™ amine derivatized beads was gradually changed from water to isopropanol to finally DCM (Figure 2.1 and 2.2). Then Fmoc-L-Lys(Boc)-OSu was reacted with the amino beads in DCM containing 0.3% TEA overnight. On the following day the beads were washed with DCM three times, and the Boc protection group was removed by incubation of beads in 50% TFA in DCM at room temperature for two hours. The beads were washed with DCM three times and one time with 3% TEA in DCM. To continue the derivatization, Boc-Lys(Boc)-OSu was added to the bead DCM solution followed by the addition of TEA (final concentration 3.0%). The reaction was carried out at room temperature with end-over-end rotation overnight. Then the Boc group was deprotected by 50% TFA as mentioned above. The Boc-Lys(Boc)-OSu conjugation step was repeated twice. Then the Fmoc groups were removed by mixing the functionalized beads in 50% piperidine DCM solution at room temperature for 30 minutes. Finally, all free amine groups were coupled with 5-carboxybenzoboroxole through EDC HOAt chemistry as described above.

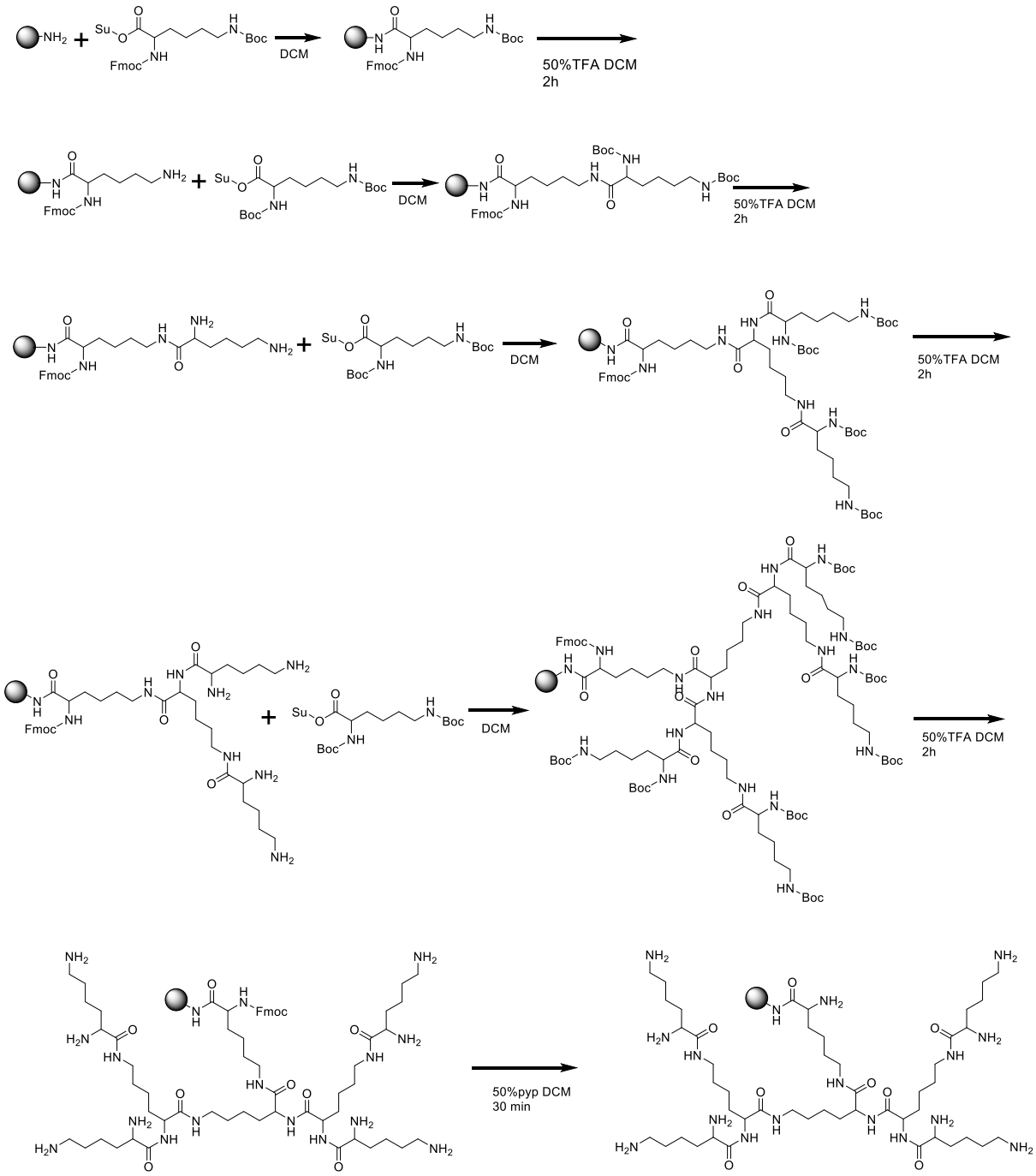


Figure 2.1 Synthesis of the dendrimer with functional amine groups.

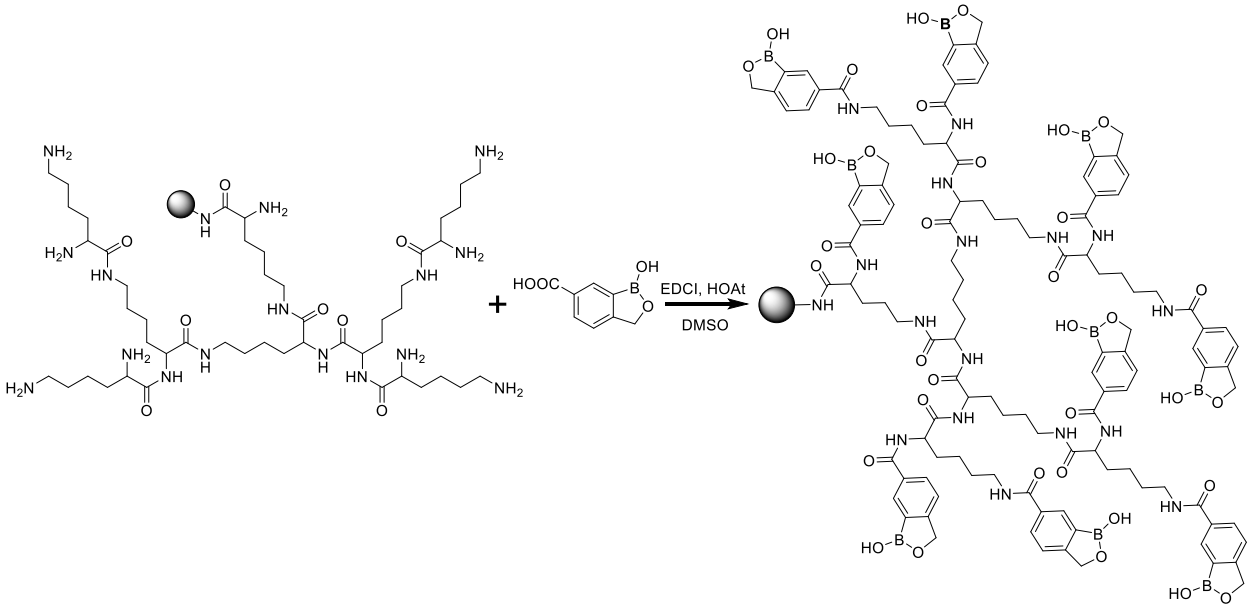


Figure 2.2 Conjugation of the boronic acid derivative, benzoboroxole, to the dendrimer.

2.2.3 Yeast cell culture and protein extraction

Yeast cells (strain BY4742, MAT alpha, derived from S288c) were grown in yeast extract peptone dextrose (YPD) media until they reached log-phase (optical density (OD) was about 1.0 at 600 nm). For biological duplicate experiments, cells were grown independently. Yeast cells were harvested by centrifugation and resuspended in a buffer containing 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.4, 150 mM NaCl, 0.5% sodium deoxycholate (SDC) and protease inhibitor cocktail (one tablet (complete mini, Roche) per 10 ml lysis buffer) at 4 ºC. Cells were lysed using the MiniBeadbeater (Biospec) at maximum speed, three cycles of 30 s each, with 2 min pauses between cycles to avoid overheating the lysates. After centrifugation, lysates were transferred to new tubes, and the protein concentration in the lysate was determined by BCA protein assay (Pierce).

2.2.4 Human cell culture, cell lysis and protein extraction

MCF7, HEK 293T and Jurkat cells (American Type Culture Collection (ATCC)) were cultured following the instructions provided by ATCC. Once MCF7 and HEK 293T cells reached 80% confluency, cells were washed with PBS twice and harvested by scraping. Jurkat cells were harvested by centrifugation and then washed with PBS. Cell pellets were suspended in the ice-cold RIPA buffer (50 mM HEPES, pH=7.4, 150 mM NaCl, 0.5% SDC, benzonase (25 U/mL), and protease inhibitor cocktail) and incubated end-over-end for 1 hour at 4 °C. After complete solubilization of nuclei and digestion of genomic DNA, the lysate was centrifuged at 25,000 *g* for 10 minutes. The supernatant was collected and the protein concentration was measured by BCA protein assay.

2.2.5 Protein extraction from mouse brain tissues

For mouse brain samples, brain tissues from two C57BL/6 mice (3 and 6 months) were frozen in liquid nitrogen and homogenized in the RIPA buffer mentioned above. The mixtures were incubated on ice for an hour, and then clarified by centrifugation at 5,000 g for 20 minutes. Half of the supernatants (~8 mg proteins per experiment) were used for protein glycosylation analysis.

2.2.6 Protein reduction, alkylation and digestion

Lysates from yeast, human cells, or mouse brain tissue were reduced with 5 mM dithiothreitol (DTT) (56 °C, 25 minutes) and alkylated with 15 mM iodoacetamide (RT, 30 minutes in the dark). Proteins were purified by the methanol-chloroform precipitation method. The purified proteins were digested with Lys-C (Wako) at a protein:enzyme ratio of ~100:1 in 50 mM HEPES, pH=8.2, 1.6 M urea, 5% ACN at 31 °C overnight, and then 10 ng/μL trypsin (Promega) for 4 h. Digestion was quenched by the addition of TFA to a final concentration of

0.1%, and precipitate was removed by centrifugation at 5,000 g for 10 min. The supernatant was collected, and peptides were purified using a Sep-Pak tC18 cartridge (Waters).

2.2.7 Glycopeptide enrichment

For boronic acid derivative experiments, mammalian peptides were dissolved in 100 mM ammonium acetate buffer and incubated for one hour with different boronic acid derivatized magnetic beads at room temperature. After incubation, the beads were washed with the binding buffer, and enriched peptides were eluted first by incubation with a solution containing ACN:H_2O:TFA (50:49:1) at 37 °C for 30 min. Then the peptides were eluted two more times through incubation with 5% formic acid at 56 °C for 5 min each time. For the enrichment of peptides from yeast, human cells or mouse brain tissues using DBA, ~10 mg of peptides were used in each experiment and incubated with DBA beads in DMSO containing 0.5% TEA, then washed five times using a buffer containing 50% DMSO and 50% 100 mM ammonium acetate (pH=11). Glycopeptides were then eluted as described above.

For lectin enrichment, ConA and WGA-conjugated agarose beads (Vector Laboratories) were washed five times using the enrichment buffer (20mM tris-base pH=7.4, 0.15 M NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, and 1 mM $MnCl_2$)[35]. Peptides were dissolved in the enrichment buffer, mixed with the lectin beads, and vortexed under 37 ^oC for an hour. The beads were then washed again with the enrichment buffer for five times before glycopeptide elution using the elution buffer (0.2 M α-methyl mannoside, 0.2 M α-methyl glucoside, 0.2 M galactose, and 0.5 M N-Acetyl-D-Glucosamine in PBS). The elution was performed twice with vortex for half an hour each, and the eluents were combined.

For HILIC enrichment, SeQuant® ZIC-HILIC SPE cartridges (the Nest Group) were washed with ten column volumes of 1.0% TFA in water, followed by three washes with the loading buffer (1.0% TFA in 80% ACN, 20% H_2O)[38-40]. Peptides were loaded onto the column

in the loading buffer using a slow flow rate. The flow-through was re-loaded onto the column once. The column was then washed with the loading buffer three times. Glycopeptides were eluted using 1.0% TFA in water three times, and the eluents were combined.

2.2.8 Glycopeptide PNGase F treatment and fractionation

The enriched samples were dried in a lyophilizer overnight. The completely dried samples were dissolved in 40 mM ammonium bicarbonate in heavy-oxygen water ($H_2{}^{18}O$) and treated with PNGase F (lyophilized powder from Sigma Aldrich) at 37 °C for 3 hours. For optimization experiments, after deglycosylation, peptide samples were purified using a stage tip. For all other experiments, enriched glycopeptides were desalted using a tC18 Sep-Pak cartridge, and then subjected to fractionation using high-pH reversed phase HPLC (pH=10). The sample was separated into 10 fractions using a 4.6×250 mm 5 μm particle reversed phase column (Waters) with a 40-min gradient of 5-50% ACN with 10 mM ammonium acetate. Every fraction was further purified with stage tip before LC-MS/MS.

2.2.9 LC-MS/MS analysis

Fractionated and purified peptide samples were resuspended in a solvent of 5.0% ACN and 4.0% FA, and 4 μL was loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 μm, 200 Å, 75 μm x 16 cm) using a WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler, Dionex). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 90-min gradient of 4-30% ACN (in 0.125% FA) and detected in a hybrid dual-cell quadrupole linear ion trap - orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher) using a data-dependent Top20 method. For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 20 MS/MS in the LTQ for the most

intense ions. The isolation window was 2 Da, which is the most commonly used, and the activation energy was 40% normalized collision energy (NCE), which was obtained through testing different NCEs to acquire the best results for the machine used here. Selected ions were excluded from further analysis for 90 s. Ions with a single or unassigned charge were not sequenced. Maximum ion accumulation times (Maximum IT) were 1000 ms for each full MS scan and 50 ms for MS/MS scans. For protein O-glycosylation analyses, the data was collected using a Q-Exactive Plus Orbitrap mass spectrometer with a two-hour LC gradient. Higher-energy collisional dissociation (HCD) was used as the fragmentation method with the following parameters: 10^6 AGC target for full MS and 2×10^5 AGC target for MS^2, 100 ms maximum IT, 2.0 Da isolation window, and 30% NCE. The dynamic exclusion time was set to 60 sec. Both full MS and MS^2 were collected in the Orbitrap cell with high mass accuracy and high resolution, which contribute to confident identification of O-glycopeptides.

2.2.10 Database searches and data filtering

The raw files were converted into mzXML format prior to the database search. The SEQUEST algorithm[53] (version 28) was used to search all MS/MS spectra against either a database containing sequences of yeast (*Saccharomyces cerevisiae*) proteins downloaded from SGD (http://www.yeastgenome.org/) or human (*Homo sapiens*) proteins downloaded from UniProt. The following parameters were used for the database search: 10 ppm precursor mass tolerance; 1.0 Da product ion mass tolerance; fully tryptic digestion; up to two missed cleavages; variable modifications: oxidation of methionine (+15.9949) and ^{18}O tag of Asn (+2.9883); fixed modifications: carbamidomethylation of cysteine (+57.0214). In order to estimate the false discovery rate (FDR) of peptide identification, both forward and reversed orientations of each protein sequence were listed in the database, and the target-decoy method was employed[54]. To distinguish between correct and incorrect peptide identifications, linear

discriminant analysis (LDA) was utilized with several parameters such as XCorr, ΔCn, and precursor mass error[55]. After scoring, peptides shorter than seven amino acid residues were discarded, and the remaining peptide spectral matches were controlled to have less than 1.0% FDR. When determining FDRs of the final data set, only glycopeptides were considered.

For O-glycopeptide identification, we used Byonic™ software. Some parameters are similar as above. For yeast intact O-glycopeptide analysis, up to ten mannoses per glycan were searched for raw files. In order to control false positive rates, every peptide was required to have ≤0.001 for 1 D PEP (one dimensional posterior error probability) and >4 for |Log Prob| (the absolute value of the log10 of the posterior error probability)[56]. The Score of identified glycopeptide must be higher than 300, and the mass accuracy is less than 10 ppm. The PEP takes into account 10 features, including the Byonic™ score, delta score, precursor mass error, digestion specificity, etc. Requiring |Log Prob| to be larger than 4 means the *P* value is $<10^{-4}$. These are very stringent criteria for filtering. For example, for protein O-GlcNAcylation analysis, after filtering, there was no reverse hit in the final datasets. For glycoproteins identified in each type of cells, we performed subcellular compartment analysis based on the protein location information downloaded from Uniprot (uniprot.org).

2.2.11 Protein glycosylation site localization

In order to evaluate the confidence of the glycosylation site assignment, a Modscore was calculated for each identified glycopeptides, which is similar to Ascore[57]. An algorithm considering all possible glycosylation sites of a peptide was used to generate the Modscore. It examines the presence or absence of MS/MS fragment ions unique to each glycosylation site and indicates the likelihood that the best site match is correct when compared with the next best match. Sites with Modscore ≥ 19 ($P \leq 0.01$) were considered to be confidently localized.

2.2.12 Data availability

The datasets generated during the current study are available in the PeptideAtlas repository (Dataset Identifier: PASS00980; Password: KV788a), https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/PASS_View?identifier=PASS00980. In total, there are 142 raw files (20 files for the yeast N-glycosylation duplicate experiments, 20 files for the MCF7 N-glycosylation duplicate experiments, 10 files for the HEK 293T N-glycosylation experiment, 10 files for the Jurkat N-glycosylation experiment, 22 files for the mouse brain N-glycosylation duplicate experiments, 20 files for the yeast O-mannosylation duplicate experiment, 20 files for the MCF7 O-GlcNAcylation duplicate experiments, 10 files for the HEK 293T O-GlcNAcylation experiment, 10 files for the Jurkat O-GlcNAcylation experiment).

2.3 Results

2.3.1 Enhancing glycopeptide enrichment with BA derivatives

Boronic acid can form reversible covalent bonds with sugars and has been extensively used for sugar detection[58-60]. Therefore, BA-based methods have great potential in universally enriching glycopeptides and glycoproteins, and the reversible nature of the interaction leaves enriched glycopeptides intact after the release. However, the interaction between BA and sugar is relatively weak, preventing the enrichment of low-abundance glycoproteins. To effectively enrich low-abundance glycoproteins, which often contain important information, it is critical to strengthen the interaction.

Reversible interactions between boronic acid and sugars have great potential to enrich glycopeptides/glycoproteins[52, 61, 62]. For global analysis of protein glycosylation, enrichment through strong interactions between boronic acid and glycopeptides is critical to cover low-abundance glycopeptides.

There are several major factors that govern the interactions between boronic acid (BA) and sugars, including the pK_a of BA, the solution pH, and steric/stereoelectronic effects[58, 63]. Although BA with a lower pK_a is expected to have greater binding affinities at neutral pH, this is not always true for glycopeptide enrichment. Therefore, several BA derivatives with various pK_a values were tested, and the optimum pH was found for each BA derivative. Previously, we demonstrated that phenylboronic acid conjugated beads were able to enrich glycopeptides from yeast whole cell lysates[52]. In yeast, high mannose glycans dominate, while glycans are more structurally diverse in mammalian cells. Here, we have designed and optimized a BA-based method to effectively enrich glycopeptides from mammalian cell lysates. The structures of several BA derivatives tested here are displayed in Figure 2.3a. Each of these BA derivatives was conjugated to magnetic beads containing either carboxylate or amine groups. After the –NH_2 or –COOH group reacts with the corresponding groups on the magnetic beads, the amide bond (–CONH–) between the beads and the benzene ring in each BA derivative should have a minimal effect on the optimum binding pH values.

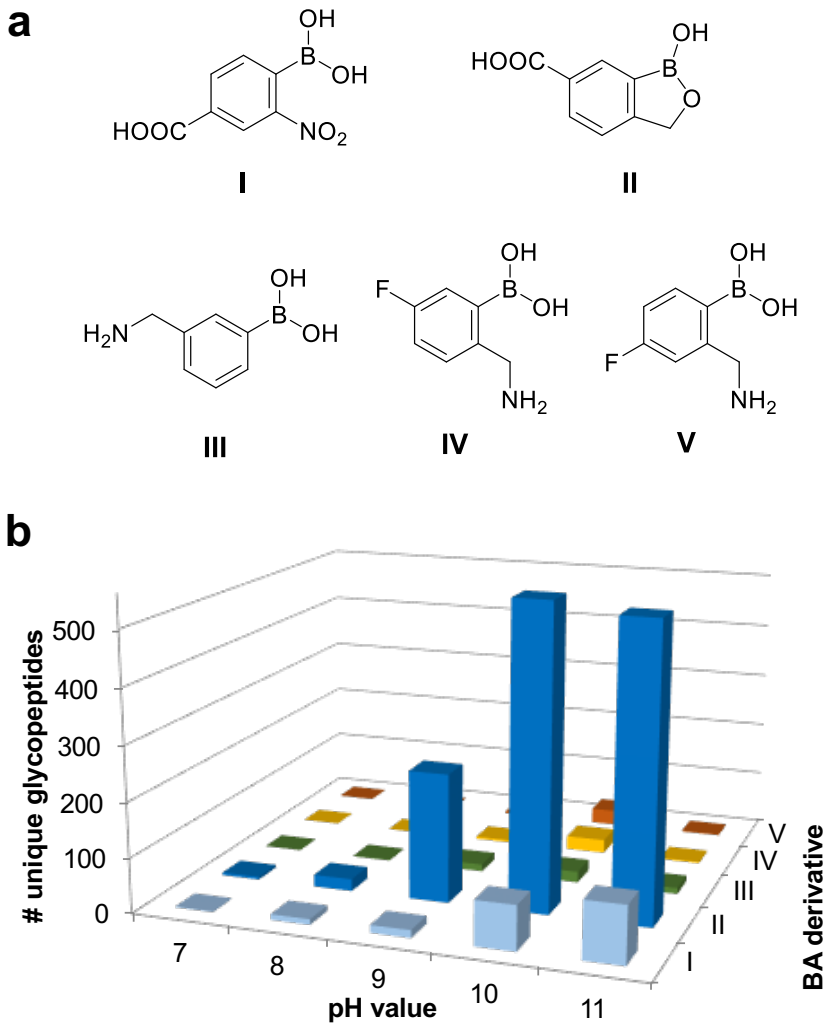


Figure 2.3 Structures of boronic acid derivatives and experimental results using different derivatives. Structures of boronic acid derivatives tested in this work (a), and the number of glycopeptides identified with each BA derivative at varying pH values from the parallel experiments (b).

In parallel experiments starting with the same amount of purified peptides from human cells (HEK 293T), we examined these BA derivatives at different pH values and compared the number of unique identified N-glycopeptides. Very few glycopeptides were identified at pH=7 or 8 with any BA derivative. For all derivatives, the optimal pH was 10 or 11, as shown in Figure 2.3b. The derivatives **IV** and **V** enriched slightly more unique glycopeptides compared to phenylboronic acid (**III**). Although the pK_a of derivative **I** (9.2) is similar to that of

phenylboronic acid (9.0), enrichment with **I** resulted in the identification of more unique glycopeptides. One possible reason is that the adjacent nitrogen may form an extra hydrogen bond with a nearby hydroxyl group on the glycan, which enhances the interactions between the BA derivative and glycans and facilitates the enrichment.

Among these five boronic acids tested, derivative **II** (benzoboroxole) allowed the identification of the greatest number of glycopeptides. The interactions between benzoboroxole and sugars were reported to be stronger than those between phenylboronic acid and sugars[59, 64, 65]. For example, the binding constant (K_a) for the reaction between benzoboroxole and fructose is 606 M^{-1} at neutral pH, which is nearly ten times higher than that between phenylboronic acid and fructose (79 M^{-1}) under identical conditions[64]. The current experimental results are very consistent with previous findings, and stronger interactions between BA and glycans can more effectively enrich glycopeptides. For the first time, the method based on benzoboroxole was systematically optimized for site-specific and global analysis of glycoproteins in combination with MS, and the results were dramatically improved compared to any other boronic acids tested here.

2.3.2 Synergistic interactions to increase glycopeptide coverage

Strengthening the interactions between benzoboroxole and glycopeptides will further increase the coverage of low-abundance glycopeptides. One glycan typically contains multiple monosaccharides, which allows one glycopeptide to interact with multiple benzoboroxole molecules. The synergistic effect for the interactions between multiple BA derivative molecules and glycans is expected to further facilitate the enrichment of glycopeptides, especially those with low abundance. Here we synthesized a dendrimer as the platform for synergistic interactions because the number of benzoboroxole molecules bound to a dendrimer

can be easily adjusted. More importantly, the dendrimer branches also provide structural flexibility to enhance the synergistic interactions.

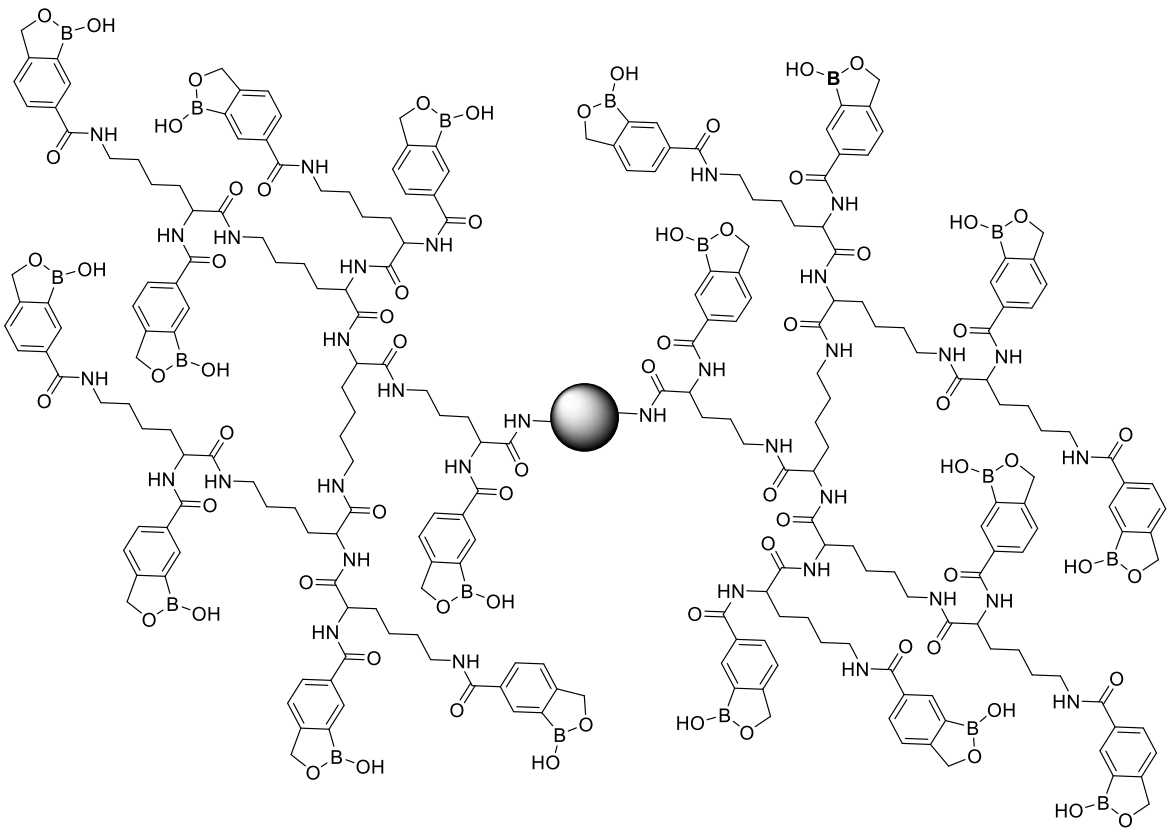


Figure 2.4 The structure of BA derivative II (benzoboroxole) conjugated dendrimer.

The dendrimer was first synthesized and bound to magnetic beads, and next the BA derivative, benzoboroxole, was conjugated to the dendrimer (Figure 2.1 and 2.2). Many benzoboroxole molecules were bound to one dendrimer, as shown in Figure 2.4, and the number of benzoboroxole molecules on one dendrimer bead is proportional to the dendrimer size. In this case, several sugars from one glycan may interact with multiple benzoboroxole molecules simultaneously (Figure 2.5).

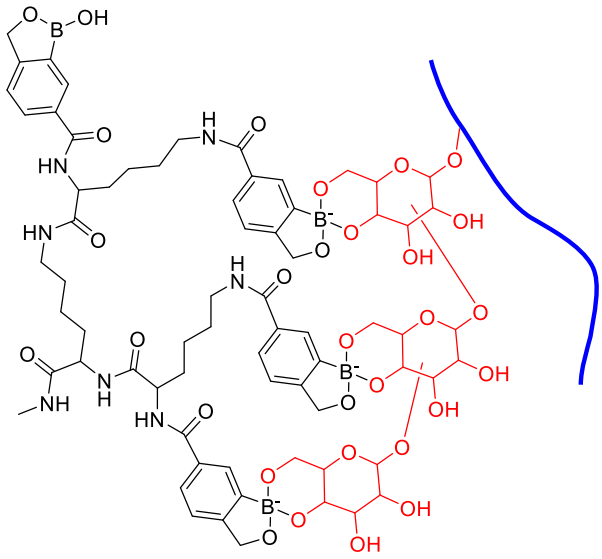


Figure 2.5 An example of the synergistic interactions between multiple benzoboroxole molecules in a dendrimer and several sugars within one glycan of a glycopeptide.

Dendrimer size is expected to have a large impact on the synergistic interactions, and the effect of dendrimer size was systematically evaluated in parallel experiments, where the number of benzoboroxole molecules on the beads attempted to remain the same, and the amount of starting materials (peptides from HEK 293T cells) was also the same. In Figure 2.6, when the number of cycles is zero, the magnetic beads are directly conjugated with benzoboroxole without a dendrimer. The dendrimer size increases with the number of rounds of synthesis, as well as the number of benzoboroxole molecules after conjugation. With dendrimer beads synthesized through one to four rounds of the reaction, the number of total N-glycopeptides, unique N-glycopeptides, and N-glycoproteins increased linearly (Figure 2.6). After four rounds of synthesis, the numbers are very comparable, and the specificity results have a similar trend (Figure 2.7). Once the number of benzoboroxole molecules on a single bead reaches the threshold, larger dendrimers with more benzoboroxole molecules do not affect the synergistic interactions, which occurs after four rounds of synthesis.

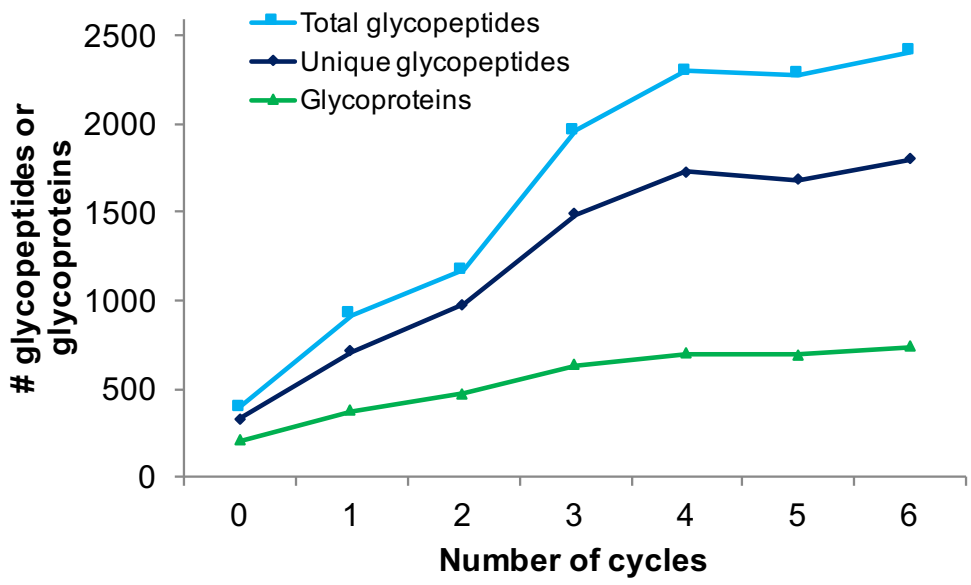


Figure 2.6 The effect of number of synthesis cycles and corresponding dendrimer size on the enrichment of glycopeptides.

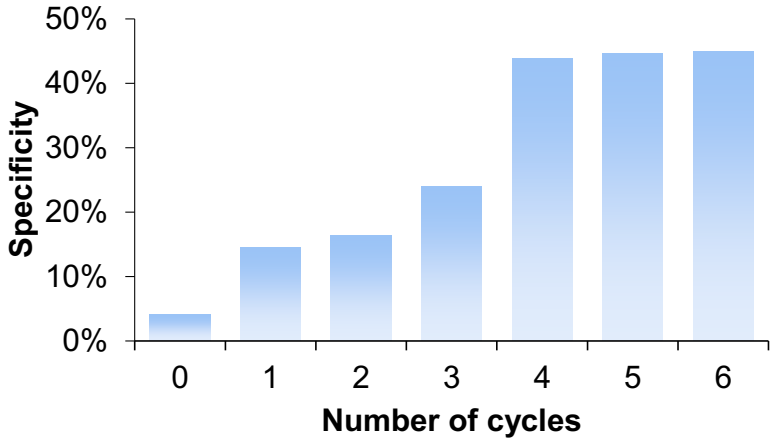


Figure 2.7 Specificity of the N-glycopeptide identifications increases with the number of the dendrimer synthesis cycles, and it levels off after the fourth cycle. The overall specificity of glycopeptide enrichment should be higher considering that O-glycopeptides were also enriched.

Since the enrichment reaction is quick and the conditions are mild, prolonging the reaction time does not have negative effects on glycopeptide identification. As shown in Figure 2.8, similar number of unique N-glycopeptides and glycoproteins were identified when the

incubation time varied between 10 minutes to 3 hours. We also assessed the residual N-glycans after PNGase F treatment, and duplicate experiments were performed to examine the percentage of residual N-glycans. The results demonstrated that the N-glycan removal efficiency with PNGase F within three hours was very high (Figure 2.9). Briefly, peptides from MCF7 whole cell protein digestion were subject to enrichment with the DBA beads. The enriched glycopeptides were then treated with PNGase F in $H_2{}^{18}O$ for three hours. The purified peptides were analyzed using an online LC-MS/MS system with a Q-Exactive Plus mass spectrometer, and both full MS and MS/MS were recorded in the Orbitrap cell. Higher-energy collisional dissociation (HCD) was used as the fragmentation method. We searched for the deglycosylated peptides (2.9883 Da mass shift on N) and the N-glycan-containing peptides using Byonic. As a result, 44 unique glycan-containing peptides and 1,866 deglycosylated peptides were identified in the first experiment; 45 unique glycan-containing peptides and 1,871 unique deglycosylated peptides were identified in the second experiment. Overall, N-glycopeptides with residual N-glycans are only around ~2%, demonstrating that the 3-hour PNGase F treatment was effective to remove N-glycans.

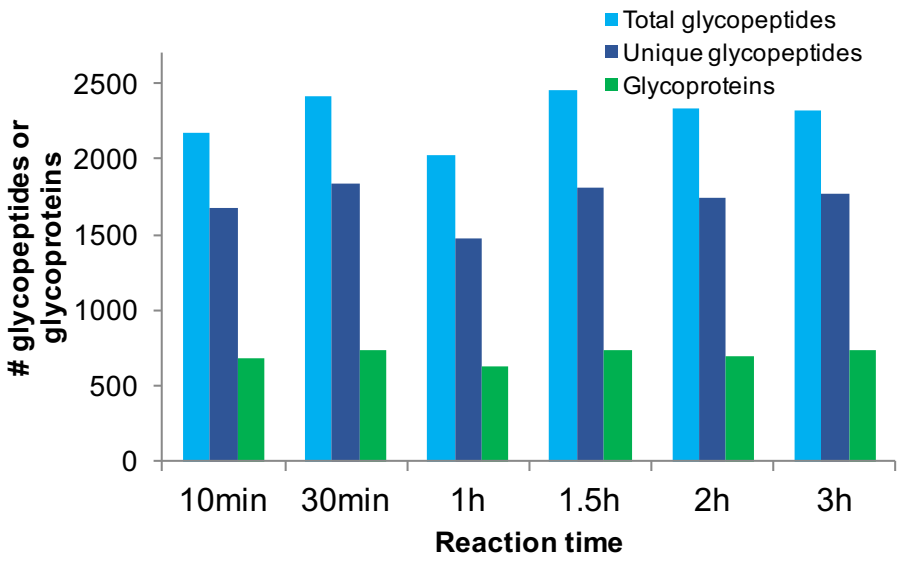


Figure 2.8 The effect of reaction time on the N-glycopeptide identification.

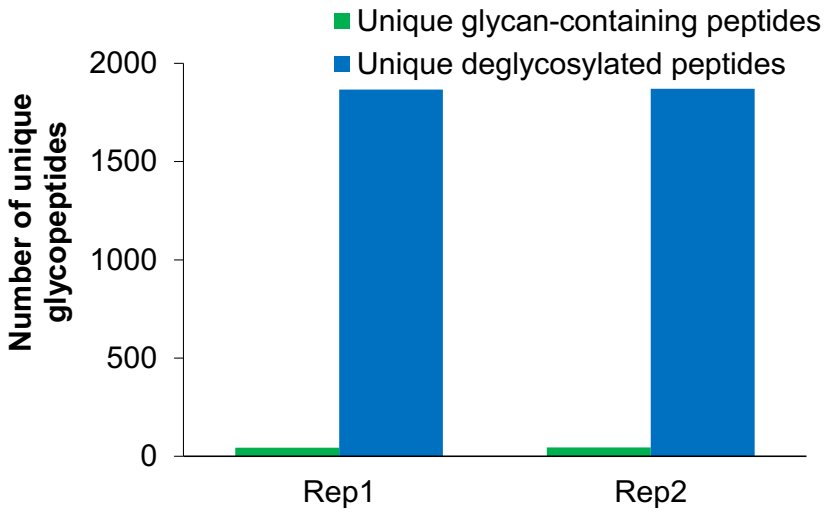


Figure 2.9 Duplicate experimental results for assessing residual N-glycans after PNGase F treatment. Only about 2% N-glycopeptides contained residual glycans after the three-hour treatment.

We performed the control experiments previously[66] and found that the effect of spontaneous deamidation is negligible under the treatment conditions (pH=7.5 and 37 ^{o}C) for three hours. For all our experiments for protein N-glycosylation analysis, we strictly controlled the treatment time within three hours. Although a longer treatment time may lead to more complete removal of N-glycans and result in the identification of more N-glycosylation sites, spontaneous deamidation will cause higher false positive rates for protein N-glycosylation site identification.

2.3.3 Further optimization of experimental conditions for DBA enrichment

Effect of different solvents on glycopeptide enrichment. A variety of solvent combinations were tested for glycopeptide enrichment with DBA beads (Figure 2.10). The pH of all aqueous solutions was adjusted to 11 using an ammonium acetate buffer. For each combination, the

binding step of enrichment was performed for an hour in the first solution, and then the beads were washed five times in the second solution. The combination of "DMSO+DMSO" provided the highest enrichment efficiency with the identification of the most N-glycopeptides and glycoproteins. This is consistent with *Le Chatelier*'s principle because water is the product of the reaction between the boronic acid derivative and sugars. Without water, the reaction shifts toward the direction of bond formation and becomes more complete.

Washing buffer for glycopeptide enrichment optimization. Based on the results from Figure 2.10, several washing buffers were tested, and the results are in Figure 2.11. We performed the enrichment in DMSO containing 0.5% trimethylamine (TEA) for one hour, and then washed the beads with different buffer combinations. The enriched peptides were subsequently deglycosylated and analyzed by LC-MS/MS. The washing buffer containing 50% DMSO and 50% H_2O (pH=11) outperformed all other combinations. The addition of water helped remove non-specifically bound peptides and increased the number of identified glycopeptides and glycoproteins.

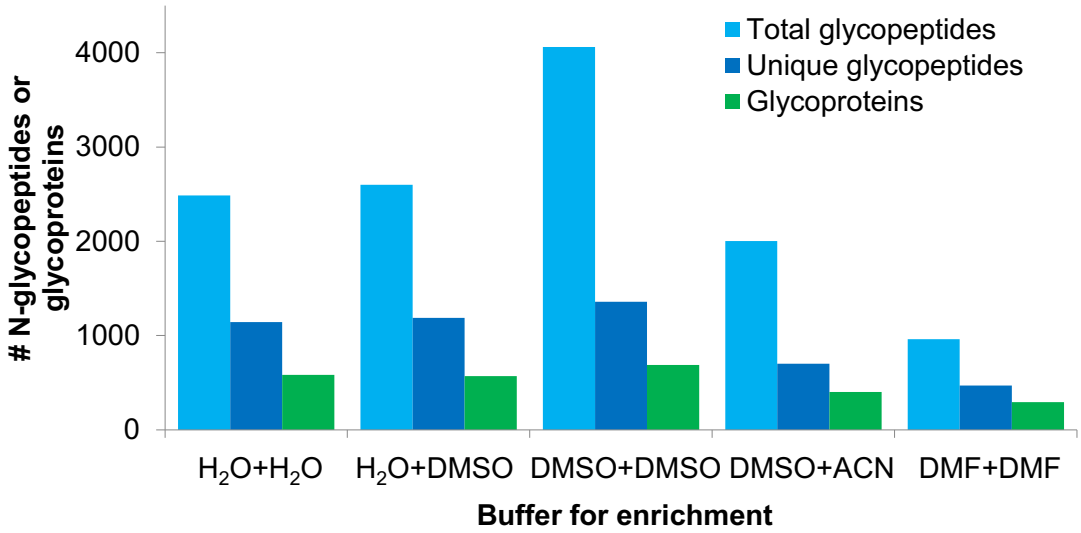


Figure 2.10 Effect of solvents on glycopeptide enrichment from a human cell lysate (HEK 293T).

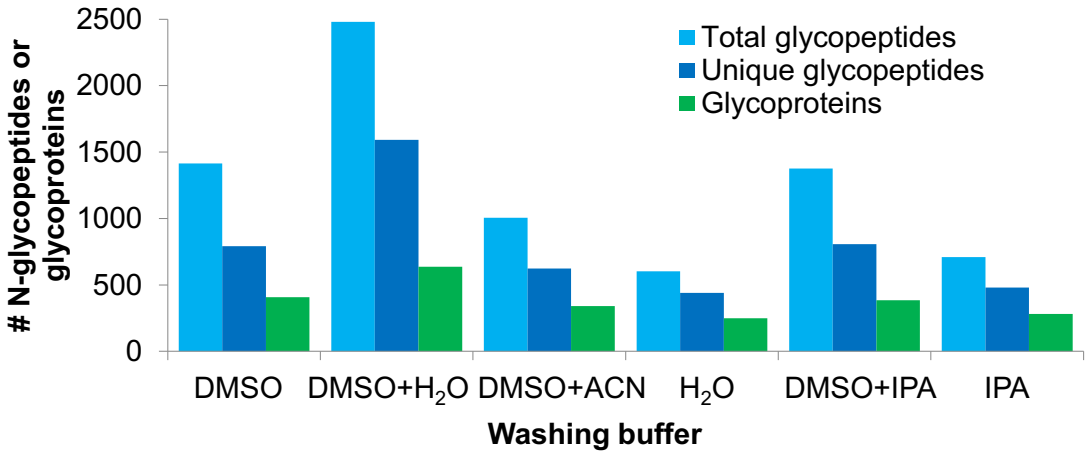


Figure 2.11 Washing buffer optimization for glycopeptide enrichment.

Number of washes for glycopeptide enrichment optimization. Based on the previous results, we further optimized the number of washes (0-9 washes). All parallel experiments started with about 0.25 mg mammalian peptides, which were enriched with the DBA beads in DMSO containing 0.5% TEA for one hour, and then the number of times the beads were washed with 50% DMSO and 50% H_2O (pH=11) was varied. From 0 to 4 washes, a linear trend was found for N-glycopeptide and glycoprotein identifications because increasing the number of washes removed non-specifically bound peptides. After washing four times, there was no obvious change (Figure 2.12). These results indicate that the interactions between DBA and glycans are very strong because washing more times did not result in the loss of glycopeptides.

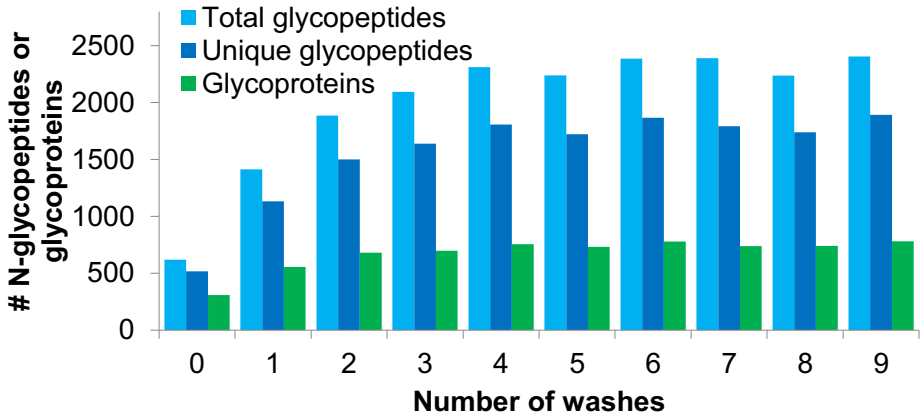


Figure 2.12 The effect of washing times on glycopeptide enrichment.

Effect of sample size on the identification of glycopeptides and glycoproteins. Different amount of cultured MCF 7 cells were used to evaluate the effect of sample size on the N-glycopeptide identification with the DBA enrichment. Duplicate experiments were performed. Cells in each group were harvested and the final protein amounts in the eight groups were around 10, 30, 60, 100, 200, 300, 500, and 1000 μg, respectively. After protein precipitation and digestion, the peptides were subject to DBA enrichment. The enriched glycopeptides were then purified and analyzed by LC-MS/MS. The data is presented in Figure 2.13.

The lowest number of glycoproteins we identified in one MS run was about 200 from the 10 μg group among the samples tested here. When the sample amount is very small, the sample loss coming from every step may be a problem. For instance, a very small volume of solvent (lysis buffer or digestion buffer) was used to transfer the sample from tube to tube, which could result in a considerable (relatively higher percentage) sample loss. However, even for 10 μg proteins, we were still able to identify about 200 glycoproteins. More samples allowed us to identify higher numbers of unique glycopeptides and glycoproteins. After the protein amount reached ~300 μg, the increasing trend of the number of identified glycopeptides and glycoproteins slowed down, and both the 500 μg and 1000 μg groups yielded almost the same results. Besides the sample loss, the MS sensitivity is also a major contribution factor for the results that fewer glycoproteins were identified using a low sample amount. A machine with higher sensitivity would allow us to identify more glycoproteins using the same amount of material or the same number of glycoproteins using a lower amount of material. Of note, normally the protein digestion efficiency and peptide purification efficiency are lower than 100%, and therefore, the resulting peptide amounts subjected to the DBA enrichment in the current experiment should be slightly lower than the sample amounts shown in the figure.

2.3.4 Comparison DBA with existing lectin- and HILIC-based methods

To test the effectiveness of the Dendrimer-conjugated Boronic Acid derivative (DBA) enrichment, triplicate parallel experiments were performed to compare the current method with the commonly used lectin (combining WGA and ConA) and zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) enrichment methods. Each experiment started from the same amount of peptides from MCF7 cell whole lysates (Figure 2.14). For these parallel experiments, except the enrichment method, every other step was kept the same. Prior to this comparison, we compared 0.1% and 1% TFA as ion-pairing reagent for the ZIC-HILIC experiment and found that 1% TFA had slightly better performance (Figure 2.14b).. Therefore, we used 1% TFA in the comparison experiment. From the parallel experiments, the greatest number of unique N-glycopeptides were identified using the current DBA method, and more unique N-glycopeptides were identified with ZIC-HILIC than the lectin-based method.

Regarding the specificity, we compared the numbers of glycopeptides and non-glycopeptides identified in each of the parallel experiments, and the results showed that the DBA method had the highest specificity (Figure 2.14c). We reasoned that although ZIC-HILIC allows for enrichment of a broader spectrum of glycopeptides than lectin, the principle of the ZIC-HILIC method is based on the hydrophilic property difference between glycopeptides and non-glycopeptides. Therefore, some hydrophilic but non-glycosylated peptides can also be enriched, lowering the enrichment specificity. Based on the number of unique glycopeptides identified, DBA outperformed the other two methods, while ZIC-HILIC had better performance than lectin. Furthermore, the current method also has the highest enrichment specificity.

2.3.5 Global characterization of protein N-glycosylation in yeast

Using the newly developed method, we performed biological duplicate experiments for the global analysis of protein N- and O-glycosylation in yeast. For N-glycoprotein analysis, we identified 881 sites on 400 proteins in one experiment and 836 sites on 404 proteins in the other. Overall, 1,044 N-glycosylation sites (Figure 2.15a) on 501 proteins (Figure 2.15b) were identified. For the first time, over 1,000 protein N-glycosylation sites were identified in yeast. To ensure that the sites were confidently identified, very stringent criteria were applied during analysis. First, the false positive rate at the N-glycopeptide level was well-controlled under 1.0%, based on the target-decoy method[54]. Additionally, all N-glycosylation sites were required to contain the motif NX[S/T/C], where X is any amino acid except proline. The N-glycosylation site was also required to contain heavy oxygen (^{18}O) as a tag[45]. To minimize possible spontaneous deamidation during PNGase F treatment in heavy-oxygen water, the reaction was run for only three hours. Our previous results demonstrated that within three hours under mild conditions, spontaneous asparagine deamidation is negligible[66].

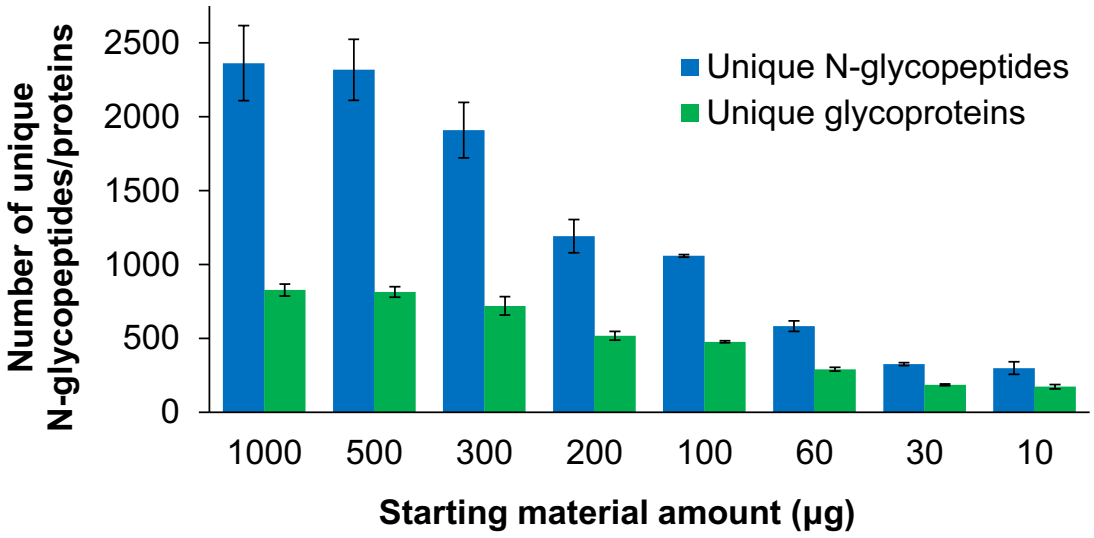


Figure 2.13 Evaluation of the effect of sample size on the identification of glycopeptides and glycoproteins with the DBA enrichment followed by LC-MS analysis.

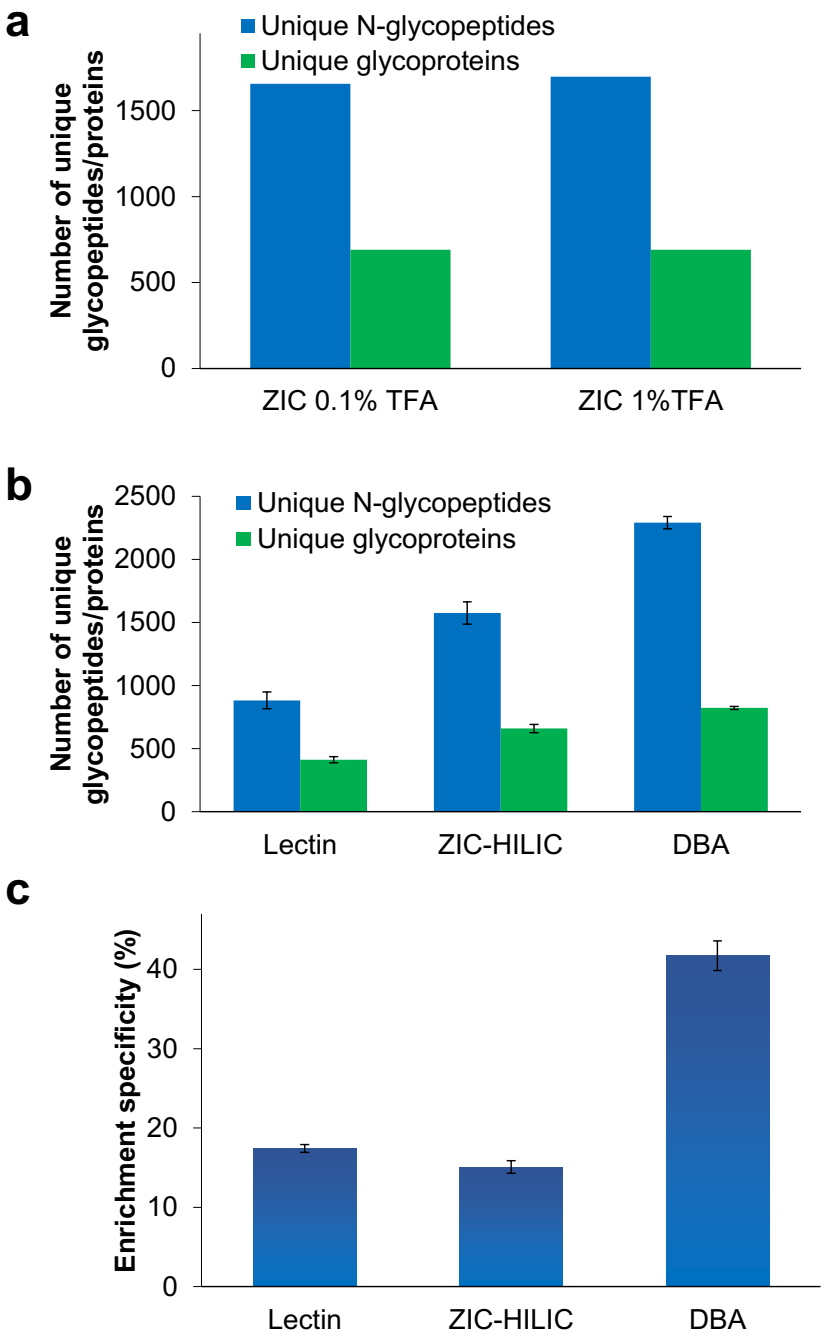


Figure 2.14 Comparison of three enrichment methods (Lectin, ZIC-HILIC and DBA). (a) Optimization of the concentrations of TFA as the ion-pairing reagent for ZIC-HILIC enrichment. (b) The numbers of unique glycopeptides and glycoproteins identified using each of the three methods from parallel experiments. (c) Comparison of the enrichment specificity for the three enrichment methods.

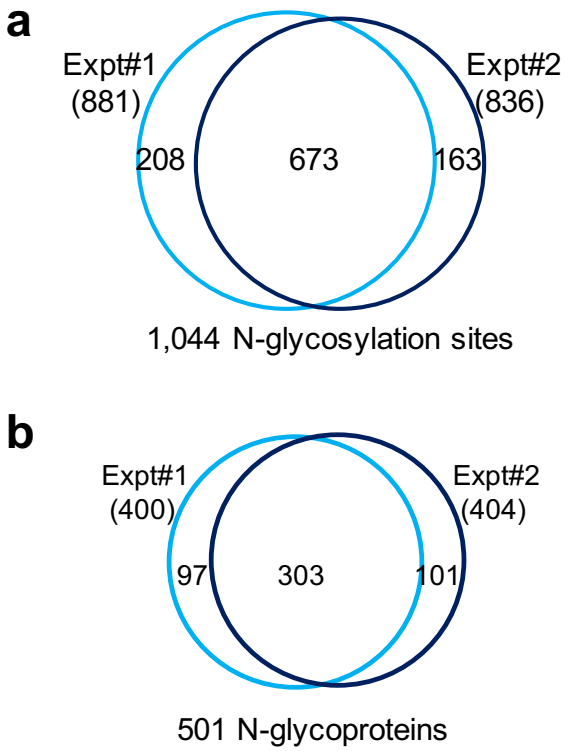


Figure 2.15 (a) N-glycopeptides and (b) N-glycoproteins identified from the yeast duplicate experiments

In order to demonstrate that low-abundance glycoproteins can be identified with the current method, we compared the abundance distributions of identified N-glycoproteins and all proteins in the whole yeast proteome, and they were very similar (Figure 2.16a). We reanalyzed our previous dataset using phenylboronic acid magnetic beads in yeast[52] with the same criteria as above, and 716 N-glycosylation sites on 297 proteins were identified. The abundance distributions for both datasets are shown in Figure 2.16b. More N-glycoproteins were identified in each bin with the current method, especially for low-abundance N-glycoproteins (abundance from the literature[67]). For example, for proteins with abundances less than 2,000 copies per cell, about twice as many N-glycoproteins were identified in this work (158 *vs.* 84), which clearly

demonstrated that the current method is more effective in enriching low-abundance glycopeptides due to strengthened interactions from the BA derivative and synergistic interactions of DBA.

2.3.6 Analyzing protein O-mannosylation in yeast

The reversible covalent interactions can leave enriched glycopeptides with intact glycans for site identification and glycan structure elucidation. In yeast, O-glycans consist of only mannose, but the number of mannose per glycan varies. The current enrichment method also enables us to globally analyze O-glycoproteins. In order to increase the identification confidence of intact O-glycopeptides, high-energy collisional dissociation (HCD) was employed for glycopeptide fragmentation, and the tandem mass spectra were recorded in the Orbitrap cell. Several important machine parameters, such as (automatic gain control) AGC target for MS and MS^2, normalized collision energy, and maximum ion accumulation time for MS^2, were optimized (Figure 2.17). We used Byonic^{TM} to search the raw files for the identification of protein O-mannosylation.

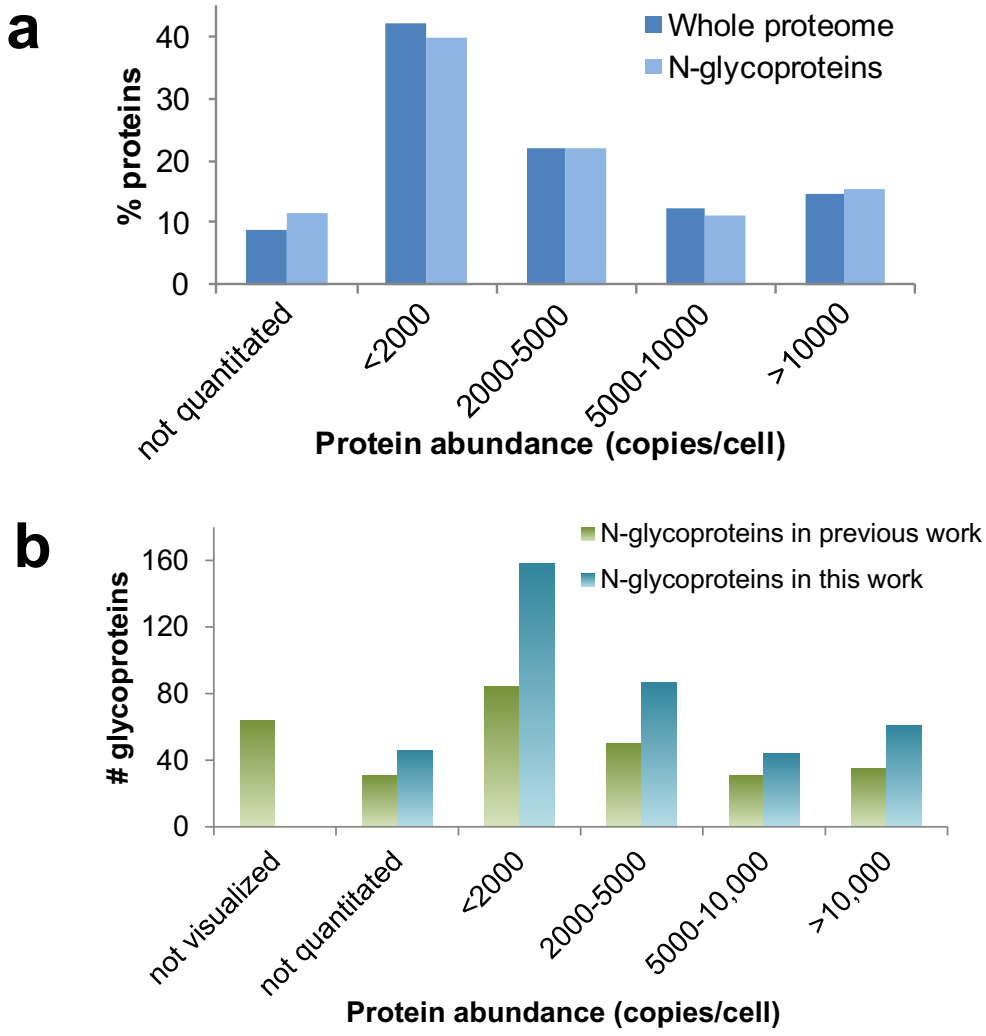


Figure 2.16 (a) Abundance distributions of the whole proteome and N-glycoproteins identified here. (b) Comparison of the abundance distributions of yeast N-glycoproteins identified in this work and identified previously with the phenylboronic acid beads in 2014[52].

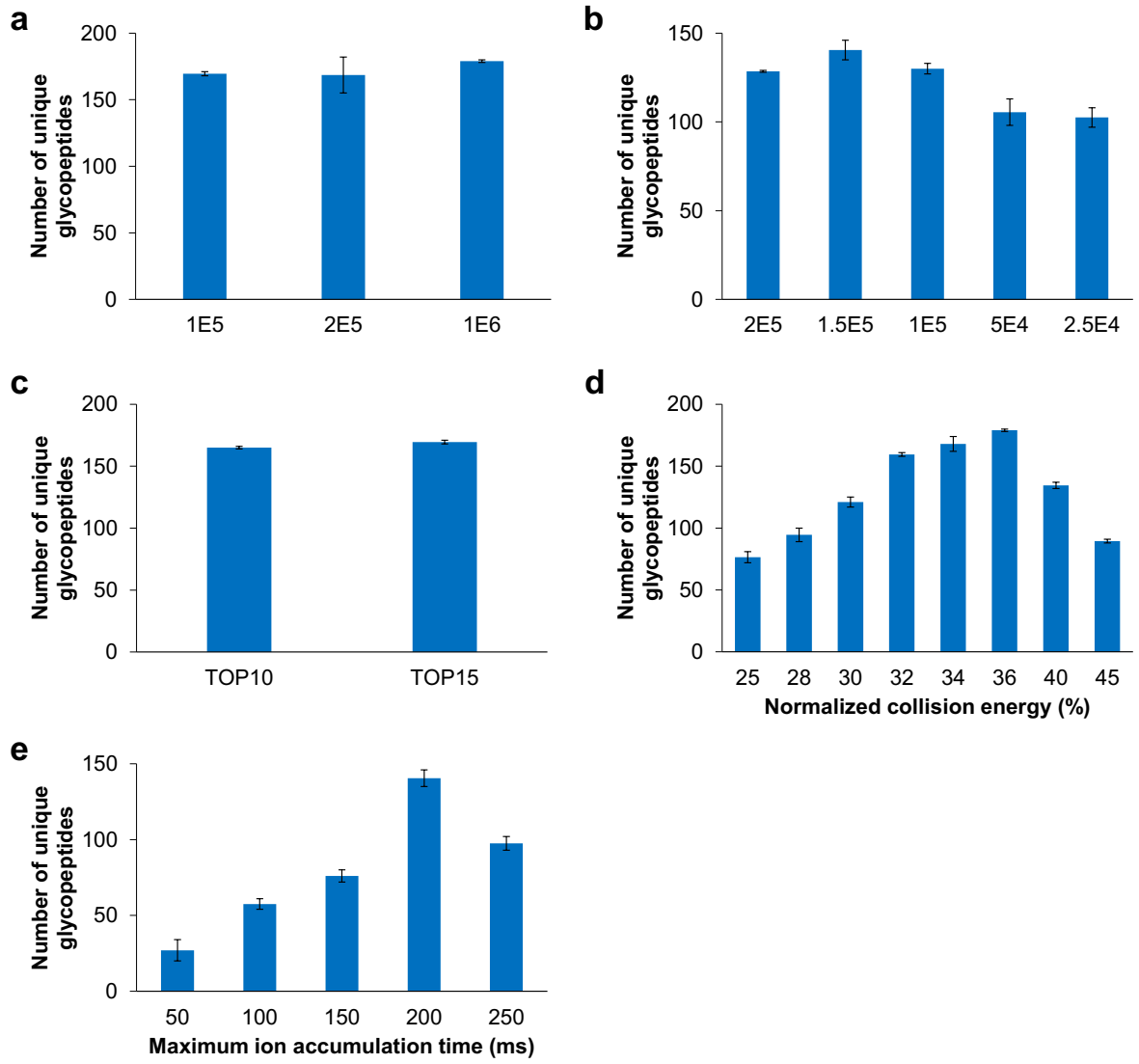
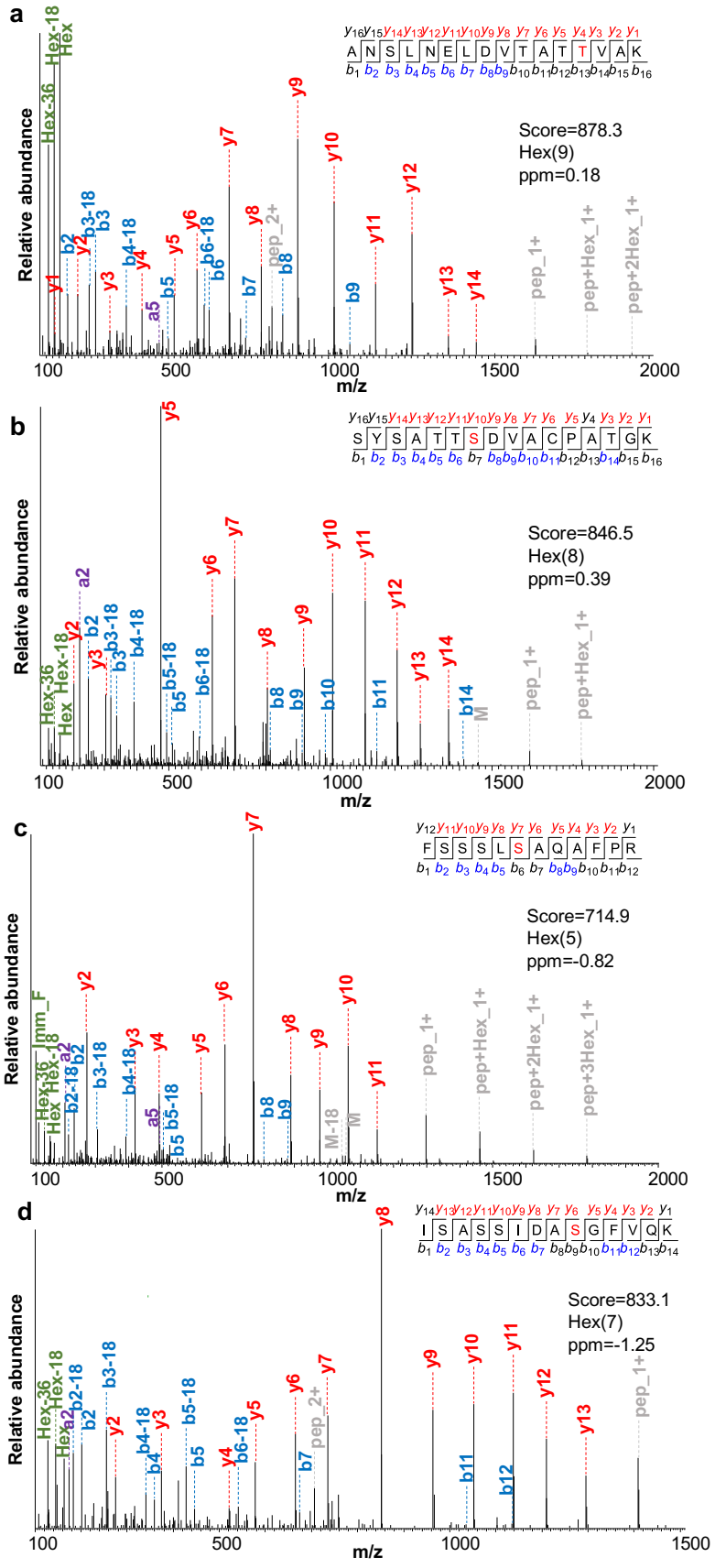


Figure 2.17 Machine parameters were optimized for yeast intact O-glycopeptide analysis using the Orbitrap cell to record tandem mass spectra of glycopeptides. (a) AGC target for full MS, (b) AGC target for MS^2, (c) comparison of Top10 and Top15 methods, (d) normalized collision energy, (e) maximum ion accumulation time for MS^2.


a
Relative abundance
m/z
Score=878.3
Hex(9)
ppm=0.18
b
Score=846.5
Hex(8)
ppm=0.39
c
Score=714.9
Hex(5)
ppm=-0.82
d
Score=833.1
Hex(7)
ppm=-1.25

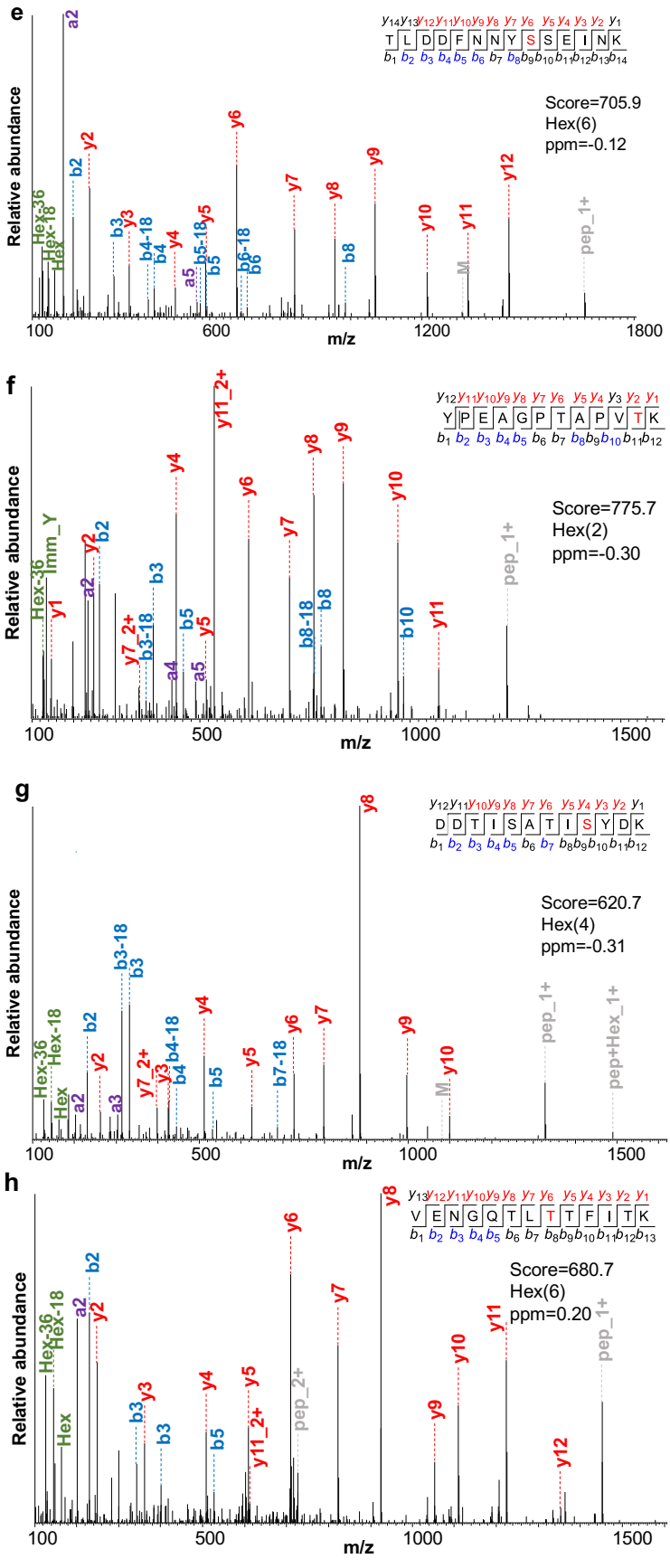

e
Relative abundance
m/z
Score=705.9
Hex(6)
ppm=-0.12
f
Relative abundance
m/z
Score=775.7
Hex(2)
ppm=-0.30
g
Relative abundance
m/z
Score=620.7
Hex(4)
ppm=-0.31
h
Relative abundance
m/z
Score=680.7
Hex(6)
ppm=0.20

Figure 2.18 Examples of O-mannosylated peptides identified in this work. (a) Glycopeptide ANSLNELDVTATT[Hex_9]VAK from protein GAS3. (b) Glycopeptide SYSAT[Hex_8]TSDVACPATGK from protein GAS1. (c) Glycopeptide FSSSLS [Hex_5]AQAFPR from protein EXG2. (d) Glycopeptide ISASSIDAS[Hex_7]GFVQK from protein SED4. (e) Glycopeptide TLDDFNNYS[Hex_6]SEINK from protein GAS1. (f) Glycopeptide YPEAGPTAPVT[Hex_2]K from protein YD056. (g) Glycopeptide K.DDTIS [Hex_4]ATISYDK from protein GAS3. (h) Glycopeptide R.VENGQTLT[Hex_6]TFITK from protein PRY2.

Several examples of the O-mannosylated peptides with different glycans identified here are displayed in Figure 2.18. Here, we identified 987 unique O-glycopeptides from 206 proteins in the first experiment and 971 unique O-glycopeptides from 196 proteins in the second experiments. In total, 234 O-glycoproteins were identified, and 168 proteins were identified in both experiments. The overlap was very high (81.5 and 85.7%), which further demonstrated that the identification of glycopeptides and glycoproteins were highly confident. The current results are proof-of-concept to demonstrate that the glycopeptide enrichment based on the reversible covalent interactions can keep enriched glycopeptides intact for site identification and glycan structure elucidation.

The distribution of the number of mannose per glycan is in Figure 2.19. The number of unique glycopeptides with one mannose is the highest, and the second are those with four mannoses. For glycopeptides with glycans containing more than four mannoses, the number decreases with the increasing number of mannoses. The site localization confidence is lower than that of N-glycosylation due to the neutral loss of O-glycans and the presence of many serine and threonine residues on O-glycopeptides (Figure 2.20). Compared to the whole yeast proteome, both S and T were more frequent in the identified unique O-glycopeptides, and the occurrence of T was almost two times as many (9.0 vs. 11.8% for S and 5.9 vs. 10.7% for T).

Conversely, the frequency of N (N-glycosylation sites) in the identified O-glycopeptides was lower than the whole proteome (6.1 vs. 4.5%).

In total, 234 O-glycoproteins were identified, and about one third were also N-glycosylated (Figure 2.21). O-glycoproteins located on the cell wall (*P*=4.25E-32) are the most highly enriched when clustered using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[68] (Figure 2.22). Seventy-three O-glycoproteins belong to the endomembrane system, and 55 are located in the ER. Clustering of O-glycoproteins based on molecular function indicates that proteins related to hydrolase activity (acting on glycosyl bonds) and transferase activity (transferring glycosyl groups) are most highly enriched (Figure 2.23). Based on reversible covalent interactions between DBA and glycans, protein O-glycosylation can be confidently identified, providing valuable glycan structural information.

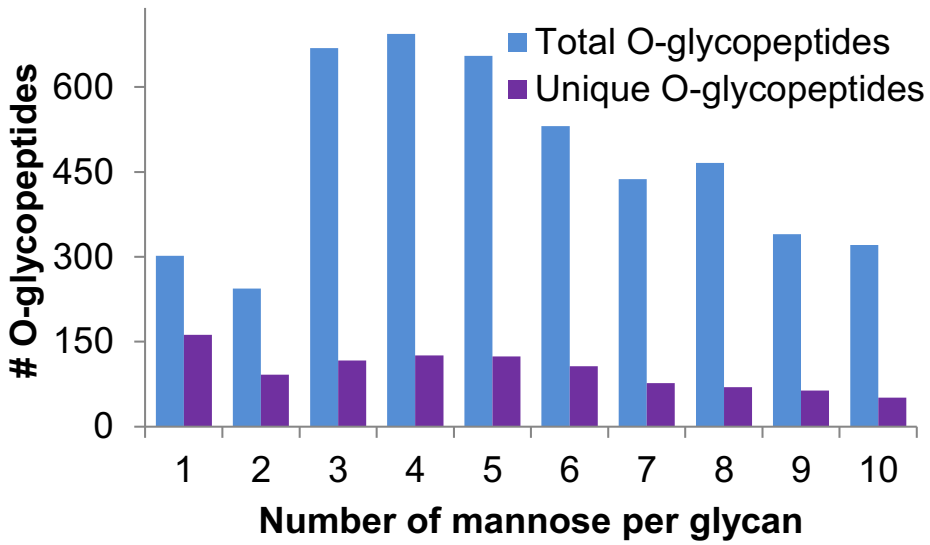


Figure 2.19 Distribution of the number of mannose residues per glycan on all identified O-glycopeptides.

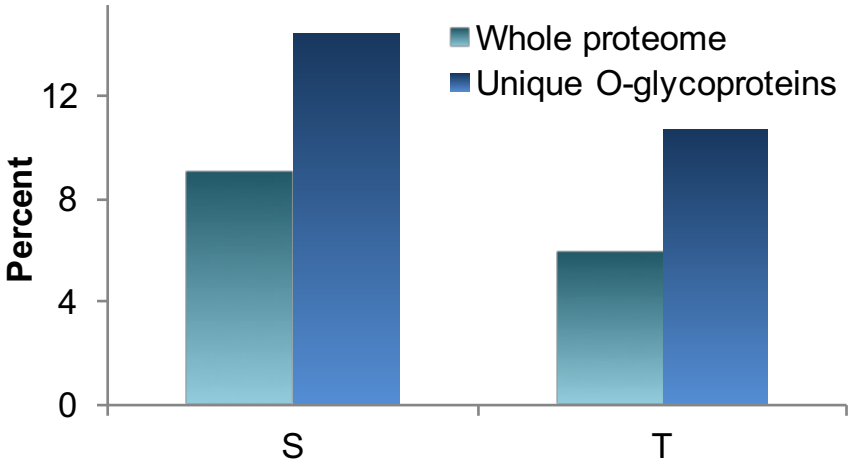


Figure 2.20 Percentages of S, T and N in O-glycopeptides compared to the whole proteome.

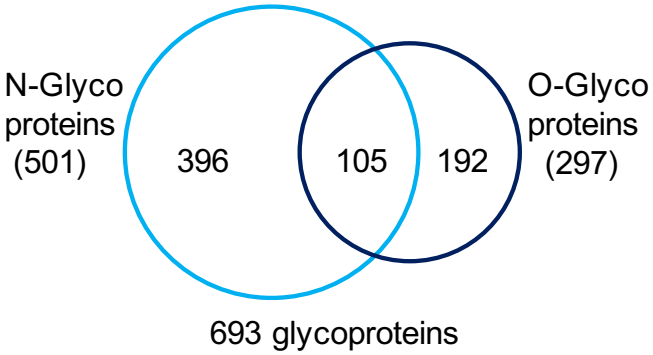


Figure 2.21 Comparison of O- and N-glycoproteins identified in yeast cells.

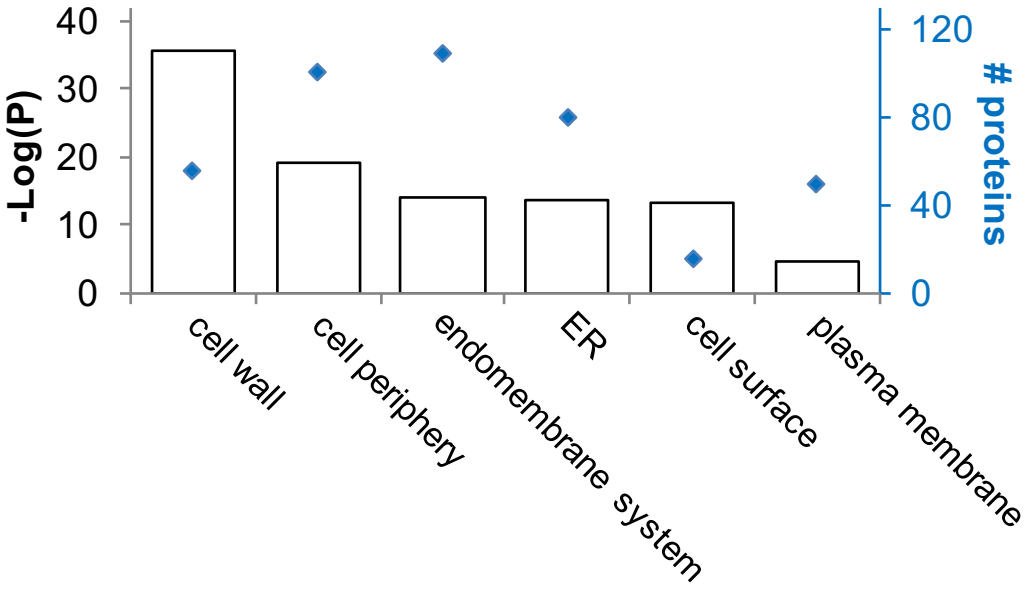


Figure 2.22 Clustering of O-glycoproteins based on cellular compartment.

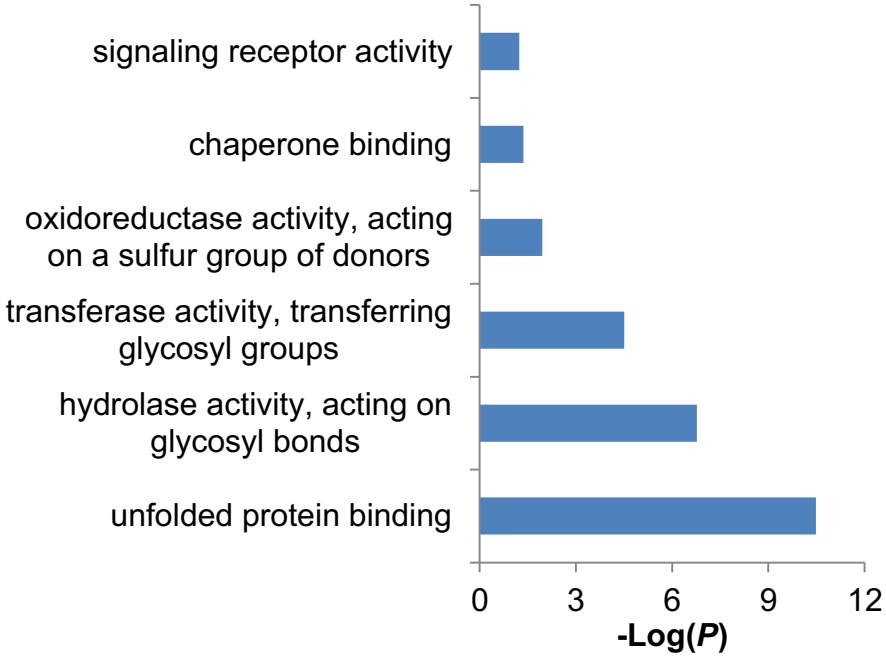


Figure 2.23 Clustering of identified O-glycoproteins in yeast based on molecular function.

2.3.7 Global analysis of protein N-glycosylation in human cells

Due to the diversity of glycan structures, it is more challenging to globally analyze glycoproteins in human cells. The new DBA method was applied to globally analyze protein N-glycosylation in different types of human cells. Biological duplicate experiments were performed for MCF7 cells, and the number of glycosylation sites and glycoproteins identified in each experiment is shown in Figure 2.24. With the well-controlled FDR of <1.0% at the glycopeptide level and stringent criteria described above, we identified 2,710 N-glycosylation sites on 1,127 proteins in one experiment and 2,815 sites on 1,156 proteins in the other. Overall, 2,340 common sites were identified in both experiments, which represent 86.3% and 83.1% of the total sites identified from each experiment, respectively. As expected, the overlap at the glycoprotein level was even higher: 981 common glycoproteins were identified. A total of 3,185 glycosylation sites were identified on 1,302 proteins in MCF7 cells.

The method was also employed to globally analyze protein N-glycosylation in HEK 293T and Jurkat cells; 3,052 sites were identified on 1,301 proteins in HEK 293T cells, and 2,120 sites on 948 proteins were found in Jurkat cells. The comparison of identified sites is in Figure 2.25.

We further tested the effect of the dendrimer on glycopeptide enrichment by comparing DBA vs. benzoboroxole conjugated magnetic beads without the dendrimer (designated as BA). With the DBA beads, we were able to identify 88% more N-glycosylation sites and 79% more glycoproteins compared to the benzoboroxole conjugated magnetic beads (BA) (Figure 2.26a). The abundance distributions of all glycoproteins identified using either the DBA or BA beads are displayed in Figure 2.26b (abundances from an online database (PaxDb)[69]). Besides the number of glycoproteins identified using the DBA beads was higher than that with the BA beads in each abundance category, the DBA method was especially superior for glycoproteins with very low abundance (less than 10 ppm). For low-abundance proteins, over twice as many

N-glycoproteins were identified with the DBA beads (84 *vs.* 34 glycoproteins for <0.1 ppm, and 402 *vs.* 196 for 0.1-1.0 ppm). These results explicitly demonstrate that the synergistic interactions between multiple BA derivative molecules and glycans can greatly increase the coverage of low-abundance glycopeptides.

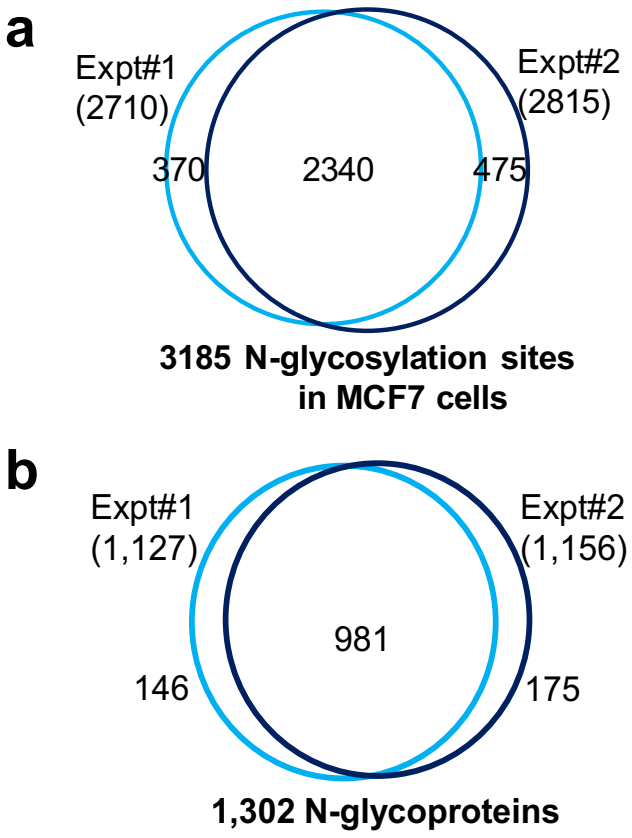


Figure 2.24 (a) The N-glycosylation sites and (b) the glycoproteins identified from the MCF-7 cell duplicate experiments.

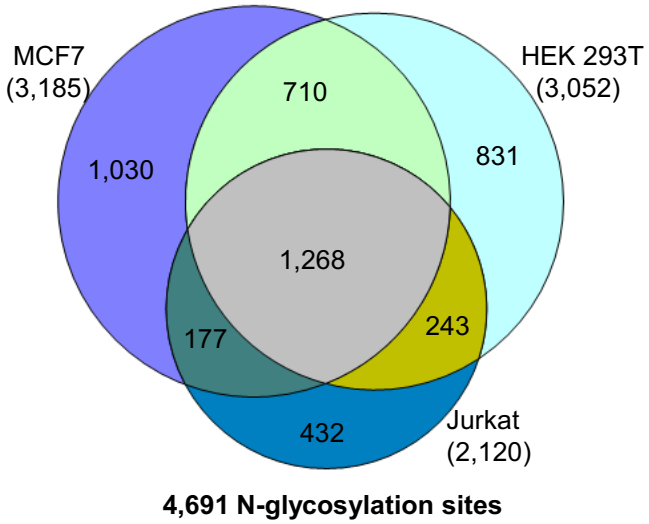


Figure 2.25 Comparison of N-glycosylation sites identified in MCF7, HEK 293T and Jurkat cells.

Combining the results from the three human cell lines, we identified a total of 4,691 N-glycosylation sites on 1,906 proteins (Figure 2.28a). More than 10% of proteins (238) are highly glycosylated and contain at least five sites (Figure 2.27). In consideration of different cell types, there is a decent overlap among identified N-glycoproteins in human cell experiments (Figure 2.28a). One example highlighting differences between cell types are the N-glycoproteins (180) identified only in Jurkat cells, many of which are related to immune cell-specific activities, such as cell activation and cell immune response (Figure 2.27b).

By clustering 1,906 N-glycoproteins according to molecular function, proteins related to glycosyl-transferase activity are the most highly enriched with a *P* value of 8.5E-35 (Figure 2.29a), and 108 N-glycoproteins belong to this category. In yeast, this group of proteins is the second most highly enriched. The following groups of N-glycoproteins are also highly enriched in human cells: receptor binding, signaling receptor activity, growth factor binding proteins, glycosaminoglycan binding, cell adhesion molecule binding, and active transmembrane transporter activity.

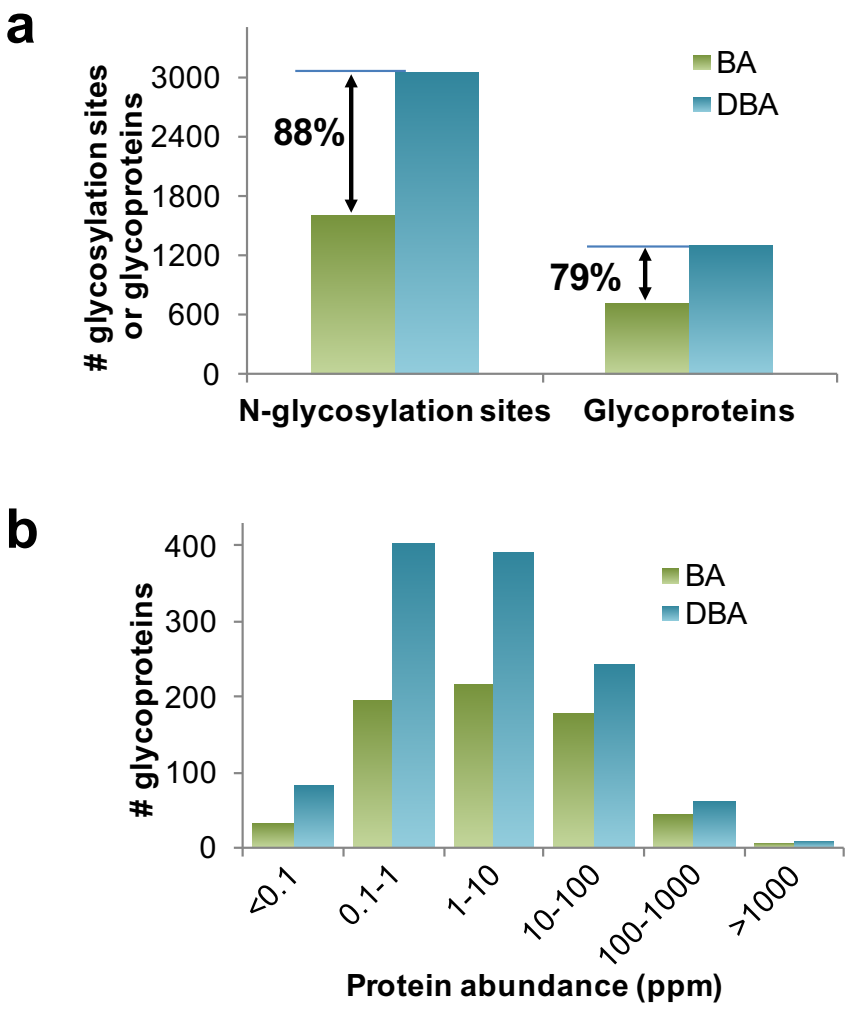


Figure 2.26 (a) Comparison of unique glycosylation sites and glycoproteins identified with the boronic acid derivative magnetic beads (designated as BA) and with the dendrimer beads conjugated with the boronic acid derivative (DBA). (b) Abundance distributions of N-glycoproteins identified with the BA or DBA beads.

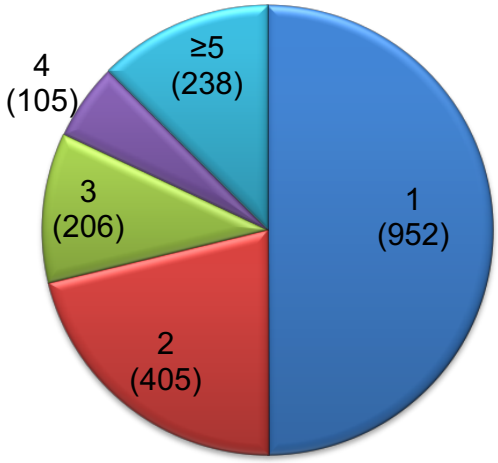


Figure 2.27 The distribution of unique N-glycosylation sites per glycoprotein in human cells.

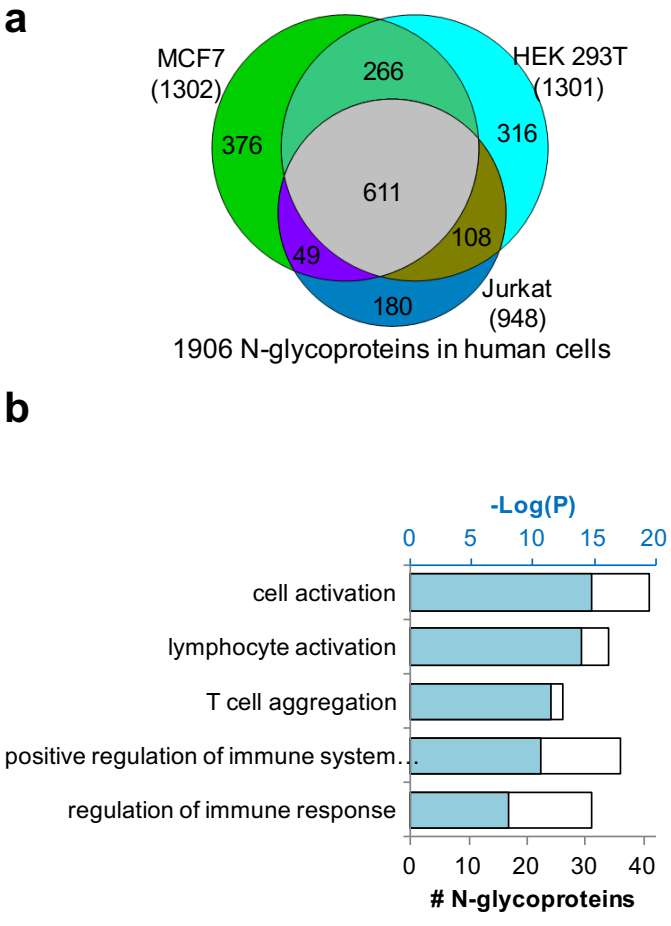


Figure 2.28 (a) Overlap of N-glycoproteins in three different types of cells (MCF7, HEK 293T and Jurkat), and (b) Protein clustering results for 180 N-glycoproteins identified exclusively in Jurkat cells

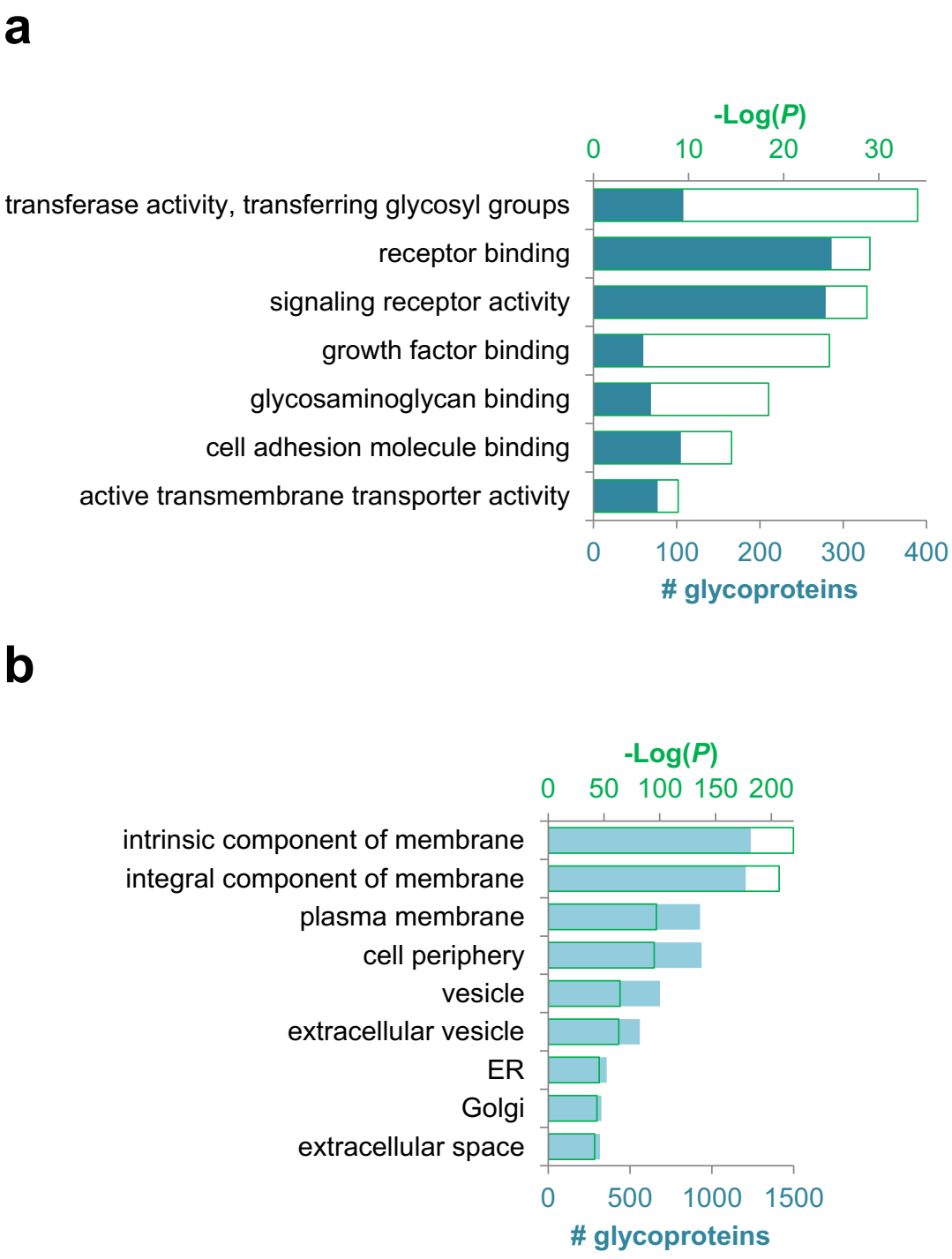


Figure 2.29 Clustering of N-glycoproteins based on (a) molecular function and (b) cellular compartment.

Many glycoproteins are known to be membrane proteins. Here, 1,251 out of 1,906 N-glycoproteins are membrane proteins, which are highly enriched with an extremely low *P* value of 1.6E-192. Glycoproteins in the cell periphery, vesicle, ER, Golgi, and extracellular space are all enriched with very low *P* values (Figure 2.29b). Based on the information available on UniProt (uniprot.org), 524 of identified membrane proteins are type I membrane proteins, 177 are type II, and 348 proteins contain multiple transmembrane domains (Figure 2.30). A total of 301 receptors were identified among these N-glycoproteins (Figure 2.31a); glycosylation site locations for receptors identified as type I and II membrane proteins are shown in Figure 2.31b. All sites (1,079) were located in the extracellular space, which corresponds very well with the belief that glycans are located on the extracellular side of surface membrane proteins.

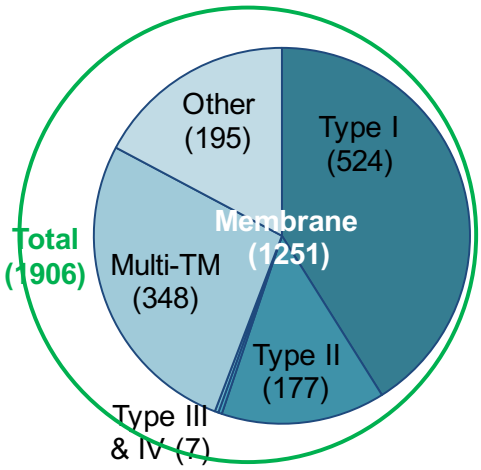


Figure 2.30 Distribution of membrane proteins (Type I, II, III & IV, and multi-pass transmembrane (TM)) among all identified N-glycoproteins.

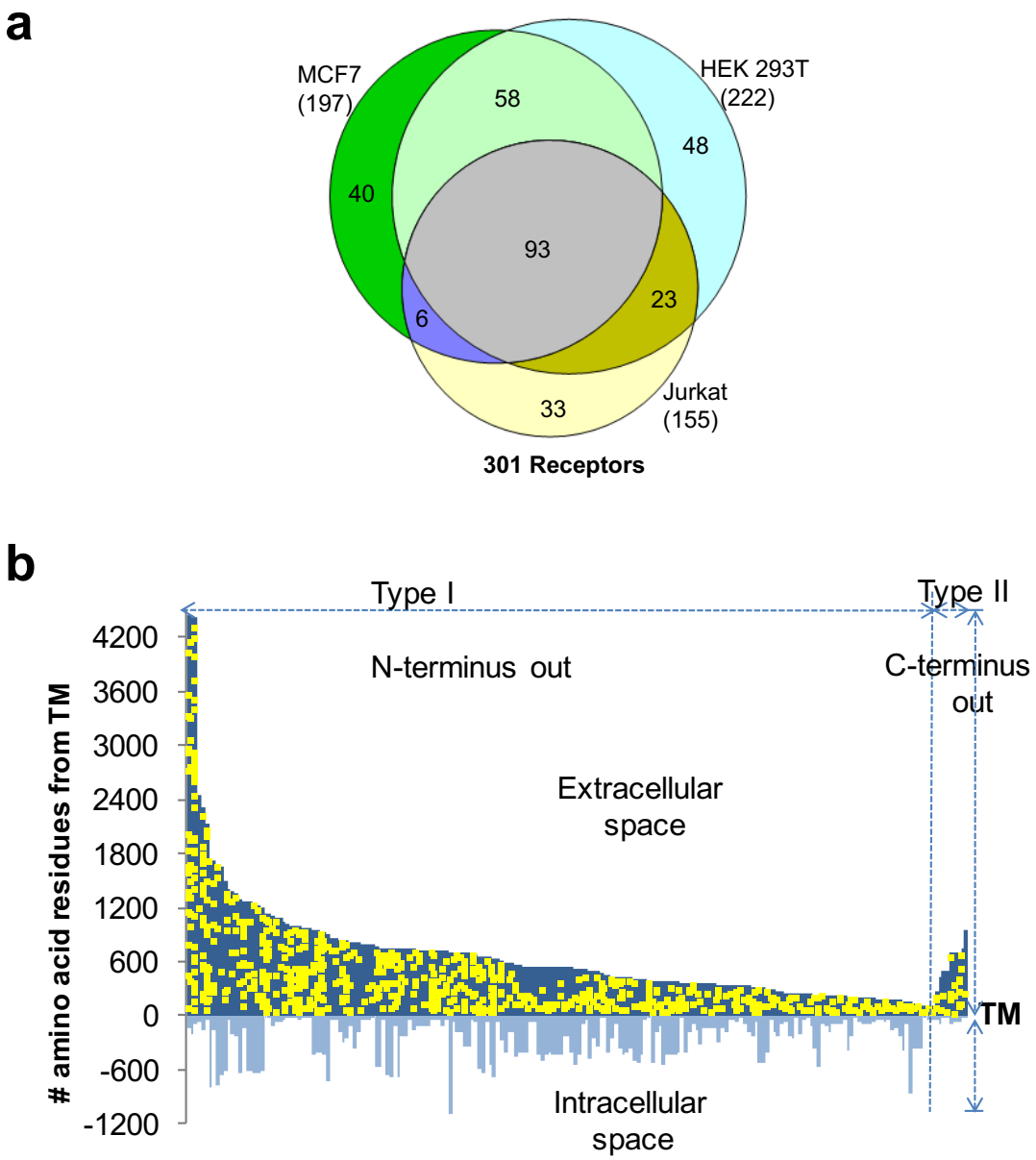


Figure 2.31 (a) The number of receptors (N-glycoproteins) identified in each type of human cells, and (b) N-glycosylation site locations on 301 receptors with X-axis as the TM domain. Each glycoprotein sequence was aligned against the transmembrane (TM) domain, and the glycosylation sites are indicated as yellow dots. All sites are located in the extracellular space.

Domain analysis shows that many N-glycoproteins contain different types of Ig domains (such as I-set, V-set and C1-set). Besides Ig domains, other domains related to cell-cell adhesion are also highly enriched, including fibronectin type III, cadherin and laminin G

(Figure 2.32). Domains corresponding with receptor activities, such as PTK (protein tyrosine kinase) and EGF (epidermal growth factor)-like domains, are also highly enriched.

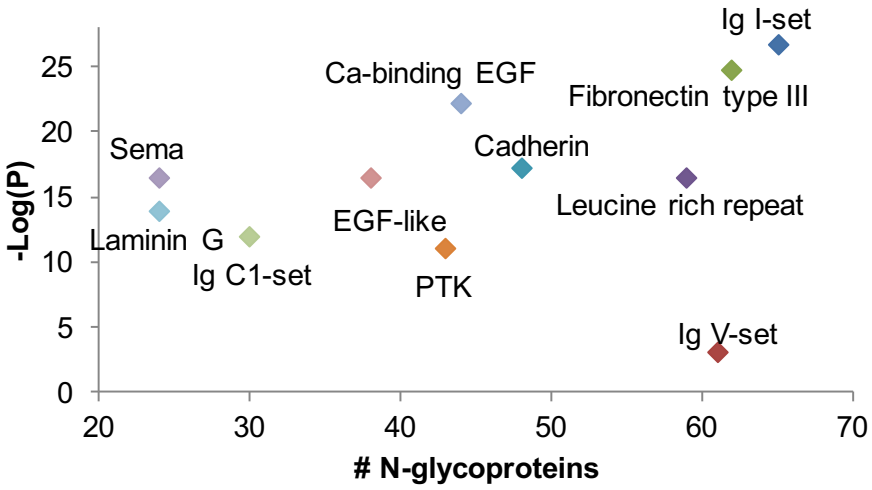


Figure 2.32 Domain analysis of N-glycoproteins showing the number of N-glycoproteins containing the most highly-enriched domains and their corresponding P values.

2.3.8 Analysis of protein N-glycosylation in mouse brain tissues

Without sample restriction, the current method can be applied for glycoprotein analysis in any other samples, including animal tissues and clinical samples. Here, we further applied this method to analyze protein N-glycosylation in mouse brain tissues, and biological duplicate experiments were performed. After protein extraction and digestion, glycopeptides were enriched using the DBA beads, and enriched glycopeptides were fractionated, followed by analysis with an online LC-MS system.

In the first experiment, we identified 3,583 sites on 1,434 glycoproteins, and very similar results were obtained in the second experiment (3,685 sites on 1,443 proteins). In total, 4,195 sites were identified on 1,608 proteins, and 3,073 common sites and 1,269 glycoproteins were found in both experiments, as shown in Figure 2.33. Considering the large-scale analysis and the experiments being biologically duplicate, the overlap is very high at both the site (85.8

and 83.4% compared to the both experimental results, respectively) and protein (88.5 and 88.0%) levels, which is consistent with the above results from duplicate experiments using human cells. The highly reproducible results further demonstrate that the current method is effective.

Glycoproteins identified in the mouse brain tissues were clustered using DAVID based on biological process. About one quarter of identified glycoproteins (396) are related to cell surface receptor signaling pathway, which is the most highly enriched with a *P* value of 1.1E-61. Proteins related to brain-specific functions such as nervous system development (*P*=4.1E-61), axon development (*P*=1.9E-54), and synapse assembly (*P*=2.6E-30) were also highly enriched, as shown in Figure 2.34.

2.3.9 Synergistic interactions to identify protein O-GlcNAcylation

Protein O-GlcNAcylation was discovered more than three decades ago, and it has been reported to be involved in many cellular events, from regulating cell signaling to gene expression[70-72]. Using BA-based methods, it is challenging to enrich O-GlcNAcylated proteins because only one sugar (GlcNAc) is bound to S or T, and this sugar does not contain a *cis*-1,2-diol. Although boronic acid can interact with sugars without *cis*-1,2-diols, such as glucose and GlcNAc, the interaction is weak[58], and enrichment is therefore less effective.

In this work, we identified 510 total glycopeptides with HexNAc(1) and 304 unique glycopeptides located on 131 proteins in HEK 293T cells with the DBA enrichment Figure 2.35. In striking contrast, with the BA derivative magnetic beads, only 18 total glycopeptides with HexNAc and 13 unique glycopeptides were found on 12 proteins. Among 131 glycoproteins, 81 were located in the nucleus (Figure 2.36), and typically, these proteins are O-GlcNAcylated because only glycoproteins with O-GlcNAc have been reported in the nucleus.

Similarly, 131 O-glycoproteins with HexNAc(1) were identified in MCF7 cells, and 119 O-glycoproteins were found in Jurkat cells.

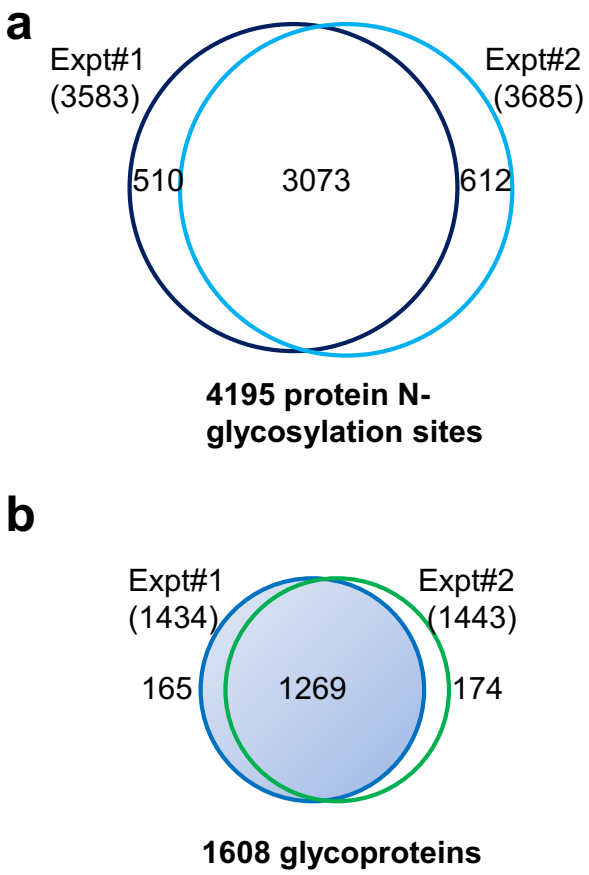


Figure 2.33 The number of protein N-glycosylation sites (a) and glycoproteins (b) identified in mouse brain tissues from biological duplicate experiments.

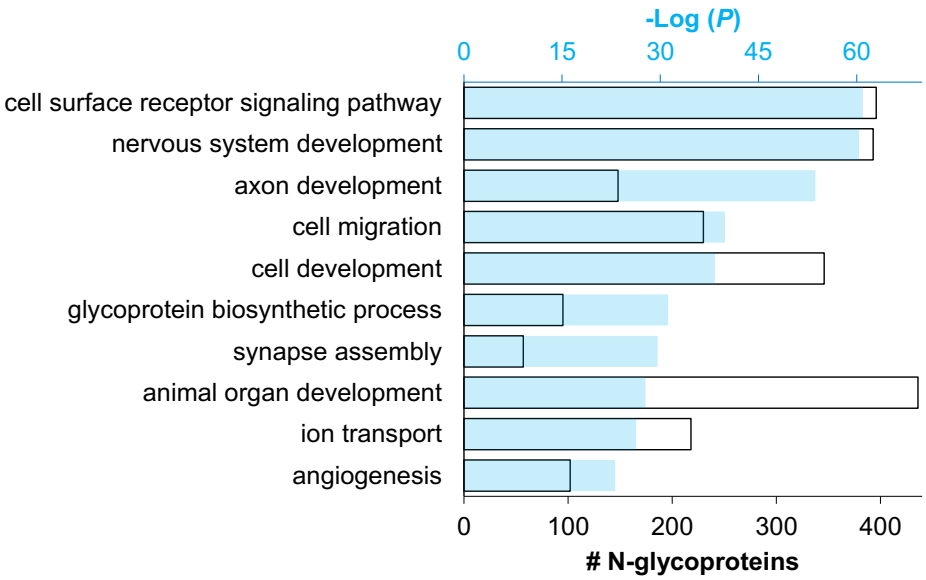


Figure 2.34 Clustering of glycoproteins identified in mouse brain tissues based on biological process.

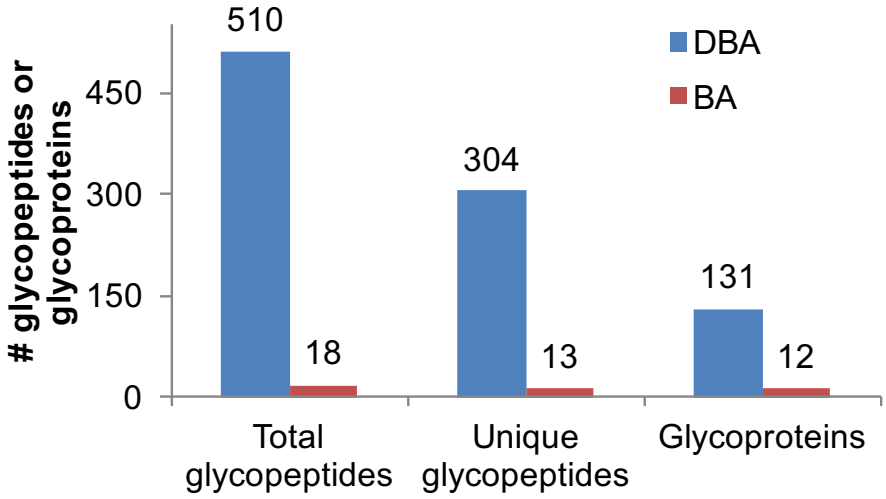


Figure 2.35 Comparison of glycoproteins with one HexNAc identified with BA and DBA, which clearly shows that the results from DBA are substantially better.

The effective enrichment of O-GlcNAcylated peptides may be attributed to the synergistic interactions with DBA beads. As discussed above, multiple sugars from one glycan

synergistically interact with different benzoboroxle molecules on a single dendrimer bead. Although there is no *cis*-1,2-diol in GlcNAc, multiple hydroxyl groups in each GlcNAc may form reversible covalent bonds with several benzoboroxle molecules on a dendrimer bead, as shown in Figure 2.37. The synergistic interactions can dramatically facilitate the enrichment of O-GlcNAcylated peptides with DBA. The results are highly reproducible in different types of human cells (HEK 293T, MCF7 and Jurkat). The greatest number of identified glycoproteins (about 50%) are located in the nucleus of each cell type Figure 2.38 and about 30% of them are in the cytoplasm. Glycoproteins in the nucleus and the cytoplasm are normally O-GlcNAcylated. In addition, ~12% of them are in the ER/Golgi. Only a small portion of glycoproteins (~7%) are secreted proteins, which are likely to be O-GalNAcylated.

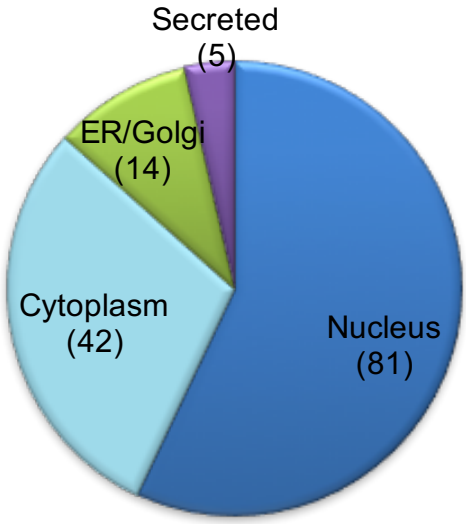


Figure 2.36 Distribution of O-glycoproteins modified with HexNAc(1) identified in HEK 293T cells based on cellular compartment.

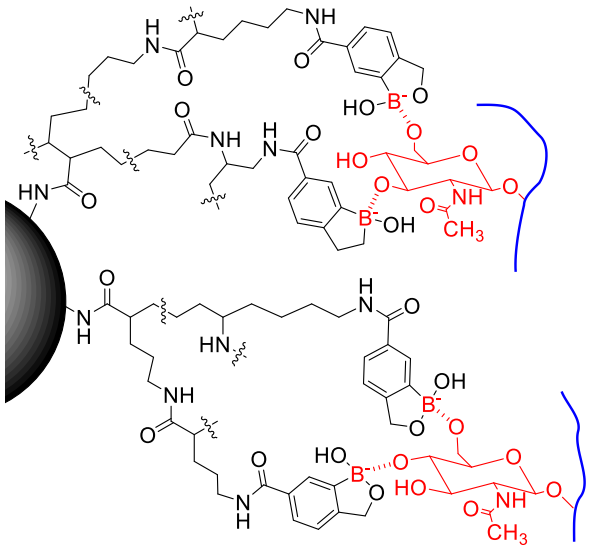


Figure 2.37 Proposed mechanism of the interactions between DBA and GlcNAc benefiting from synergistic interactions.

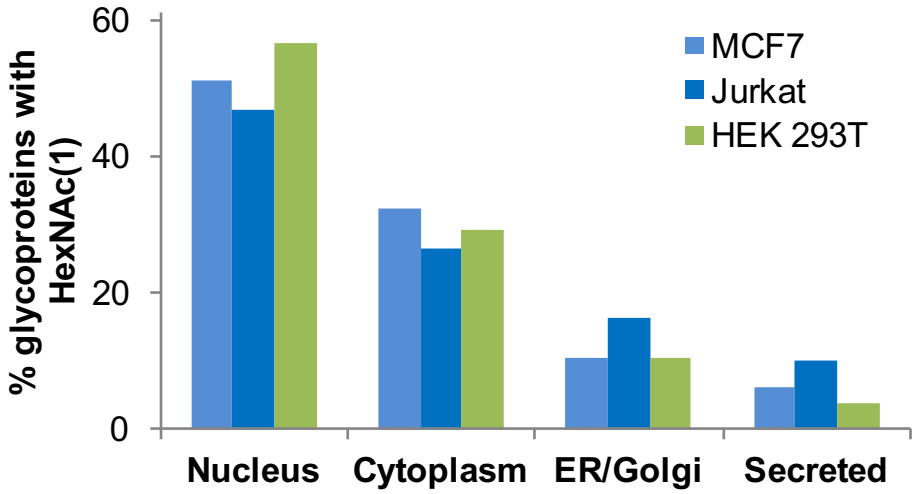


Figure 2.38 Cellular compartment distribution of glycoproteins containing one HexNAc identified in the three types of cells.

2.4 Discussion

Based on universal and reversible covalent interactions between boronic acid and sugars, BA-based enrichment methods have great potential in enriching glycopeptides for global analysis of protein glycosylation. However, the relatively weak interactions prevent the enrichment of glycopeptides with low abundance. In order to effectively enrich glycopeptides in complex biological samples, it is critical to strengthen the interactions between BA and glycans.

In this work, we enhanced the interactions between BA and glycans through two ways. First, we employed the BA derivative (benzoboroxole) to form stronger interactions with glycans, which was able to dramatically increase the coverage of low-abundance glycopeptides, as shown in Figure 2.3b and Figure 2.17. Second, based on the common features of a glycan containing multiple monosaccharides and one sugar bearing several hydroxyl groups, we benefited from synergistic interactions by conjugating many benzoboroxole molecules onto a dendrimer bead. The synergistic interactions between several benzoboroxole molecules on a bead and different sugars within a glycan make the enrichment much more effective, which is clearly demonstrated from the current results (Figure 2.26). The dendrimer provides an excellent platform to conjugate many benzoboroxole molecules onto the same bead. The dendrimer size is readily adjustable, and correspondingly, the number of benzoboroxole molecules can be controlled on each bead. Furthermore, the dendrimer provides structural flexibility to form stronger interactions with glycans.

The reversible nature of the interactions between BA and glycans allows enriched peptides to be released with intact glycans. The direct analysis of intact glycopeptides provides valuable information about protein glycosylation sites and glycan structures. We systematically analyzed O-mannosylated proteins and their glycan structures in yeast, and overall, 234 O-glycoproteins were identified. With stringent criteria for analysis, the identifications of O-

glycopeptides and O-glycoproteins are highly confident. However, compared to protein N-glycosylation site identification, O-glycosylation sites were less confidently localized because of the possible neutral loss of glycans during intact O-glycopeptipe analysis and high percentages of S and T in glycopeptides.

Synergistic interactions can enhance not only the interactions between benzoboroxole and glycans containing multiple monosaccharides but also the interactions with O-GlcNAcylated peptides. It is well-known that BA can form stronger interactions with sugars containing *cis*-1,2-diols because two covalent bonds are formed. The interaction between BA with glucose or GlcNAc without *cis*-1,2-diols is much weaker[58]. Here, due to the flexible nature of the dendrimer, one GlcNAc may form multiple covalent bonds with different benzoboroxle molecules, as shown in Figure 2.37. Compared to BA beads, DBA is much more effective in enriching O-GlcNAcylated peptides (Figure 2.35).

Cluster of differentiation (CD) molecules are those located on the cell surface that provide immunophenotyping targets for cell classification[73]. Two examples of glycoproteins identified in Jurkat cells are shown in (Figure 2.39), and the majority of identified N-glycosylation sites are located in extracellular domains. In our experiment, 188 CD proteins were identified as N-glycoproteins. There were more CDs identified in Jurkat cells (137) than MCF7 (115) or HEK 293T (129) cells (Figure 2.40), despite the fact that the total N-glycoproteins identified in Jurkat cells were fewer. However, this result is consistent with the fact that more CDs are relevant to immune-related cells, including Jurkat cells. CDs with site-specific information may be more meaningful for cell classification and serve as effective biomarkers for disease detection.

Benefiting from the common features of a glycan containing multiple monosaccharides and one sugar bearing several hydroxyl groups, the current method can dramatically enhance the interactions between boronic acid and glycans, which is critical in analyzing glycoproteins

with low abundance. Furthermore, there are several other advantages. First, this method is quick and easy to operate. As shown in Figure 2.8, the results from 10-min incubation are almost the same as those from two- or three-hour incubation. Glycopeptides are captured under basic conditions and released in an acidic solution. Second, this method is highly reproducible and robust. Third, because the enrichment is based on the reversible interactions, the enriched glycopeptides remain intact, which allows us to analyze glycan structures and also to identify protein O-glycosylation, as demonstrated by the analyses of protein O-mannosylation in yeast and O-GlcNAcylation in human cells. Fourth, because there are no sample restrictions, this method can be extensively applied to analyze different types of samples, from whole cell lysates to clinical and plant samples.

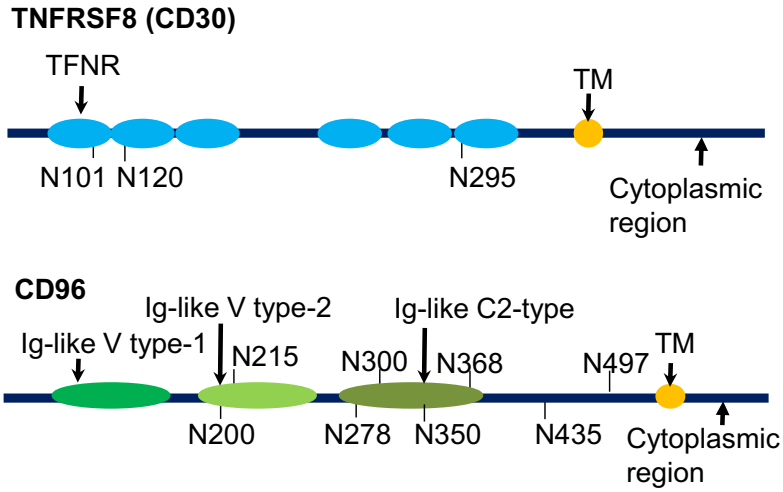


Figure 2.39 Two examples of glycoproteins (CD30 and CD96) with domain and glycosylation site information in Jurkat cells.

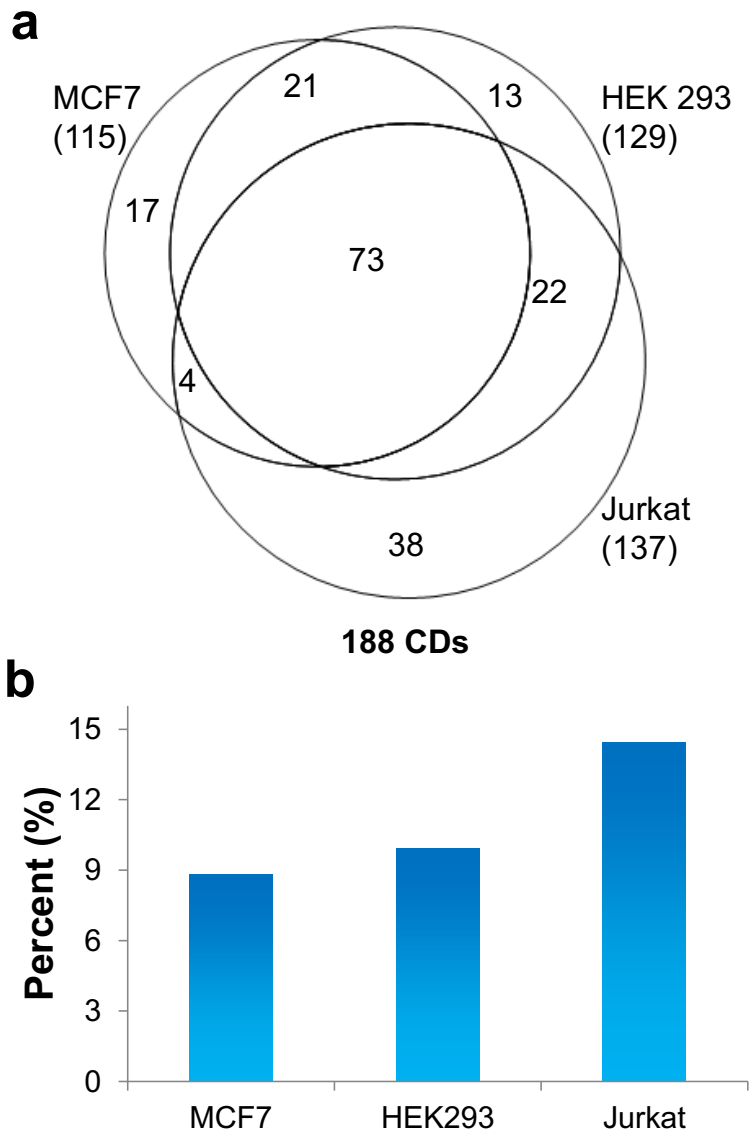


Figure 2.40 The numbers of CD N-glycoproteins (a), and the percentage of CD glycoproteins with respect to all N-glycoproteins (b) identified in each type of human cells.

2.5 Conclusions

The current method is based on the universal and reversible interactions between hydroxyl groups in glycans and boronic acid. The experimental results for yeast and human cells and mouse tissue demonstrated that this method is highly effective in enriching glycopeptides, especially for those with low abundance, and the reversible nature of the

interactions keep enriched glycopeptides intact for both site identification and glycan structure analysis. Due to the biological importance of glycoproteins, their global analysis will aid in a better understanding of glycoprotein functions and the molecular mechanisms of diseases, and the discovery of glycoproteins as drug targets and disease biomarkers.

2.6 References

1. Spiro, R.G. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).
2. Varki, A. et al. Essentials of glycobiology (2nd Edition). (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; 2008).
3. Mahal, L.K., Yarema, K.J. & Bertozzi, C.R. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **276**, 1125-1128 (1997).
4. Wolfert, M.A. & Boons, G.J. Adaptive immune activation: glycosylation does matter. *Nat. Chem. Biol.* **9**, 776-784 (2013).
5. Drake, P.M. et al. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin. Chem.* **56**, 223-236 (2010).
6. Reis, C.A., Osorio, H., Silva, L., Gomes, C. & David, L. Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol.* **63**, 322-329 (2010).
7. Ju, T.Z., Otto, V.I. & Cummings, R.D. The Tn antigen-structural simplicity and biological complexity. *Angew. Chem.-Int. Edit.* **50**, 1770-1791 (2011).
8. Ohtsubo, K. & Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).
9. Wada, Y. et al. Comparison of methods for profiling O-glycosylation human proteome organisation human disease glycomics/proteome initiative multi-institutional study of IgA1. *Mol. Cell. Proteomics* **9**, 719-727 (2010).
10. 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).
11. Kailemia, M.J., Park, D. & Lebrilla, C.B. Glycans and glycoproteins as specific biomarkers for cancer. *Anal. Bioanal. Chem.* **409**, 395-410 (2017).
12. Qiu, Y.H. et al. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J. Proteome Res.* **7**, 1693-1703 (2008).
13. 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).
14. Siuti, N. & Kelleher, N.L. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **4**, 817-821 (2007).
15. Yates, J.R., Ruse, C.I. & Nakorchevsky, A. in Annual Review of Biomedical Engineering, Vol. 11 49-79 (Annual Reviews, Palo Alto; 2009).
16. Trinidad, J.C., Specht, C.G., Thalhammer, A., Schoepfer, R. & Burlingame, A.L. Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol. Cell. Proteomics* **5**, 914-922 (2006).
17. Wu, R.H. et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* **8**, 677-683 (2011).

18. Ge, Y., Rybakova, I.N., Xu, Q.G. & Moss, R.L. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12658-12663 (2009).

19. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* **20**, 301-305 (2002).

20. Mertins, P. et al. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods* **10**, 634-637 (2013).

21. Ludwig, K.R., Sun, L.L., Zhu, G.J., Dovichi, N.J. & Hummon, A.B. Over 2300 phosphorylated peptide identifications with single-shot capillary zone electrophoresis-tandem mass spectrometry in a 100 min separation. *Anal. Chem.* **87**, 9532-9537 (2015).

22. Huang, H., Lin, S., Garcia, B.A. & Zhao, Y.M. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115**, 2376-2418 (2015).

23. Rexach, J.E. et al. Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. *Nat. Chem. Biol.* **6**, 645-651 (2010).

24. Khatri, K. et al. Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. *J. Proteome Res.* **13**, 4347-4355 (2014).

25. Segu, Z.M., Hussein, A., Novotny, M.V. & Mechref, Y. Assigning N-glycosylation sites of glycoproteins using LC/MSMS in conjunction with Endo-M/exoglycosidase mixture. *J. Proteome Res.* **9**, 3598-3607 (2010).

26. Kaji, H. et al. Large-scale identification of N-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. *J. Proteome Res.* **11**, 4553-4566 (2012).

27. Ramachandran, P. et al. Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* **5**, 1493-1503 (2006).

28. Neubert, P. et al. Mapping the O-mannose glycoproteome in *saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **15**, 1323-1337 (2016).

29. Wang, X.S. et al. A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol. Cell. Proteomics* **15**, 2462-2475 (2016).

30. Yang, Y. et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 10 (2016).

31. Zheng, J.N., Xiao, H.P. & Wu, R.H. Specific identification of glycoproteins bearing the Tn antigen in human cells. *Angew. Chem.-Int. Edit.* **56**, 7107-7111 (2017).

32. Olsen, J.V. et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).

33. Lemeer, S. & Heck, A.J.R. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **13**, 414-420 (2009).

34. Phanstiel, D.H. et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat. Methods* **8**, 821-U884 (2011).

35. Zielinska, D.F., Gnad, F., Wisniewski, J.R. & Mann, M. Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907 (2010).

36. Trinidad, J.C., Schoepfer, R., Burlingame, A.L. & Medzihradszky, K.F. N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* **12**, 3474-3488 (2013).

37. 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).

38. Wohlgemuth, J., Karas, M., Eichhorn, T., Hendriks, R. & Andrecht, S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Anal. Biochem.* **395**, 178-188 (2009).

39. Mysling, S., Palmisano, G., Hojrup, P. & Thaysen-Andersen, M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* **82**, 5598-5609 (2010).

40. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O.N. & Roepstorff, P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **3**, 556-566 (2004).

41. Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* **12**, 561-567 (2015).

42. Sun, S.S. et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **34**, 84-88 (2016).

43. Khoury, G.A., Baliban, R.C. & Floudas, C.A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the Swiss-Prot database. *Scientific Reports* **1**, Article Number: 90 (2011).

44. 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).

45. Kaji, H. et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **21**, 667-672 (2003).

46. Hang, H.C., Yu, C., Kato, D.L. & Bertozzi, C.R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14846-14851 (2003).

47. Nilsson, J. et al. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **6**, 809-U826 (2009).

48. Wollscheid, B. et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol.* **27**, 378-386 (2009).

49. Steentoft, C. et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977-982 (2011).

50. Zielinska, D.F., Gnad, F., Schropp, K., Wisniewski, J.R. & Mann, M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell* **46**, 542-548 (2012).

51. Zhang, L.J. et al. Boronic acid functionalized core-satellite composite nanoparticles for advanced enrichment of glycopeptides and glycoproteins. *Chem.-Eur. J.* **15**, 10158-10166 (2009).

52. Chen, W.X., Smeekens, J.M. & Wu, R.H. A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol. Cell. Proteomics* **13**, 1563-1572 (2014).

53. 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).

54. 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).

55. Huttlin, E.L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010).

56. Bern, M.W. & Kil, Y.J. Two-dimensional target decoy strategy for shotgun proteomics. *J. Proteome Res.* **10**, 5296-5301 (2011).

57. 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).

58. Jin, S., Cheng, Y.F., Reid, S., Li, M.Y. & Wang, B.H. Carbohydrate recognition by boronolectins, small molecules, and lectins. *Med. Res. Rev.* **30**, 171-257 (2010).

59. Arnaud, J., Audfray, A. & Imberty, A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem. Soc. Rev.* **42**, 4798-4813 (2013).

60. Li, D.J. et al. A high boronate avidity monolithic capillary for the selective enrichment of trace glycoproteins. *J. Chromatogr. A* **1384**, 88-96 (2015).

61. Sparbier, K., Wenzel, T. & Kostrzewa, M. Exploring the binding profiles of ConA, boronic acid and WGA by MALDI-TOF/TOF MS and magnetic particles. *J. Chromatogr. B* **840**, 29-36 (2006).

62. Xu, G.B., Zhang, W., Wei, L.M., Lu, H.J. & Yang, P.Y. Boronic acid-functionalized detonation nanodiamond for specific enrichment of glycopeptides in glycoproteome analysis. *Analyst* **138**, 1876-1885 (2013).

63. Peters, J.A. Interactions between boric acid derivatives and saccharides in aqueous media: Structures and stabilities of resulting esters. *Coord. Chem. Rev.* **268**, 1-22 (2014).

64. Dowlut, M. & Hall, D.G. An improved class of sugar-binding boronic acids, soluble and capable of complexing glycosides in neutral water. *J. Am. Chem. Soc.* **128**, 4226-4227 (2006).

65. Adamczyk-Wozniak, A., Cyranski, M.K., Zubrowska, A. & Sporzynski, A. Benzoxaboroles - old compounds with new applications. *J. Organomet. Chem.* **694**, 3533-3541 (2009).

66. Xiao, H.P., Tang, G.X. & Wu, R.H. Site-specific quantification of surface N-glycoproteins in statin-treated liver cells. *Anal. Chem.* **88**, 3324-3332 (2016).

67. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature* **425**, 737-741 (2003).

68. 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).

69. Wang, M. et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell. Proteomics* **11**, 492-500 (2012).

70. Wells, L., Vosseller, K. & Hart, G.W. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291**, 2376-2378 (2001).

71. Xu, S.L. et al. Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1536-E1543 (2017).

72. Alfaro, J.F. et al. 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.* **109**, 7280-7285 (2012).

73. Zola, H. et al. CD molecules 2006 - human cell differentiation molecules. *J. Immunol. Methods* **319**, 1-5 (2007).

CHAPTER 3. QUANTITATIVE ANALYSIS OF GLYCOPROTEINS BY COMBINING BORONIC ACID ENRICHMENT AND MS-BASED PROTEOMICS

Partially adapted with permission from American Chemical Society

Xiao, H. P., and Wu, R. H. Simultaneous Quantitation of Glycoprotein Degradation and Synthesis Rates by Integrating Isotope Labeling, Chemical Enrichment, and Multiplexed Proteomics. *Analytical Chemistry,* 2017, 89, 10361-10367. Copyright 2017 American Chemical Society.

Partially adapted with permission from The Royal Society of Chemistry

Xiao, H. P., and Wu, R. H. Quantification of tunicamycin-induced protein expression and N-glycosylation changes in yeast. *Analyst,* 2016, 141, 3731-3745. Copyright The Royal Society of Chemistry 2016.

3.1 Simultaneous Quantitation of Glycoprotein Degradation and Synthesis Rates by Integrating Isotope Labelling, Chemical Enrichment and Multiplexed Proteomics

3.1.1 Introduction

Protein glycosylation plays vital roles in a variety of cellular processes.[1-4] The functions of glycoproteins are intrinsically related to their dynamics, and the presence of glycans on proteins create a steric hindrance that prevents proteases from approaching,[5] thus impacting protein dynamics. Modern mass spectrometry (MS)-based proteomics has offered a unique opportunity for global analysis of glycoproteins,[6-15] but it is still challenging due to the low

abundance of glycoproteins and the heterogeneity of glycan structures.[16-23] Studying protein glycosylation and its dynamics not only advances our knowledge of the underlying mechanisms of many cellular activities and diseases, but also enables us to identify glycoproteins as disease biomarkers and drug targets.[24-28]

There have been many reports about the global analysis of glycoproteins, and considerable progress has been made in recent years.[6, 29-34] However, few of them focused on glycoprotein dynamics on a large scale in spite of its importance.[35] Stable isotope labelling with amino acid in cell culture (SILAC) has been widely used for protein turnover study.[36, 37] Using pulse-chase SILAC, the newly-synthesized proteins can be distinguished from the existing background through incorporation of the heavy (or light) isotopic amino acid residues. The labelling can allow us to generate valuable information about protein degradation and synthesis through mass spectrometric analysis.

In this work, we combined pulse-chase SILAC, chemical enrichment of glycopeptides, and multiplexed proteomics to globally quantify the degradation and synthesis rates of glycoproteins simultaneously. Pulse-chase labelling allowed us to track the protein abundance changes, and in combination with chemical enrichment of glycopeptides we were able to quantify glycoprotein dynamics. After enrichment, we labelled the enriched glycopeptides from multiple time points with the tandem mass tag (TMT) reagents[38] for quantitation with MS-based proteomics and used their abundances to calculate the degradation and synthesis rates of glycoproteins.

3.1.2 Experimental section

3.1.2.1 Cell culture, heavy isotope labeling, and time course-based cell collection

MCF-7 cells (ATCC) were grown in a humidified incubator at 37 °C and 5.0% CO_2 in high glucose Dulbecco's modified eagle's medium (DMEM) (Sigma-Aldrich) with 10% fetal

bovine serum (FBS) (Corning) for each of the triplicate experiments. Heavy isotope labeling of cells was performed with $^{13}C_6$, $^{15}N_2$ L-lysine (Lys^8) and $^{13}C_6$ L-arginine (Arg^6) (Cambridge Isotopes) in SILAC DMEM with 10% dialyzed FBS for about six generations. Media was then switched to normal media with Lys^0 and Arg^0 to begin the time-course experiment. Cells were collected separately at five time points (0, 6, 12, 24 and 48 hours).

3.1.2.2 Cell lysis and protein digestion

Cells were washed twice with phosphate buffered saline (PBS) and pelleted by centrifugation at 500 g for 3 minutes and washed twice with cold PBS. Cell pellets were lysed through end-to-end rotation at 4 ºC for 45 minutes in the lysis buffer (50 mM N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES) pH=7.4, 150 mM NaCl, 0.5% sodium deoxycholate (SDC), and 25 units/ mL benzonase and 1 tablet/ 10 mL protease inhibitor). Lysates were centrifuged, and the resulting supernatant was transferred into new tubes. Proteins were subjected to disulfide reduction with 5 mM dithiothreitol (DTT) (56 ºC, 25 minutes) and alkylation with 14 mM iodoacetamide (room temperature, 20 minutes in the dark). Detergent was removed by the methanol-chloroform protein precipitation method. The purified proteins were digested with 10 ng/ μL Lys-C (Wako) in 50 mM HEPES pH 8.6, 1.6 M urea, 5% ACN at 31 ºC for 16 hours, followed by further digestion with 8 ng/ μL Trypsin (Promega) at 37 ºC for 4 hours.

3.1.2.3 Glycopeptide enrichment, TMT labeling, and deglycosylation

Protein digestions were quenched by addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, followed by centrifugation to remove the precipitate. The peptides were desalted using a tC18 Sep-Pak cartridge (Waters) and lyophilized, subjected to boronic acid-conjugated beads-based enrichment as described previously.[39] The peptides were then eluted

twice by incubating the beads in a solution containing acetonitrile, water, and trifluoroacetic acid at a respective ratio of 50:49:1 for 30 minutes at 37 ºC. Eluates were desalted using tC18 Sep-Pak cartridges and lyophilized. Purified glycopeptides from each time point were labeled with each channel (126, 128, 129, 130, or 131) of the multiplexed TMT reagents (Thermo) following the manufacturer's protocol. Briefly, purified and lyophilized peptides were dissolved in 100 µL of 100 mM triethylammonium bicarbonate (TEAB) buffer, pH= 8.5. Each tube of TMT reagents was dissolved in 41 µL of anhydrous ACN, and 7 µL was transferred into the peptide tube with another 34 µL of ACN. The reaction was performed for 1 hour at room temperature, quenched by adding 8 µL of 5% hydroxylamine and shaking for 15 min. Peptides from all tubes were then mixed, desalted using a tC18 Sep-Pak cartridge, and lyophilized overnight. The dried peptides were deglycosylated with three units of peptide-N-glycosidase F (PNGase F, Sigma-Aldrich)[40] in 60 µL buffer containing 40 mM NH_4HCO_3 (pH=9) in heavy-oxygen water ($H_2{}^{18}O$) for 3 hours at 37 °C. The reaction was quenched by adding formic acid (FA) to a final concentration of 1%, and peptides were desalted again using tC18 Sep-Pak cartridges and dried.

3.1.2.4 Glycopeptide fractionation and LC-MS/MS analysis

Purified and dried peptides were separated by high pH reversed-phase high-performance liquid chromatography (HPLC) into 10 fractions with a 40-min gradient of 5-55% ACN in 10 mM ammonium acetate (pH=10), dried and purified using the stage-tip method, and dissolved in a 10 µL solution with 5% ACN and 4% FA. 4 µL were loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 µm, 200 Å, 100 µm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 128 min gradient.

Peptides were detected with a data-dependent Top15 method[41] in a hybrid dual-cell quadrupole linear ion trap - Orbitrap mass spectrometer (LTQ Orbitrap Elite, Thermo Scientific, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 automatic gain control (AGC) target was followed by up to 15 MS/MS in the Orbitrap again for the most intense ions. The selected ions were excluded from further analysis for 90 seconds. Ions with singly or unassigned charge were not sequenced.

3.1.2.5 Database Search and Data Filtering

All MS^2 spectra were converted into a mzXML format and searched using the SEQUEST algorithm (version 28).[42] Spectra were matched against a database containing sequences of all proteins in the Human (*Homo sapiens*) database downloaded from the UniProt. The following parameters were used during the search: 10 ppm precursor mass tolerance; 0.1 Da product ion mass tolerance; fully digested with trypsin; up to three missed cleavages; fixed modification: carbamidomethylation of cysteine (+57.0214); variable modifications: oxidation of methionine (+15.9949), O^{18} tag of asparagine (+2.9883). For heavy TMT-labeled proteins, these following fixed modifications were also added to the search: TMT plus heavy isotope for lysine (+237.1771), heavy arginine (+6.0201), N-terminal TMT (229.1629). For light TMT-labeled proteins, TMT (+229.1629) was added to both lysine and N-terminal as fixed modification. False discovery rates (FDR) of glycopeptide and glycoprotein identifications were evaluated and controlled to less than 1% by the target-decoy method[43] through linear discriminant analysis (LDA),[44] using parameters such as XCorr, precursor mass error, and charge state, to control the glycopeptide identification quality.[45] The consensus motif N#X[S/T/C] (# stands for the glycosylation site and X represents any amino acid residues other than proline) was also required to guarantee the reliability of the N-glycosylation analysis.

Peptides fewer than seven amino acid residues in length were deleted. The dataset was restricted to glycopeptides when determining FDRs for glycopeptide identification.[46]

3.1.2.6 Glycosylation Site Localization

We assigned and measured the confidence of glycosylation site localizations by calculating their ModScores, which applies a probabilistic algorithm[46] that considers all possible glycosylation sites in a peptide and uses the presence of experimental fragment ions unique to each site to assess the localization confidence. Sites with ModScore > 13 ($P < 0.05$) were considered as confidently localized. If the same glycopeptide was quantified several times, the median value was used as the glycopeptide abundance change.

3.1.3 Results and discussion

3.1.3.1 Experimental procedure for simultaneous measurement of glycoprotein degradation and synthesis rates

The experimental procedure is shown in Figure 3.1, and the detailed description is included in the Experimental Section. Briefly, MCF-7 cells were cultured with SILAC Dulbecco's Modified Eagle's Medium (DMEM) containing Lys^8 and Arg^6 for six generations for full heavy isotope incorporation, and then were equally passaged for the time-course experiments. When the cells were approaching nearly full confluency (to minimize the dilution effect from cell growth), we switched the media to normal DMEM with Lys^0 and Arg^0 and began the time course. Upon the media switch (0 h), the numbers of cells across different groups were kept as similar as possible. For each sample, there were also very similar amount of heavy isotope-labelled proteins (heavy proteins) and nearly no light isotope-labelled proteins (light proteins). We then harvested cells at each time point until the completion of the 48 h time

course. As time went by, heavy proteins were degraded and newly-synthesized proteins were theoretically all light proteins. Therefore, the abundance changes of heavy glycoproteins can be used to calculate the degradation rates while the abundance changes of light glycoproteins as a function of time are glycoprotein synthesis rates. We performed biological triplicate experiments to evaluate the reproducibility and ensure the technical rigor.

Proteins were reduced, alkylated, and digested by Lys-C and trypsin. Purified peptides were subjected to the chemical enrichment of glycopeptides through incubation with boronic acid-conjugated magnetic beads, as reported previously.[39, 47] The beads were then washed to remove nonglycopeptides, and elution was performed twice using a buffer containing water: acetonitrile: trifluoroacetic acid= 49:50:1. Glycopeptides from each time point were purified using C18 cartridges, labelled with the TMT reagents and mixed. Glycopeptides were treated with PNGase F in heavy oxygen water ($H_2{}^{18}O$) to create a common tag on the N-glycosylation sites for MS analysis.[48, 49] After purification, the deglycosylated peptides from each experiment were separated into 10 fractions using high-pH reversed-phase high performance liquid chromatography. Each fraction was further purified by the stage-tip method, followed by LC-MS analysis.

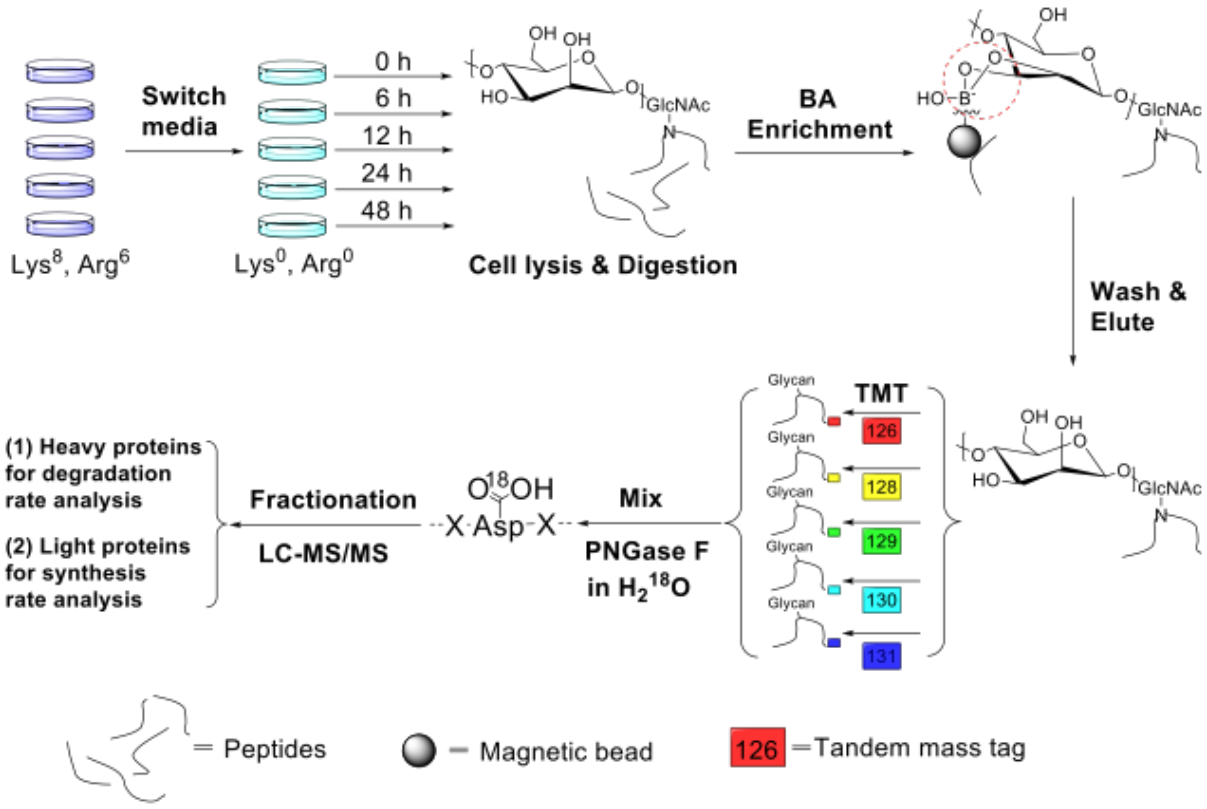


Figure 3.1 The experimental procedure for the simultaneous quantification of the glycoprotein degradation/synthesis rates.

3.1.3.2 Glycoprotein identification

After glycopeptides were treated with PNGase F in heavy oxygen water ($H_2{}^{18}O$), a common and unique tag (+2.9883 D) on the N-glycosylation sites was created for MS analysis. The deamidation of asparagine happens *in vivo* and *in vitro*. The tag containing the heavy oxygen can allow us to distinguish the real N-glycosylation sites from spontaneous deamidation of asparagine. However, the spontaneous deamidation of asparagine could occur during the PNGase F treatment in heavy water, resulting in the false positive identification. In order to minimize this, we carried out the reaction for only three hours. As tested previously, the spontaneous deamidation of asparagine is negligible for three hours under the mild enzymatic reaction conditions.[48]

The glycopeptides were filtered to <1% false discovery rate. Additionally they are required to have the consensus motif, i.e. N#X[S/T/C] (# stands for the glycosylation site and X represents any amino acid residues other than proline). Among biological triplicate experiments, we identified 1,373, 1,342, and 1,280 unique light glycopeptides (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241), respectively. They overlapped very well, and 790 glycopeptides were found in all three experiments (Figure 3.2a). Totally, 1,875 unique light glycopeptides were identified. Slightly fewer number of heavy glycopeptides were identified, i.e. 866, 1,048 and 1,097 in each of the three experiments (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241). Finally, 1,515 unique heavy glycopeptides were identified with site-specific information, and the comparison is displayed in (Figure 3.2b).

3.1.3.3 Calculation of the glycoprotein degradation and synthesis rates

We calculated the degradation/synthesis rates based on the abundance changes of glycopeptides as a function of time simulated by the following exponential decay/growth equation (1) or (2), as performed previously: [50, 51]

Based on the abundance changes of heavy glycopeptides, the degradation rates were calculated:

$$P_h(t) = P_{h0} * \exp(-k_d t) \quad (1)$$

According to the abundance changes of light glycopeptides, the synthesis rates were obtained using the following equation:

$$P_l(t) = P_{l0} * \exp(k_s t) \quad (2)$$

where P_{h0} or P_{l0} is the abundance of the heavy or light glycopeptide at the first time point (represented by the intensity of the reporter ion), $P_h(t)$ or $P_l(t)$ is the abundance of the heavy or light glycopeptide at each subsequent time point, t stands for time. k_d is the degradation rate constant while k_s is the synthesis rate constant. The different time points at 0, 6, 12, 24, and 48

h were also designed to provide convenience for the exponential simulation and to be compatible with the measurement of relatively long half-lives of glycoproteins. [48, 49]

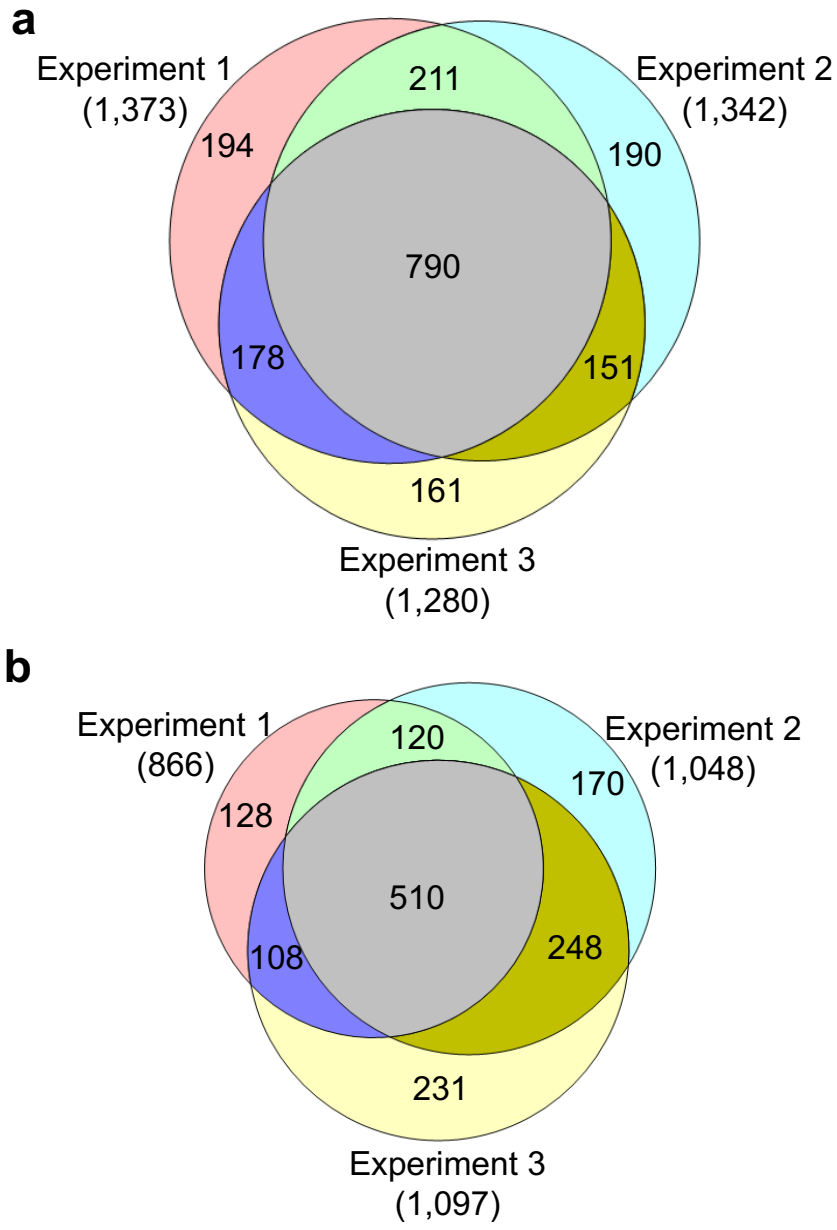


Figure 3.2 The overlap of the unique glycopeptides identified in the biological triplicate experiments: (a) light glycopeptides; (b) heavy glycopeptides.

An example of glycopeptide quantification is shown in Figure 3.3. Glycopeptide KWGHN#ITEFQQR is from protein ERO1A, an oxidoreductase involved in disulfide bond formation and preventing the accumulation of reactive oxygen species in the endoplasmic reticulum (ER). This glycopeptide was confidently identified with an XCorr of 3.8, and the glycosylation site was localized on N280, which was also reported on the UniProt (www.uniprot.org). We quantified its synthesis rate based on the reporter ion intensities. Through using this method, we quantified the synthesis rates of 847 glycoproteins (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241) and the degradation rates of 704 glycoproteins (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241) from the three experiments.

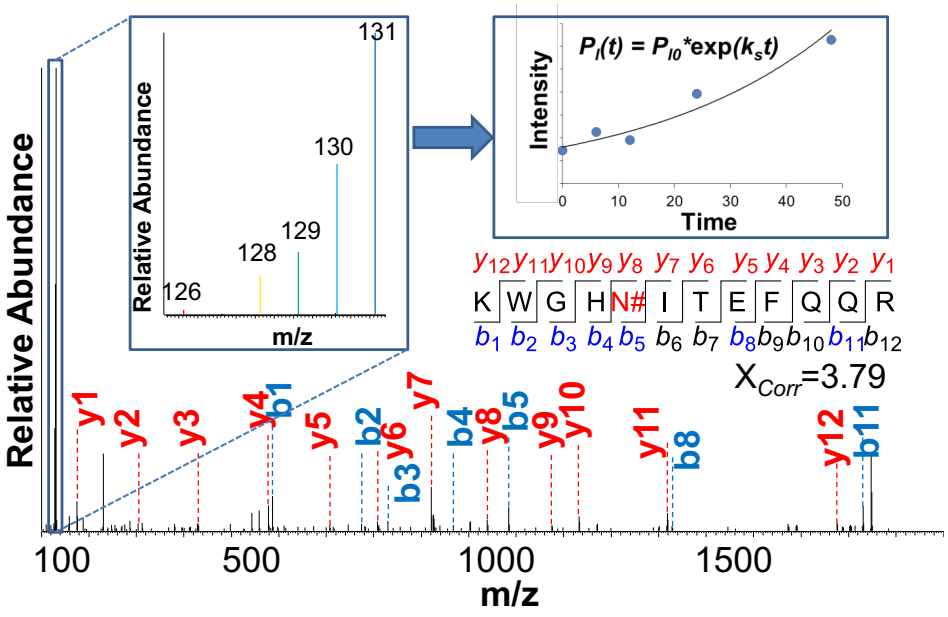


Figure 3.3 An example of glycopeptide identification and quantification.

3.1.3.4 Evaluation of the experimental reproducibility

The reproducibility was evaluated based on the biological triplicate experiments (Figure 3.4 and 3.5). The comparison of the synthesis rates from triplicate experiments is shown in Figure 3.5. A total of 1,330 (71%) unique light glycopeptides were identified in at least two

experiments (Figure 3.2a) displaying high reproducibility given that all experiments were performed independently from cell culture to LC-MS/MS analysis. In addition, the calculated synthesis rates of the light glycoproteins from the three experiments were in reasonably good agreement (Figure 3.5).

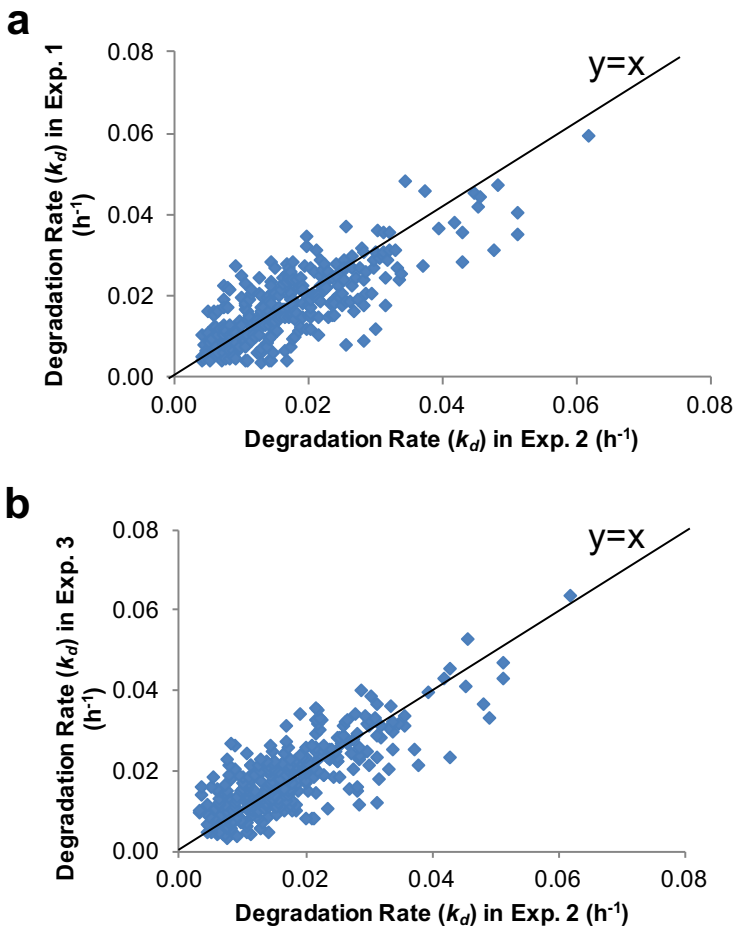


Figure 3.4 Reproducibility evaluation of the heavy glycopeptides/glycoproteins: (a) Comparison of the degradation rates of the N-glycoproteins quantified in the experiments 1 & 2, and (b) Comparison of the degradation rates of the N-glycoproteins quantified in the experiments 2 & 3.

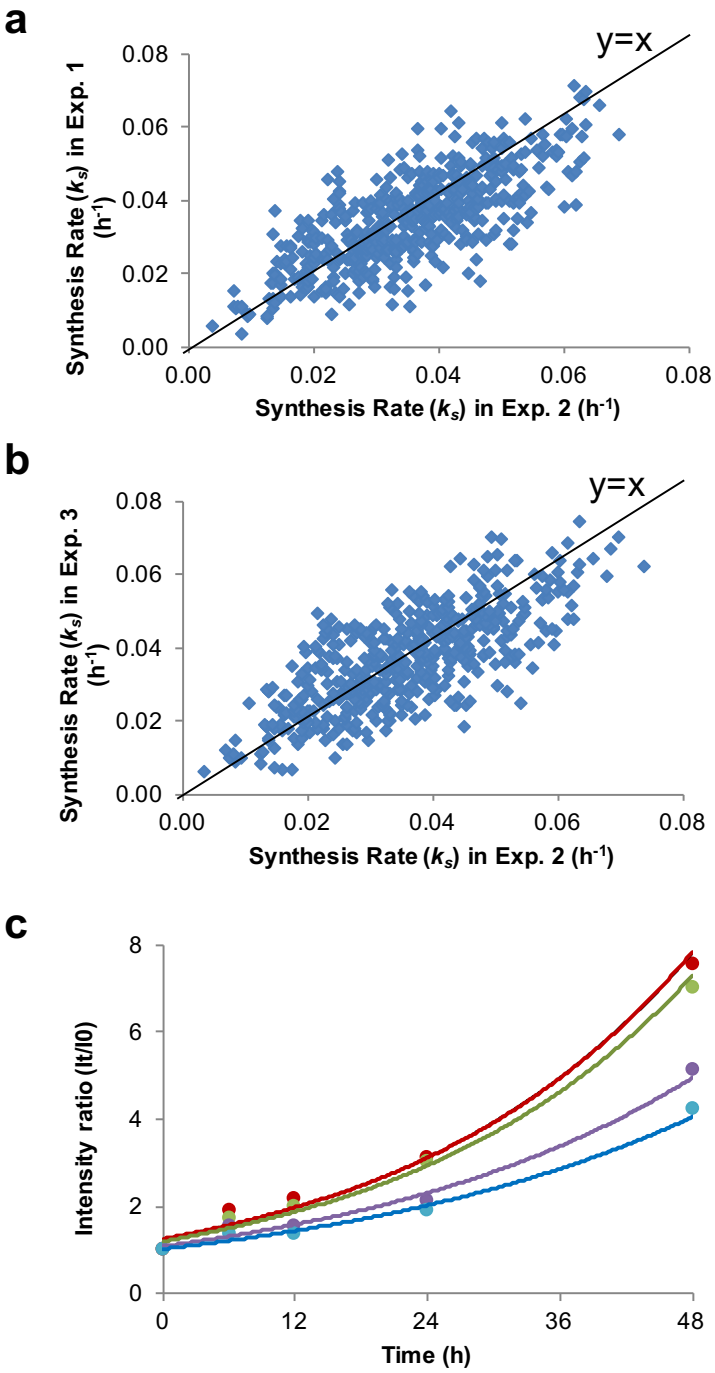


Figure 3.5 Comparison of the synthesis rates of the glycoproteins quantified in experiments 1 & 2. (b) Comparison of the synthesis rates of the glycoproteins quantified in experiment 2 & 3. (c) Examples of glycopeptide quantification: red- KPN#ATAEPTPPDR from protein MRC2, green- RELYN#GTADITLR from protein RPIEZO1, purple- TCDWLPKPN#MSASCK from protein PSAP, and blue- QPMAPNPCEANGGQGPCSHLCLINYN#R from protein LRP1.

The heavy glycopeptide identification and quantification were also proved to be reproducible (Figure 3.2b and 2.44). Compared to the protein synthesis rates, their degradation

rates are generally lower, resulting in a more condensed distribution pattern for the heavy glycoproteins, which is further discussed below.

We quantified both the synthesis (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241) and degradation rates (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241) for 639 glycoproteins (Figure 3.6). We illustrated the dynamic abundance changes over the time course for several light glycopeptides as examples in Figure 3c. Due to the fact that the raw intensities for each peptide can vary within a range of several orders of magnitudes, we used the normalized ratio of intensity at each time point, i.e. the intensity (I_t) divided by the initial intensity (I_0). Therefore, the values of all glycopeptides are 1 at the first time point. The glycopeptides are from proteins MRC2, PIEZO1, PSAP, and LRP1, representing glycoproteins from a variety of subcellular locations and with various molecular functions.

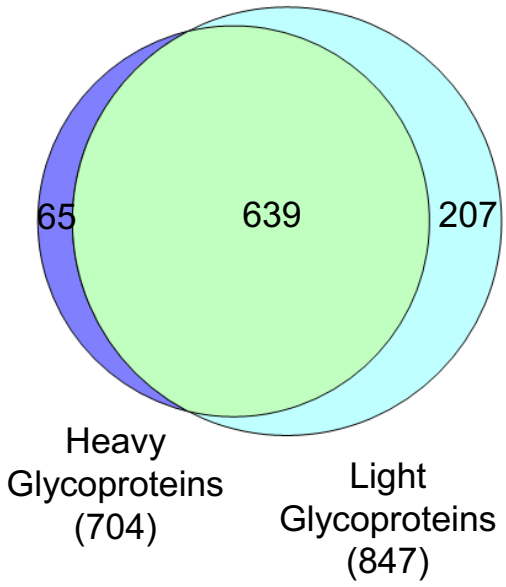


Figure 3.6 The overlap of the glycoproteins with the degradation and synthesis rates quantified.

Since the time course lasted for 48 h, the accumulation of light glycoproteins made their abundance higher than heavy glycoproteins, rendering slightly fewer heavy glycoproteins being identified. Although the current time course was relatively long, a small group of proteins

with a very slow degradation or synthesis rate ($k<0.0034\ h^{-1}$, $t_{1/2}>200$ h) were still not able to be accurately quantified, thus we annotated the rates for these proteins as "very slow" in the Supplementary Tables. For glycoproteins quantified in multiple experiments, we used their median synthesis and degradation rates for further data analysis in this work.

3.1.3.5 Clustering of glycoproteins

We clustered the identified glycoproteins according to cellular compartment using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Figure 3.7a).[52] Many categories, ranging from cell surface to organelle membranes, were highly enriched with low *P* values and high glycoprotein counts. We also clustered the glycoproteins with a relatively high synthesis rate ($k_s>0.03\ h^{-1}$) according to molecular function . Interestingly, the highly enriched categories are receptor binding and transportation. The most highly enriched are those with glycosylation enzymatic activity, including transferase activity (transferring hexosyl groups) and hydrolase activity (hydrolyzing O-glycosyl compounds) (Figure 3.7b)

Here we quantified the synthesis rates of 83 CD glycoproteins (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241), and twelve of them also belong to the family of cell adhesion molecules (CAMs), which are listed in Table 3.1**.** Many of them are integrins, which are typically a group of important transmembrane receptors that participate in the interactions between cells and extracellular matrix. For instance, ITGB1 was confidently quantified in the current experiments based on 11 unique glycopeptides. It is known to be conjoining with integrin alpha subunits to form various cell-surface receptors, such as forming a laminin receptor with integrin alpha subunit 6 (ITGA6). The latter is also quantified with a synthesis rate of 0.0380 h^{-1}. The four integrins (ITGB1, ITGB2, ITGAV, and ITGA6)

quantified all have similar synthesis rates, ranging from 0.0377 h^{-1} to 0.0483 h^{-1}, correlating well with the fact that they adjoin one another to form the receptor complex on the cell surface.

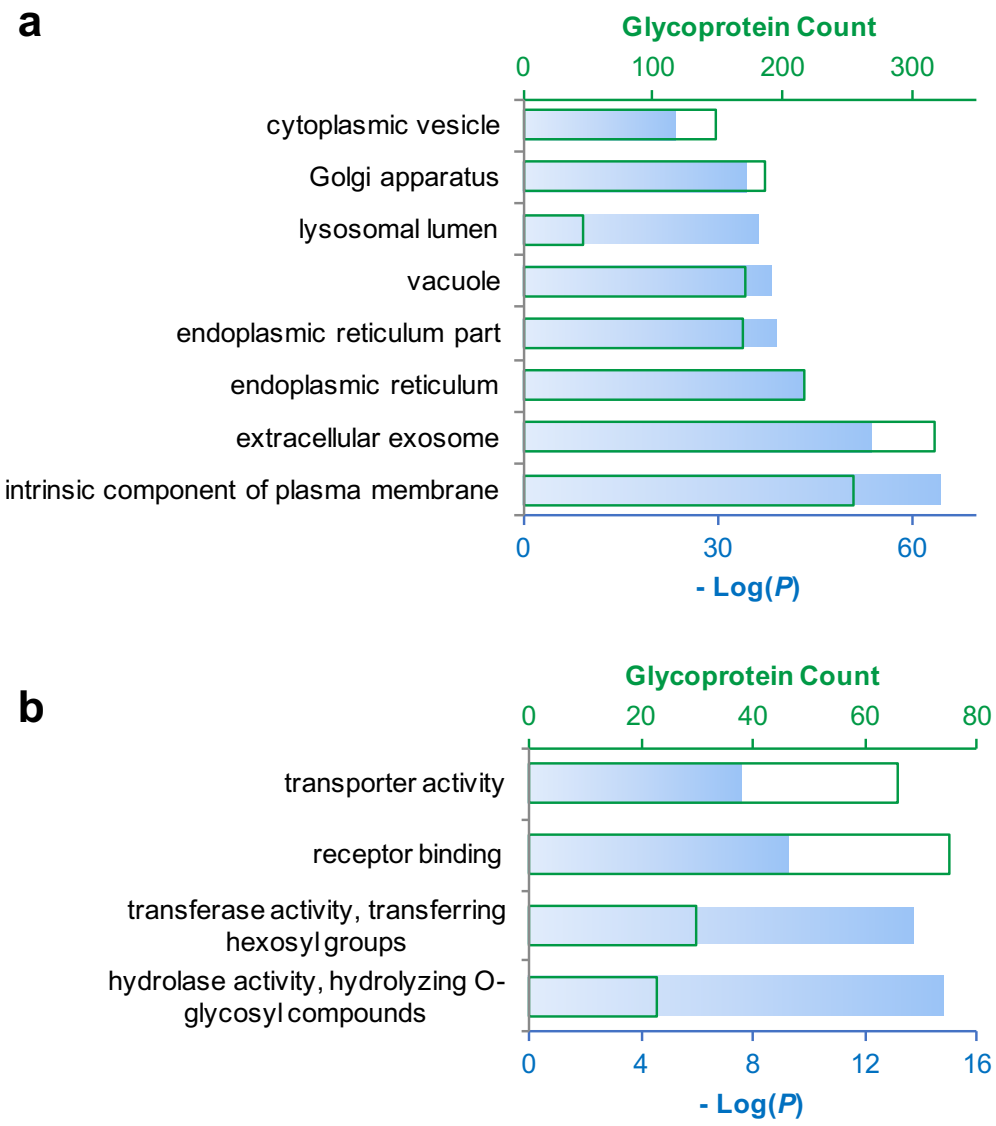


Figure 3.7 Clustering of (a) the quantified glycoproteins according to cellular compartment and (b) the glycoproteins with a relatively higher synthesis rate based on molecular function.

Table 3.1 The 12 glycoproteins that are both CD and CAM molecules.

CD Name	Gene Symbol	UniProt ID	Unique Glycopeptide Hits	Median Synthesis Rate (h^{-1})	Annotation
CD58	CD58	P19256	1	0.0552	Lymphocyte function-associated antigen 3
CD54	ICAM1	P05362	3	0.0538	Intercellular adhesion molecule 1
CD166	ALCAM	Q13740	9	0.0517	CD166 antigen
CD29	ITGB1	P05556	11	0.0483	Integrin beta-1
CD276	CD276	Q5ZPR3	4	0.0427	CD276 antigen
CD51	ITGAV	P06756	8	0.0412	Integrin alpha-V
CD171	L1CAM	P32004	6	0.0388	Neural cell adhesion molecule L1
CD49f	ITGA6	P23229	5	0.0380	Integrin alpha-6
CD18	ITGB2	P05107	3	0.0377	Integrin beta-2
CD62l	SELL	P14151	1	0.0272	L-selectin
CD275	ICOSLG	O75144	3	0.0205	ICOS ligand
CD155	PVR	P15151	1	0.0046	Poliovirus receptor

3.1.3.6 Comparison of the difference between the synthesis and degradation rates

We then analyzed the difference between the synthesis and degradation rates for glycoproteins (listed in a table online at doi.org/ 10.1021/acs.analchem.7b02241), and the rate differences are plotted in Figure 3.8. To ensure the analysis confidence, we only analyzed the glycoproteins with their synthesis/degradation rates accurately quantified (without "very slow" annotation) in at least two experiments, where 400 proteins fall into this category. Since cells were still growing throughout the time course, we anticipated that the majority of the proteins would have a faster synthesis rate than degradation rate. This was proved by the results (Figure 3.8a) as 352 proteins (88%) had higher synthesis rates.

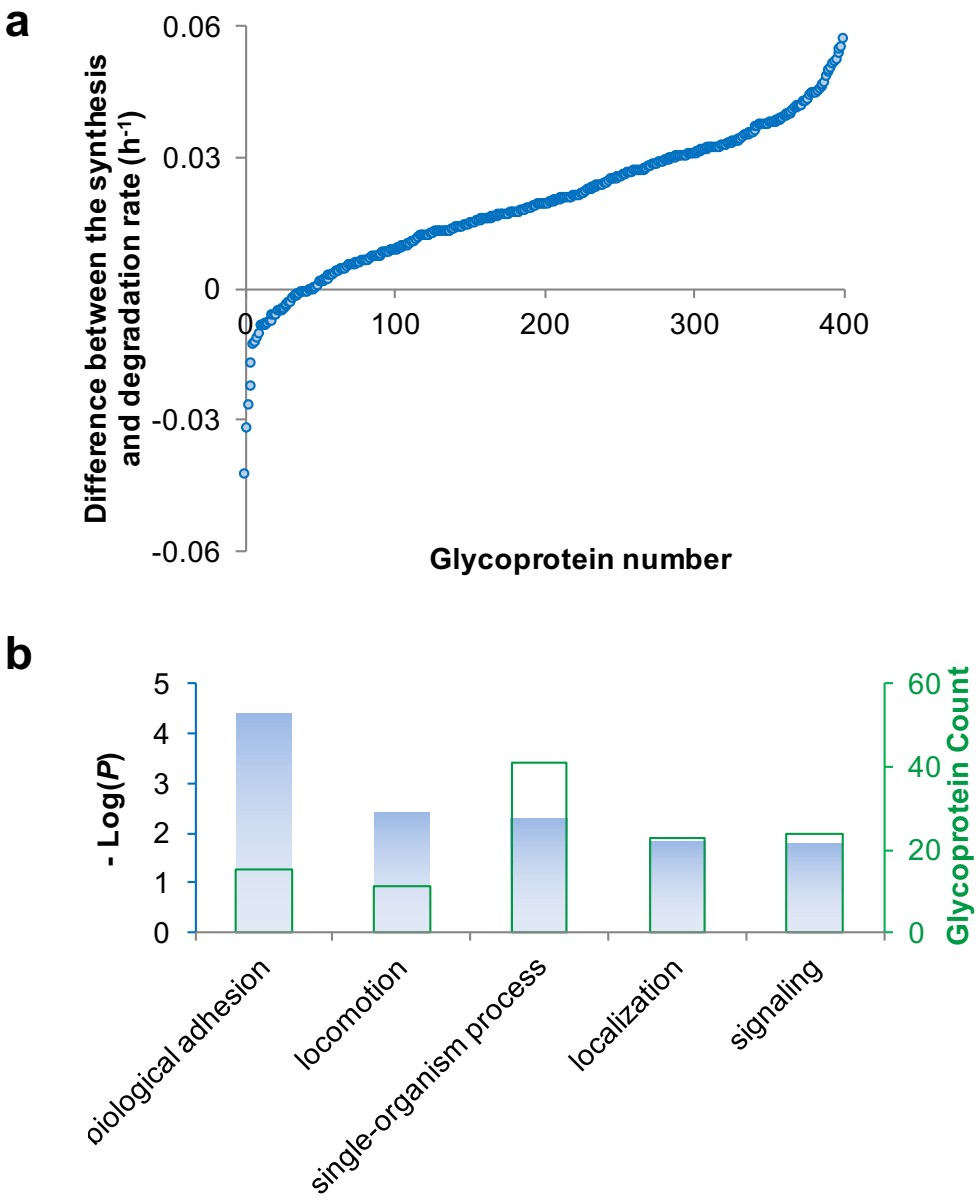


Figure 3.8 The difference between the synthesis and degradation rates for 400 glycoproteins with both rates quantified; (b) the biological processes in which 48 glycoproteins with a lower synthesis rate are involved in.

Interestingly, 48 proteins had lower synthesis rates, and they mostly participated in the biological processes of adhesion, locomotion, localization, and signaling (Figure 3.8b). We reasoned that at the end of the time course, as the cells were approaching a static state due to high confluence, these biological processes were not supposed to robustly continue, and thus the related glycoproteins were down-regulated. Another possible explanation arises from the current approach to calculate these two rates. The absolute glycoprotein degradation rate was calculated based on the abundance changes of heavy glycoproteins, and the contribution from heavy protein synthesis was negligible since heavy isotopic lysine and arginine were not supplied during the 48 h time course. However, light glycoproteins were used to calculate the synthesis rates while they were synthesized and degraded simultaneously. Since the starting light protein abundance was extremely low, we neglected the contribution from light protein degradation, but this contribution accumulated throughout the whole time course. Therefore, the real protein synthesis rates are likely slightly higher than the rates obtained in this work.

3.1.4 Conclusions

Evolution has endowed cells the ability to synthesize proteins in a conservative and low-risk pattern.[53] In this study, we integrated pulse-chase SILAC, chemical enrichment of glycopeptides, and multiplex proteomics to simultaneously investigate the glycoprotein synthesis and degradation rates in human cells on a large scale. Rigorous criteria were applied for glycopeptide filtering, and 3,390 unique heavy and light glycopeptides with site-specific information led to the simultaneous quantitation of the degradation and/or synthesis rates of many glycoproteins. We quantified the synthesis rates of 847 N-glycoproteins and the degradation rates of 704 N-glycoproteins, and demonstrated this method to be reproducible based on the results from the biological triplicate experiments. The glycoproteins related to binding, transportation, and enzyme activity were determined to have higher synthesis rates.

The majority of the quantified glycoproteins were synthesized faster than degraded due to the cell growth. In combination with pulse-chase SILAC and glycopeptide enrichment, we can simultaneously quantify the synthesis and degradation rates of glycoproteins. This method can be extensively applied to investigate glycoprotein dynamics, which will aid in a better understanding of glycoprotein functions and the molecular mechanisms of biological events.

3.2 Quantification of Tunicamycin-Induced Protein Expression and N-Glycosylation Changes in Yeast

3.2.1 Introduction

Glycosylation is a prevalent protein modification in eukaryotic cells that plays essential roles in regulating protein folding, trafficking and stability.[27, 54, 55] Aberrant glycosylation is frequently related to human disease, including cancer and infectious diseases.[27, 56-62] In eukaryotic cells, N-glycosylation typically begins with the synthesis of the dolichol-linked precursor oligosaccharide ($GlcNAc_2Man_9Glc_3$), followed by *en bloc* transfer of the precursor oligosaccharide to newly synthesized peptides in the endoplasmic reticulum (ER).[1, 63] Then oligosaccharide is further trimmed and modified by many enzymes in the Golgi apparatus.[64] The pathway for N-glycosylation synthesis is conserved from yeast to mammalian cells.[65] Although yeast primarily contains high-mannose glycans which differ from those in mammalian cells,[66] it can still be used as an excellent model system to study protein N-glycosylation.[4]

Tunicamycin (TM), a glucosamine-containing antibiotic, blocks N-linked glycosylation by inhibiting the formation of the N-acetylglucosamine-dolichol-phosphate intermediate and thus traps cells in the G1 phase of the cell cycle.[4, 67] TM was originally isolated and utilized for its antiviral activity by suppressing viral glycoprotein synthesis and membrane genesis.[68, 69] Now tunicamycin is extensively used for protein N-glycosylation manipulation. In yeast, the presence of TM has been reported to disrupt the formation of the external glycoprotein invertase, acid phosphatase, and cell wall mannan.[67] Although the mechanism responsible for TM-initiated inhibition of protein N-glycosylation has long been appreciated, a comprehensive and quantitative analysis of the affected proteome and glycoproteome in cells has yet to be conducted.

In recent years, MS-based proteomics methods have become increasingly powerful to systematically study protein expression and modification changes in complex biological samples.[29, 70-77] However, it is still challenging to investigate low-abundance proteins, which requires effective fractionation or other sample preparation.[78-82] Furthermore, the global analysis of glycoproteins in complex biological samples is extraordinarily difficult because of the high heterogeneity of glycans and low abundance of many glycoproteins.[32] Enrichment of glycopeptides and the generation of a common mass tag on glycosylation sites are required prior to MS analysis. In our recent study, based on a common feature of glycans, i.e. multiple hydroxyl groups in each glycan, boronic acid-based enrichment was used to effectively enrich glycopeptides in yeast whole cell lysates.[39] By incorporating this enrichment method, it is possible to comprehensively quantify protein glycosylation changes with quantitative proteomics.

Using yeast as a model system, we systematically investigated the cell response to TM at the proteome and N-glycoproteome levels. We quantified 4,259 proteins, which nearly covers the entire yeast proteome. Many proteins related to several glycan metabolism and glycolysis-related pathways were down-regulated in TM-treated cells. We also globally quantified protein N-glycosylation changes as a result of the TM treatment. Among down-regulated glycoproteins, those related to glycosylation, glycoprotein metabolic processes, carbohydrate processes, and cell wall organization were highly enriched. The current results clearly demonstrate that there are dramatic protein expression and N-glycosylation changes resulting from the tunicamycin treatment.

3.2.2 Experimental section

3.2.2.1 Yeast strains, SILAC labeling, and TM treatment conditions

Yeast (*Saccharomyces cerevisiae*) cells were seeded in "heavy" (Lys^8 ($^{13}C_6$ and $^{15}N_2$); Arg^6 ($^{13}C_6$) Cambridge isotopes) or "light" (Lys^0, Arg^0) media (synthetic complete medium with lysine and arginine drop-out) and cultured overnight. Tunicamycin (TM) (Cayman Chemicals) stock solution (10 mg/mL) was prepared by dissolving TM in dimethyl sulfoxide (DMSO). When the cell population had undergone more than ten doubling times and reached the exponential growth phase (OD=0.3 at 600 nm), TM (2 μg/mL) was added into the "heavy" media while the "light" cells were treated by the same amount of DMSO as a vehicle control. After treatment for three hours, cells were harvested and mixed at a 1:1 ratio based on measured protein concentrations.

3.2.2.2 Cell lysis, protein extraction and digestion

Cells were washed twice with deionized water, pelleted by centrifugation at 4,000 g for 5 minutes, and then resuspended in lysis buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH=7.6, 150 mM NaCl, 0.5% sodium deoxycholate (SDC), 20 U/mL benzonase, and 1 protease inhibitor tablet per 10 mL buffer). Cell lysis was performed using a MiniBeadbeater (Biospec), three 30 second cycles at maximum speed, with 2 minute pauses on ice in between each cycle. Lysates were then centrifuged at 15,000 g for 10 minutes and the resulting supernatant was transferred into a new tube. The protein concentration was measured by a bicinchoninic acid (BCA) assay (Pierce) and proteins were subjected to disulfide reduction with 5 mM dithiothreitol (DTT) (56 ºC, 25 minutes) and alkylation with 14 mM iodoacetamide (RT, 20 minutes in the dark). Detergents were removed by methanol-chloroform protein precipitation. The purified proteins were digested with 10 ng/μL Lys-C (Wako) in buffer containing 50 mM HEPES pH=8.6, 1.6 M urea, and 5% ACN,

at 31 ºC for 16 hours, followed by further digestion with 8 ng/uL Trypsin (Promega) at 37 ºC for 4 hours.

3.2.2.3 Peptide separation, fractionation, and glycopeptide enrichment

Protein digestions were acidified by the addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, followed by centrifugation to remove the precipitate. Then peptides were desalted using a tC18 Sep-Pak cartridge (Waters). Purified peptides were aliquoted into two portions: ~0.5 mg for protein analysis and a ~8 mg for glycosylation analysis. For protein analysis, lyophilized peptides were fractionated into 20 fractions by high pH reversed-phase high-performance liquid chromatography (HPLC) with a 40 minute gradient of 5-55% ACN in 10 mM ammonium acetate (pH=10), and then desalted again using stage-tips. For glycosylation analysis, the separation and enrichment of glycopeptides was carried out by utilizing the covalent interaction between boronic acid and glycans containing multiple hydroxyl groups, as described previously.[39] Peptides were directly subjected to glycopeptide enrichment without HPLC fractionation. Then we separated the glycopeptides into three fractions during the later stage-tip step. Briefly, the peptide mixture was dissolved in 200 mM ammonium acetate buffer (pH=10), and incubated with boronic acid-conjugated magnetic beads at 37 ºC for 1 h. The beads were then washed five times with the binding buffer to remove non-specifically bound peptides. Glycopeptides were eluted by incubating the beads in a solution containing acetonitrile, water, and trifluoroacetic acid at a respective ratio of 50:49:1 for 30 minutes at 37 ºC. Eluates were desalted using tC18 Sep-Pak cartridges and lyophilized overnight.

3.2.2.4 PNGase F treatment for glycopeptides

Glycopeptides were deglycosylated with five units of peptide-*N*-glycosidase F (PNGase F, Sigma-Aldrich) in 100 μL buffer containing 50 mM NH_4HCO_3 (pH=9) in heavy-

oxygen water ($H_2{}^{18}O$) for 3h at 37 °C.[83, 84] The reaction was quenched by adding formic acid (FA) to a final concentration of 1%. Peptides were further purified via stage tip and separated into 3 fractions using 20%, 50% and 80% ACN containing 1% HOAc.

3.2.2.5 LC-MS/MS analysis

All purified and dried peptide fractions were dissolved in a solvent containing 5% ACN and 4% FA, and a fraction of each sample was loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 μm, 200 Å, 100 μm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). For protein analysis, peptides were separated by reversed-phase liquid chromatography using an UltiMate 3000 binary pump with a 90-minute gradient of 4-30% ACN containing 0.125% FA. For the enriched glycopeptide samples, a 110-minute gradient of 3-25%, 8-38%, 10-50% ACN with 0.125% FA was used for each of the three fractions. Peptides were detected with a data-dependent method[41, 85] in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap (10^6 AGC target) was followed by up to 20 MS/MS in the LTQ for the most intense ions. The selected ions were excluded from further sequencing for 90 seconds. Ions with single or unassigned charges were not selected for MS/MS scans. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

3.2.2.6 Database search and data filtering

Raw mass spectra were converted into mzXML format, and then searched using the SEQUEST algorithm (version 28).[42] The following parameters were used during the search: 10 ppm precursor mass tolerance; 1.0 Da product ion mass tolerance; fully digested with

trypsin; up to three missed cleavages; fixed modifications: carbamidomethylation of cysteine (+57.0214); variable modifications: oxidation of methionine (+15.9949), ^{18}O tag on asparagine (+2.9883, for glycosylation analysis), heavy lysine (+8.0142) and heavy arginine (+6.0201). The target-decoy method[43, 86] was employed to determine the false discovery rate (FDR). Linear discriminant analysis (LDA) was then performed to control the quality of peptide identifications using parameters such as XCorr, charge state and precursor mass accuracy,[45] which is also similar to the previous report.[44] Peptides fewer than seven amino acid residues in length were considered unreliable and deleted. Peptide spectral matches were filtered to a <1% FDR. For protein analysis, the peptide-level FDR was calculated based on all identified peptides. For glycoprotein analysis, the dataset was restricted to glycopeptides when determining FDRs for glycopeptide identification.[46, 87] Furthermore, an additional protein-level filter was applied in each dataset to reduce the protein-level FDRs (<1%) for proteins and glycoproteins. Consequently, the FDRs at the peptide level were much less than 1%.

3.2.2.7 Glycosylation site localization and peptide quantification

We used ModScores to assign glycosylation sites and measure the confidence of their localizations.[46, 88] The ModScore software considers all possible glycosylation sites in a peptide and uses the presence of experimental fragments unique to each site to determine the actual glycosylation site and calculated ModScore value based on the binominal probability P ($ModScore = -10 \times Log_{10}(P)$). We considered ModScore > 13 ($P < 0.05$) as confidently localized. If the same peptide was quantified several times, the median heavy-to-light (H/L) value was used as the peptide abundance change. For peptides used for proteome analysis, we required that either its heavy or light isotope peak had a signal-to-noise ratio (S/N) greater than 3. If the S/N of the heavy peak was less than 3, then we required that the light peak had an S/N great than 5, and *vice versa*. Two criteria were applied for glycosylation site quantification: (1)

the quantified glycopeptide must contain only one glycosylation site; (2) the site must be confidently localized with a ModScore >13.

3.2.3 Results and Discussion

3.2.3.1 Tunicamycin treatment and glycoprotein enrichment

Tunicamycin has been widely used to model specific types of stress that affect protein folding in the ER.[89, 90] However, the protein abundance changes in tunicamycin-treated cells have remained unexplored on a large scale. Our first aim was to study the proteome changes resulting from the TM treatment. Because TM is known to inhibit the formation of the N-acetylglucosamine-dolichol-phosphate intermediate and thus prevents protein N-glycosylation, we also systematically investigated N-glycoproteome alterations in TM-treated yeast cells. Since many membrane proteins are known to be glycosylated, 0.5% sodium deoxycholate (SDC) was added into the lysis buffer to increase membrane protein extraction. As a detergent, SDC can disrupt and dissociate many types of protein interactions, and also increase the solubility of membrane proteins. After cell lysis, protein extraction and purification, 0.5 mg of digested peptides were separated into 20 fractions using high-pH reversed phase liquid chromatography (Figure 3.9). In combination with further separation under acidic conditions during on-line LC-MS/MS analysis, two-dimensional orthogonal separation can minimize peptide peak overlap and boost the identification of low-abundance proteins.

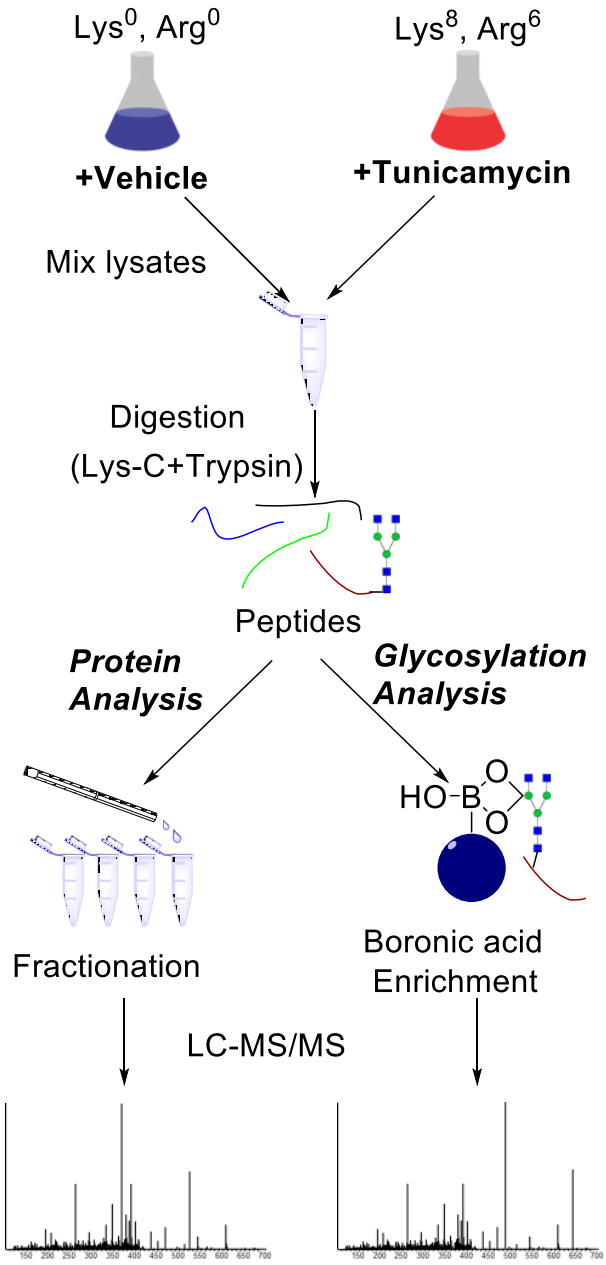


Figure 3.9 Experimental procedure for the global analysis of proteins and N-glycoproteins in TM-treated yeast cells *vs*. untreated cells.

Two technical challenges must be overcome to globally study protein N-glycosylation by using MS-based proteomics techniques: the low expression levels of many glycoproteins, and the heterogeneity of glycan structures. Therefore, an effective enrichment method and an efficient approach to generate a common tag on glycosylation sites for subsequent database

searching are required. Based on one common features of glycoproteins, i.e. glycan structures bearing multiple hydroxyl groups, we globally enriched glycoproteins and/or glycopeptides through the universal boronic acid-*cis* diol recognition.[39] Boronic acid was immobilized onto magnetic beads to capture glycopeptides, and the reversible nature of the covalent interactions between boronic acid and diols made it possible to release glycopeptides (after the removal of non-specifically bound peptides) for further analysis. After enrichment, peptides were treated with PNGase F in heavy-oxygen water ($H_2{}^{18}O$) to remove N-glycans, which converted asparagine (Asn) to aspartic acid (Asp) containing heavy oxygen and created a mass shift of +2.9883 Da.[30, 40] Heavy oxygen on Asp allows us to easily distinguish authentic N-glycosylation sites from those caused by deamidation occurring *in vitro* and *in vivo*. Treatment time was shortened to only 3 hours to minimize possible deamidation that occurs during the PNGase F cleavage process.

3.2.3.2 Examples of peptide and glycopeptide identification and quantification

In order to accurately quantify the protein expression and glycosylation changes, an Orbitrap mass spectrometer with high resolution and mass accuracy (Thermo hybrid LTQ-Orbitrap Elite MS) was used in this study. Figure 3.10 shows examples of peptide and glycopeptide identifications and quantifications. Both peptides are from the protein PDI1 (YCL043C), which is a disulfide isomerase essential for the formation of disulfide bonds in secretory and cell-surface proteins, and may unscramble non-native disulfide bonds. In addition, it participates in the processing of unfolded protein-bound $Man_8GlcNAc_2$ oligosaccharides to $Man_7GlcNAc_2$, thereby promoting degradation in unfolded protein response (http://www.yeastgenome.org). This protein was determined to be up-regulated by 2.6 fold in TM-treated yeast cells, possibly as a result of TM interrupting the proper glycosylation of various proteins, and unfolded or misfolded proteins accumulating in the ER.

Heavy isotope peaks of the peptide GLMNFVSIDAR are shown to be more than twice as intense as the light peaks in Figure 3.10a.

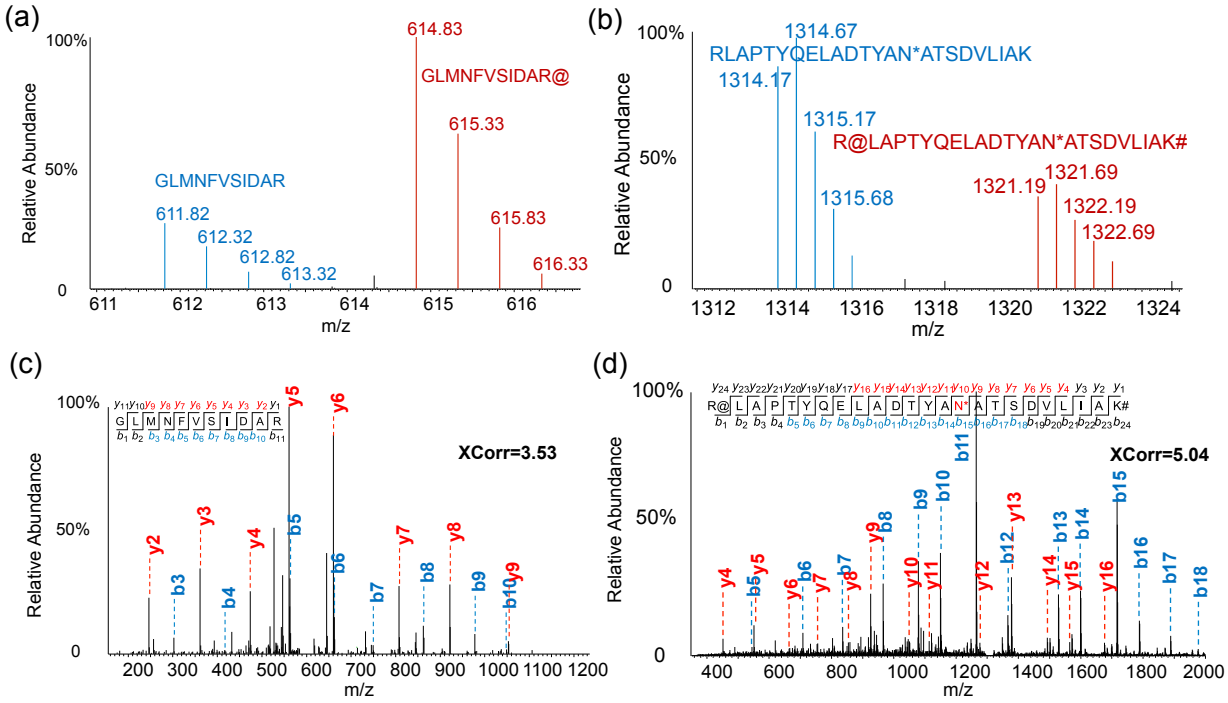


Figure 3.10 Examples of full and tandem mass spectra of peptides. (a) The full and (c) tandem mass spectra of the peptide GLMNFVSIDAR and (b) the full and (d) tandem mass spectra of the glycopeptide RLAPTYQELADTYAN*ATSDVLIAK. Both peptides are from the protein PDI1. (c) and (d) demonstrated that the two peptides were confidently identified with high XCorr values. (@-heavy arginine, #-heavy lysine, *-glycosylation site)

In TM-treated yeast cells, the glycosylation site N425 on this protein was down-regulated by 2.5 fold. The tandem mass spectra corresponding to the identification of the glycopeptide R@LAPTYQELADTYAN*ATSDVLIAK# (@-heavy arginine, #-heavy lysine and *-glycosylation site) is shown in Figure 3.10b; this glycopeptide was confidently identified with an XCorr of 5.04 and a ModScore of 1000. The two ratios of peptides and glycopeptides from the same protein are excellent examples of differential protein expression and glycosylation changes resulting from the TM treatment. Several other glycosylation sites (N82,

N117, N155 and N174) on this protein were also down-regulated (listed in a table online at doi.org/10.1039/C6AN00144K).

3.2.3.3 Global analysis of protein abundance changes

After protein samples were fractionated into 20 samples, they were measured using an online LC-MS system. With these powerful MS-based proteomics techniques, we were able to confidently quantify 4,259 yeast proteins (listed in a table online at doi.org/10.1039/C6AN00144K), which nearly covered the entire yeast proteome.[91, 92] Moreover, 95% of identified glycoproteins were also identified in the proteome experiments (Figure 3.11a). Due to their low abundances, seven glycoproteins were not identified in proteome analysis without efficient enrichment. The protein abundance change distribution is shown in Figure 3.11b and most protein abundances did not have marked changes. Overall, 400 proteins were up-regulated while 226 proteins were down-regulated by at least 2 fold in TM-treated yeast cells. We then clustered them separately according to biological process or pathway using the Database for Annotation, Visualization, and Integrated Discovery 6.7 (DAVID 6.7).[93] Several glycan metabolism pathways, including starch and sucrose metabolism, fructose and mannose metabolism, the pentose phosphate pathway, and glycolysis-related pathways were significantly enriched among up-regulated proteins (Figure 3.12a)). This phenomenon may be due to excess glycans present in cells as a result of protein glycosylation inhibition by TM. We have quantified the majority of the proteins involved in the canonical unfolded protein response pathway,[94] including Ero1 (YML130C), an essential oxidoreductase that produces disulfide bonds in the ER, which was up-regulated by 5.2 fold. Other related proteins, including Hrd3 (YLR207W), Gcn2 (YDR283C), and Ire1 (YHR079C), had increased abundances of 1.7, 1.6, and 1.8 fold, respectively.

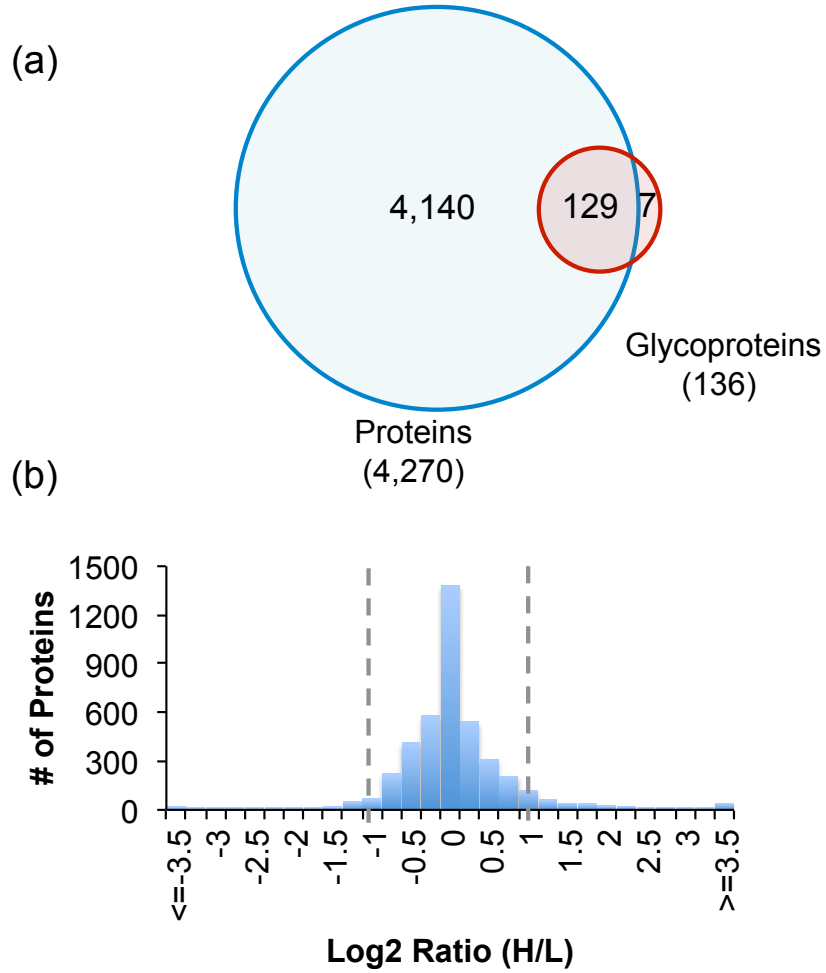


Figure 3.11 Protein identification and quantification results. (a) The overlap between proteins and glycoproteins identified in this work. (b) The ratio distribution of quantified proteins.

For down-regulated proteins, ribosome and RNA processing-related biological processes were notably enriched, which meant that protein translation was reduced. This correlates very well with previous studies in the literature.[95, 96] For example, Steffen *et al.* found that ribosomal deficiency protects yeast cells against ER stress, which was a result of many secretory proteins getting trapped in the ER due to the inhibition of their glycosylation. The treatment of the ribosomal protein gene deletion strains with TM showed significant ER stress resistance.[95] In addition, protein transportation between the Golgi and plasma membrane was also attenuated. Cell wall integrity and stress response component 4 (Wsc4) is a protein that

participates in protein transportation to the membrane, and cell wall biogenesis and degradation, and its expression was reduced to 6.6% as a result of a drug treatment. The dramatic down-regulation of this protein suggests that cell wall formation may be impacted in the TM-treated cells because protein N-glycosylation regulated protein folding and trafficking and here it was inhibited by TM. Therefore, cell wall proteins cannot be transported to the cell wall.

3.2.3.4 Site-specific glycoprotein identification

The common tag generated by PNGase F deglycosylation in heavy-oxygen water ($H_2{}^{18}O$) allowed the global and site-specific identification of protein N-glycosylation. As shown in Figure 3.10b, fragments in the tandem mass spectrum enabled us to confidently localize protein glycosylation sites. A total of 448 glycosylation sites were identified in the current experiment (listed in a table online at doi.org/10.1039/C6AN00144K). Here we assessed the confidence of site localizations with the calculation of ModScore values, which take all possible glycosylation sites in a peptide into account and uses the existing experimental fragment ions unique to each site to determine the actual glycosylation site.[39, 41, 46] For instance, two possible glycosylation sites located next to each other without adequate fragment ions to distinguish them will result in a low Modscore. A ModScore greater than 13 represents a *P* value less than 0.05, which we considered to be well-localized. Figure 3.13a shows that the majority of the glycosylation site identified in this experiment are well-localized, and 68.5% of identified sites even have a ModScore larger than 19 (corresponding to a *P* value less than 0.01).

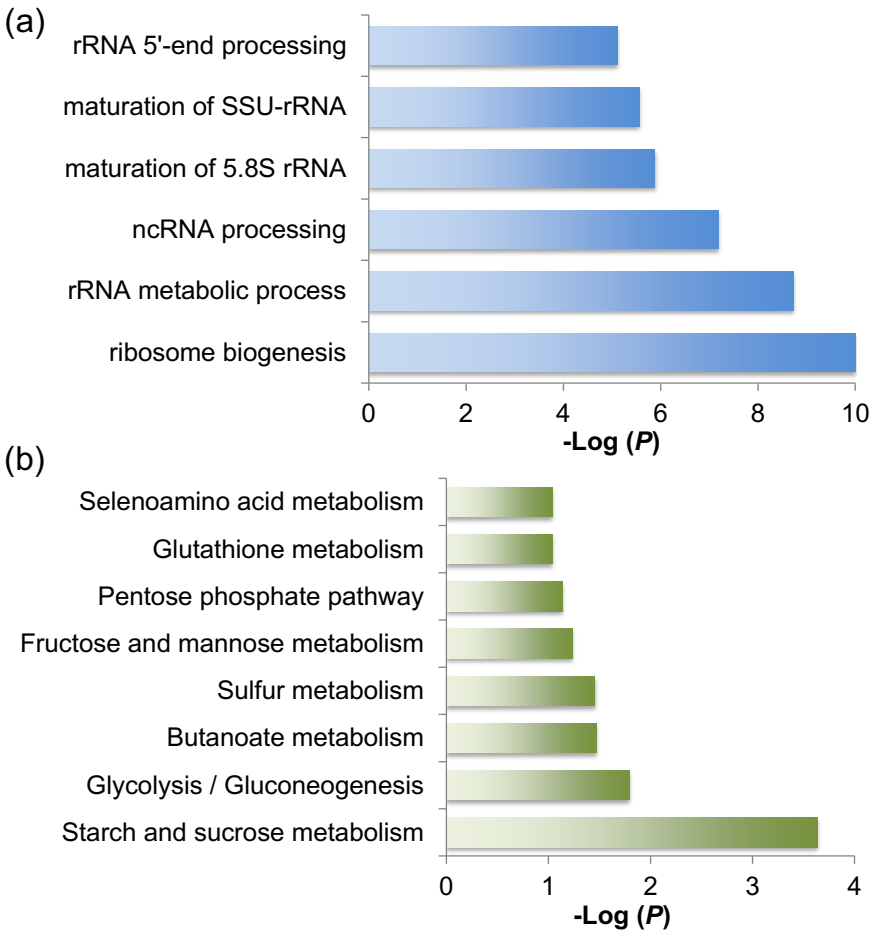


Figure 3.12 Clustering of up- and down-regulated proteins in tunicamycin-treated cells. (a) Enriched pathways for up-regulated proteins. (2) Enriched biological processes among down-regulated proteins.

Many proteins carried multiple N-glycosylation sites (Figure 3.13b) and more than 30 proteins contained at least five glycosylation sites. For example, a total of 15 possible glycosylation sites were identified from the protein Rax2 (YLR084C). Rax2 is required for the maintenance of the bipolar budding pattern, and is involved in selecting bud sites.[97] It was reported that Rax2 is a glycosylated type I membrane protein, with its long N-terminal domain in the extracellular space.[98] In TM-treated yeast cells, Rax2 was down-regulated by 2.07 fold

while its four singly glycosylated peptides were down-regulated with ratios of 0.45, 0.46, 0.51, and 0.51, respectively.

We further investigated the correlation between the number of identified glycosylation sites and the glycoprotein length (Figure 3.13c). It seems plausible that longer proteins could carry statistically more glycosylation sites which would allow a greater number of glycosylation sites to be identified. When we plotted the number of glycosylation sites as a function of the protein length, there was no significant correlation between the two.

Next we considered whether protein and glycoprotein identifications in this work were biased for highly abundant proteins. Based on the number of copies (abundances) of yeast proteins reported in the literature,[99] we plotted the abundance distribution of proteins and glycoproteins identified here with the protein abundance distribution from the literature in Figure 3.13d. The *x*-axis represents the number of protein molecules per cell, and the *y*-axis shows the percentage of proteins. Despite all three protein datasets having similar distributions over various amounts of protein molecules, we quantified a considerable amount of proteins and glycoproteins that were not quantified by the tandem affinity purification (TAP) coupled to immune-detection method in the literature.[99] This means that modern MS methods are very sensitive and can detect proteins with very low abundances. In addition, the median length of glycoproteins identified in this experiment is 581 amino acid residues, while the yeast whole proteome (http://www.yeastgenome.org/) has a median of 359 amino acid residues. This suggested that glycoproteins are generally longer than other proteins, although the number of N-glycosylation sites on each protein is not always relevant to the protein length.

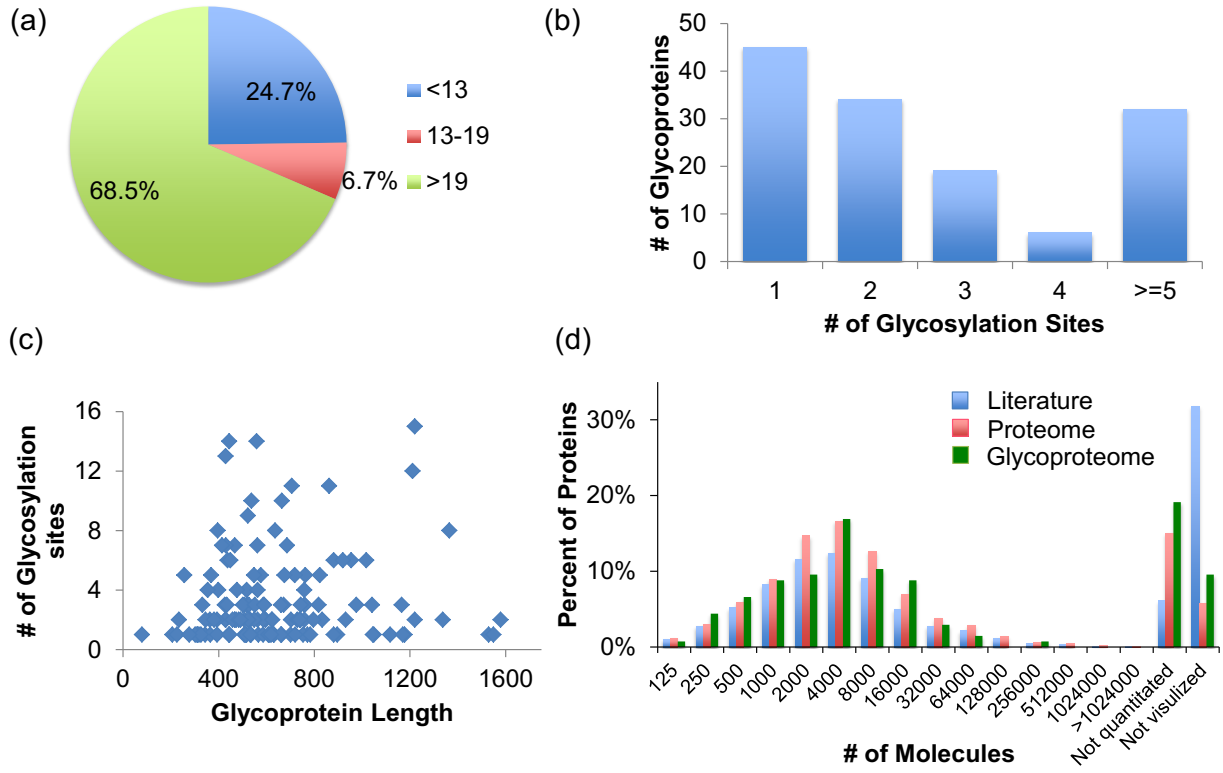


Figure 3.13 The results of site-specific N-glycosylation identification. (a) The ModScore distribution for the identified glycosylation sites. (b) The number of glycosylation sites identified in glycoproteins. (c) The correlation between the number of glycosylation sites and the length of glycoproteins. (d) The abundance distribution of proteins and glycoproteins in the literature[99] and quantified in this work.

3.2.3.5 Quantification of glycopeptides and singly-glycosylated peptides

In this work, a total of 465 unique glycopeptides were quantified, among which more than one third (162 glycopeptides) were down-regulated by more than 2 fold, while only 40 glycopeptides were up-regulated (listed in a table online at doi.org/10.1039/C6AN00144K). These results are agreeable with the known glycosylation inhibition effects of TM. The distribution of glycopeptide abundances is shown in Figure 3.14a. Glycopeptides are not expected to be up-regulated as a result of tunicamycin treatment, however this could occur because some N-acetylglucosamine-dolichol-phosphate intermediates still exist in cells for a

short period after treatment, or if the corresponding parent proteins are up-regulated. For instance, glycopeptide R.TPLVAWGAGLNK#PVHNPFPVSDN*YTENWE LSSIK#.R has an up-regulation ratio of 2.01, while the corresponding protein YKL165C were up-regulated for 3.47 fold. The regulation ratio for this peptide is determined to be 0.58 after calibration. Meanwhile, peptide K.SPVETVSDSLQFSFNGN*QTK (2.34 fold) from protein YDR055W (4.65 fold) has a ratio of 0.50 after calibration. Here we only treated cells for three hours, but more glycopeptides are anticipated to be down-regulated if cells are treated for a longer time.

Glycoproteins containing down-regulated glycopeptides were clustered according to biological processes using DAVID 6.7 (Figure 3.14b). The glycosylation, glycoprotein metabolic processes, carbohydrate processes, and cell wall organization were highly enriched. Compared to proteome analysis results, these more directly reflect the primary impact of inhibiting protein N-glycosylation by TM in yeast cells. Interestingly, several proteins containing down-regulated N-glycopeptides were related to protein O-glycosylation. A total of five glycoproteins in the current results were involved in this process, among which three were also involved in protein N-glycosylation. The other two glycoproteins, dolichyl-phosphate-mannose-protein mannosyltransferase 2 (PMT2, YAL023C) and PMT5 (YDL093W) are glycosyltransferases that specifically participate in protein O-glycosylation (especially O-linked mannosylation). The current results suggest that TM treatment could also interfere protein O-glycosylation by suppressing the N-glycosylation of important O-glycosylation transferases.

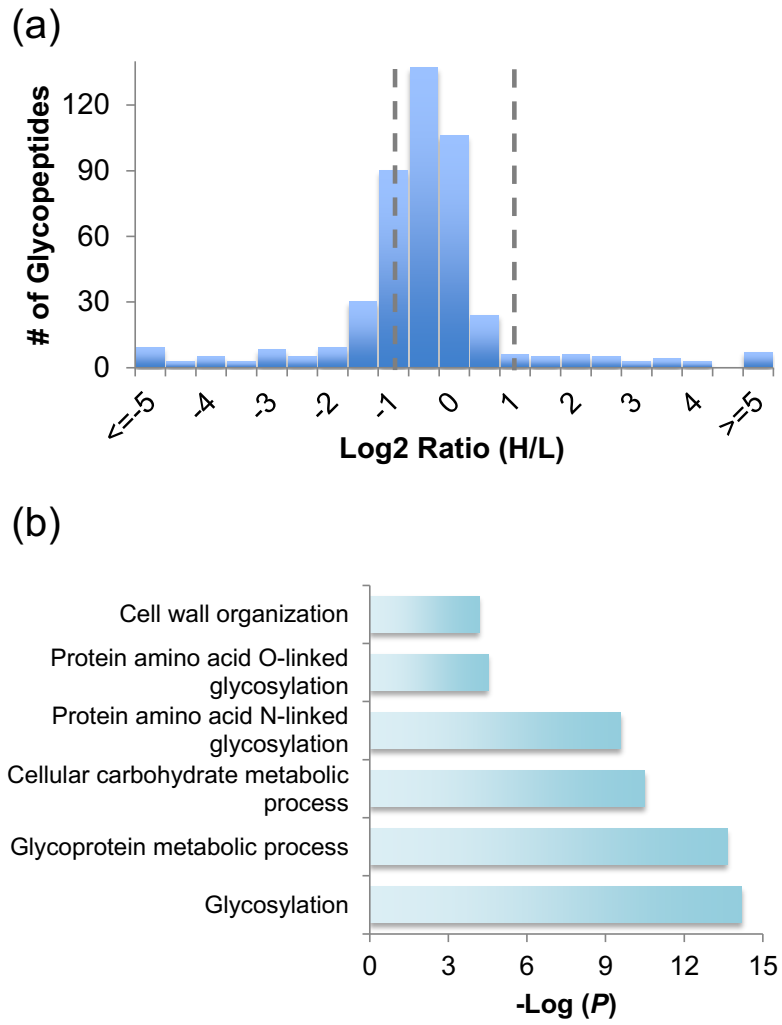


Figure 3.14 The ratio distribution of glycopeptides and glycoprotein clustering. (a) Ratio distribution of the quantified glycopeptides. (b) Clustering of the down-regulated glycoproteins according to biological processes.

Finally, we extracted all the singly-glycosylated glycopeptides with a ModScore larger than 13, and performed quantification at the glycosylation site level. The ratio distribution (Figure 3.15a) of quantified glycosylation sites is largely similar to that of quantified glycopeptides. A total of 253 sites were quantified, among which 81 were down-regulated and 18 were up-regulated. Clustering analysis also revealed that glycosylation was impacted (Figure 3.15b). The high-mannose type N-glycan biosynthesis pathway was found to be down-regulated with a *P* value of 1.24E-4; all sites quantified in this pathway were well-localized

(Table 3.2). Powerful MS-based proteomics methods allowed us to systematically and site-specifically quantify protein N-glycosylation changes in TM-treated cells, offering valuable insight into tunicamycin-cell interactions.

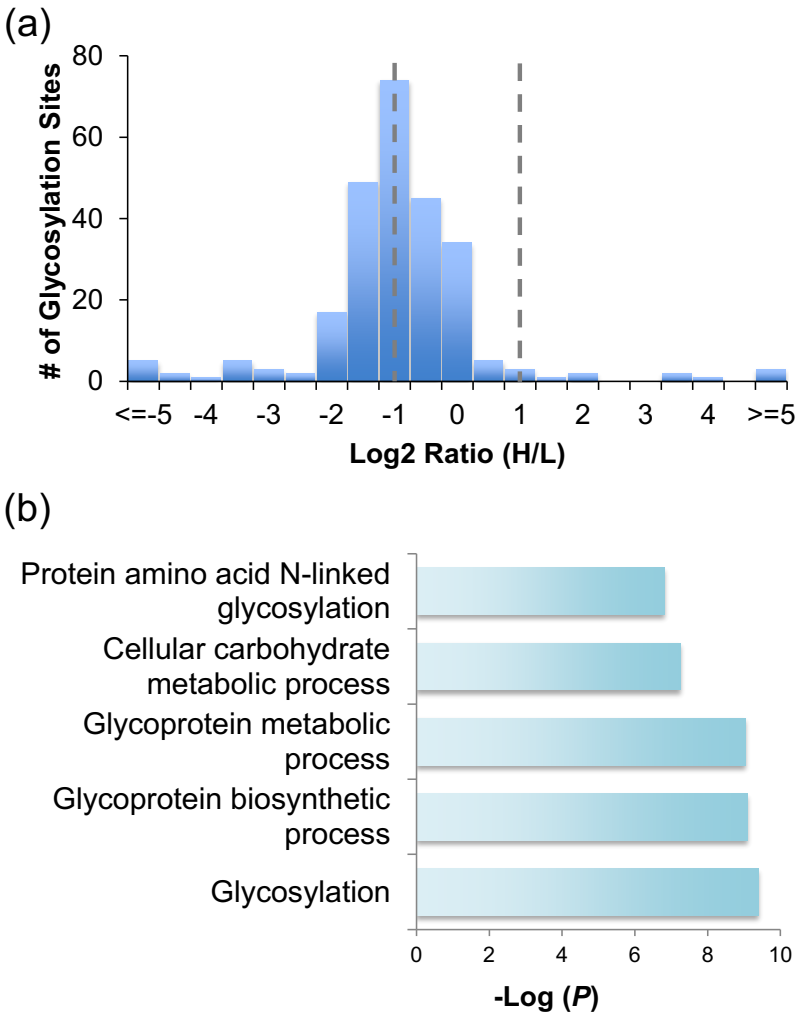


Figure 3.15 (a) Ratio distribution of the quantified glycosylation sites. (b) Clustering of the down-regulated glycosylation sites according to biological processes.

3.2.4 Conclusions

Tunicamycin has been widely used to manipulate protein N-glycosylation, but the global analysis of protein expression and N-glycosylation changes as a result of tunicamycin treatment remains unexplored. Using Baker's yeast as a model system, we systematically

investigated the protein abundance and N-glycosylation changes by powerful MS-based proteomics techniques. Through combination with quantitative proteomics, we have quantified 4,259 proteins in tunicamycin-treated yeast cells. The majority of protein abundances changed very little if at all, but nearly 10% of quantified proteins were down-regulated by >2 fold, among which proteins related to several glycan metabolism and glycolysis-related pathways were highly enriched. In addition, several proteins in the canonical unfolded protein response pathway were up-regulated because the inhibition of N-glycosylation dramatically impacts the proper folding and subsequent trafficking of some proteins.

We comprehensively quantified protein N-glycosylation changes in yeast cells induced by tunicamycin by combining boronic acid-based glycopeptide enrichment, enzymatic deglycosylation in heavy-oxygen water, and MS-based proteomics. More than one third (168) of 465 quantified unique glycopeptides were down-regulated in yeast cells with three-hour treatment. Among down-regulated glycoproteins, those related to glycosylation, glycoprotein metabolic processes, carbohydrate processes, and cell wall organization were highly enriched. The high-mannose type N-glycan biosynthesis pathway was also found to be down-regulated. For the first time, we systematically and quantitatively investigated protein expression and N-glycosylation changes in tunicamycin-treated yeast cells. These results will provide a better understanding of how cells interact with tunicamycin and how N-glycosylation is affected as a result.

Table 3.2 Down-regulated glycosylation sites involved in the high-mannose type N-glycan biosynthesis pathway (*P*=1.2E-4)

Reference	Peptide	Site	Mod Score	PPM	XCorr	H/L ratio	Annotation
YJR131W	K.YLAYLTGN*R.T	224	1000	1.02	2.16	0.01	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase (MNS1)
	R.MLGGLLSAYHLSDVLEVGN*K.T	155	1000	0.16	3.33	0.44	
YER001W	K.MFPFINN*FTTETFHEMVPK.I	254	17.0	-1.21	2.52	0.04	Alpha-1,3-mannosyltransferase (MNN1)
	K.TLN*ATFPNYDPDNFK.K	225	65.4	1.75	4.74	0.05	
	R.SPDFKPVENNYDN*STNVPQEIWFLDVSNTIHPK.W	383	38.3	-3.16	5.77	0.17	
YJL186W	K.FTDTLSGKLN*FSIPQR.E	136	1000	-0.77	3.51	0.41	Alpha-1,2-mannosyltransferase (MNN5)
YPL053C	K.SYGGN*ETTLGFMVPSYINHR.G	98	145.0	-0.2	3.76	0.48	Mannosyltransferase (KTR6)

*- glycosylation site

3.3 References

1. Varki, A. et al. Essentials of Glycobiology (2nd Edition). (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; 2008).
2. Dennis, J.W., Granovsky, M. & Warren, C.E. Protein glycosylation in development and disease. *Bioessays* **21**, 412-421 (1999).
3. Maverakis, E. et al. Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review. *J. Autoimmun.* **57**, 1-13 (2015).
4. Lehle, L., Strahl, S. & Tanner, W. Protein glycosylation, conserved from yeast to man: A model organism helps elucidate congenital human diseases. *Angew. Chem.-Int. Edit.* **45**, 6802-6818 (2006).
5. Sola, R.J. & Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci-Us* **98**, 1223-1245 (2009).
6. 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).
7. Zhu, Z.K. & Desaire, H. in Annual Review of Analytical Chemistry, Vol 8, Vol. 8. (eds. R.G. Cooks & J.E. Pemberton) 463-483 (Annual Reviews, Palo Alto; 2015).
8. Yang, Y. et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 10 (2016).
9. Wei, X., Dulberger, C. & Li, L.J. Characterization of murine brain membrane glycoproteins by detergent assisted lectin affinity chromatography. *Anal Chem* **82**, 6329-6333 (2010).
10. Breidenbach, M.A., Palaniappan, K.K., Pitcher, A.A. & Bertozzi, C.R. Mapping yeast N-glycosites with isotopically recoded glycans. *Mol Cell Proteomics* **11**, 10 (2012).
11. Khatri, K., Klein, J.A. & Zaia, J. Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Anal. Bioanal. Chem.* **409**, 607-618 (2017).
12. Loziuk, P.L., Hecht, E.S. & Muddiman, D.C. N-linked glycosite profiling and use of Skyline as a platform for characterization and relative quantification of glycans in differentiating xylem of Populus trichocarpa. *Anal. Bioanal. Chem.* **409**, 487-497 (2017).
13. Trinidad, J.C. et al. Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics* **11**, 215-229 (2012).
14. Alfaro, J.F. et al. 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.* **109**, 7280-7285 (2012).
15. Khidekel, N. et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **3**, 339-348 (2007).
16. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chem Sci* **6**, 4681-4689 (2015).
17. Banazadeh, A., Veillon, L., Wooding, K.M., Zabet-moghaddam, M. & Mechref, Y. Recent advances in mass spectrometric analysis of glycoproteins. *Electrophoresis* **38**, 162-189 (2017).

18. Gaunitz, S., Nagy, G., Pohl, N.L.B. & Noyotny, M.V. Recent advances in the analysis of complex glycoproteins. *Anal Chem* **89**, 389-413 (2017).

19. Chandler, K.B. & Costello, C.E. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis* **37**, 1407-1419 (2016).

20. Tan, Z.J. et al. Large-scale identification of core-fucosylated glycopeptide sites in pancreatic cancer serum using mass spectrometry. *J. Proteome Res.* **14**, 1968-1978 (2015).

21. Zheng, J.N., Xiao, H.P. & Wu, R.H. Specific identification of glycoproteins with the Tn antigen in human cells. *Angew. Chem. Int. Ed.*, in press, DOI: 10.1002/anie.201702191 (2017).

22. Liu, T. et al. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **4**, 2070-2080 (2005).

23. Wang, L. et al. Mapping N-Linked glycosylation sites in the secretome and whole cells of aspergillus niger using hydrazide chemistry and mass spectrometry. *J. Proteome Res.* **11**, 143-156 (2012).

24. Storr, S.J. et al. The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum. *Glycobiology* **18**, 456-462 (2008).

25. Thaysen-Andersen, M., Packer, N.H. & Schulz, B.L. Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol Cell Proteomics* **15**, 1773-1790 (2016).

26. Plomp, R., Bondt, A., de Haan, N., Rombouts, Y. & Wuhrer, M. Recent advances in clinical glycoproteomics of immunoglobulins (Igs). *Mol Cell Proteomics* **15**, 2217-2228 (2016).

27. Kurcon, T. et al. miRNA proxy approach reveals hidden functions of glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7327-7332 (2015).

28. Alvarez-Manilla, G. et al. Glycoproteomic Analysis of embryonic stem cells: identification of potential glycobiomarkers using lectin affinity chromatography of glycopeptides. *J. Proteome Res.* **9**, 2062-2075 (2010).

29. Zielinska, D.F., Gnad, F., Wisniewski, J.R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907 (2010).

30. Wollscheid, B. et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378-386 (2009).

31. Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* **12**, 561-567 (2015).

32. Sun, S.S. et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat Biotechnol* **34**, 84-88 (2016).

33. Zeng, Y., Ramya, T.N.C., Dirksen, A., Dawson, P.E. & Paulson, J.C. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* **6**, 207-209 (2009).

34. Xiao, H.P. & Wu, R.H. Global and site-specific analysis revealing unexpected and extensive protein S-GlcNAcylation in human cells. *Anal Chem* **89**, 3656-3663 (2017).

35. Wang, X.S. et al. A Novel Quantitative Mass spectrometry platform for determining protein o-glcnacylation dynamics. *Mol Cell Proteomics* **15**, 2462-2475 (2016).

36. Fierro-Monti, I. et al. A novel pulse-chase SILAC strategy measures changes in protein decay and synthesis rates induced by perturbation of proteostasis with an Hsp90 Inhibitor. *Plos One* **8** (2013).

37. Hinkson, I.V. & Elias, J.E. The dynamic state of protein turnover: It's about time. *Trends Cell Biol* **21**, 293-303 (2011).

38. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).

39. Chen, W.X., Smeekens, J.M. & Wu, R.H. A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol Cell Proteomics* **13**, 1563-1572 (2014).

40. Kaji, H. et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* **21**, 667-672 (2003).

41. 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).

42. 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).

43. 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).

44. 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).

45. Huttlin, E.L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010).

46. 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).

47. Xiao, H.P., Smeekens, J.M. & Wu, R.H. Quantification of tunicamycin-induced protein expression and N-glycosylation changes in yeast. *Analyst* **141**, 3737-3745 (2016).

48. Xiao, H.P., Tang, G.X. & Wu, R.H. Site-specific quantification of surface N-glycoproteins in statin-treated liver cells. *Anal Chem* **88**, 3324-3332 (2016).

49. Xiao, H.P. & Wu, R.H. Quantitative investigation of human cell surface N-glycoprotein dynamics. *Chem Sci* **8**, 268-277 (2017).

50. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

51. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem Sci* **7**, 1393-1400 (2016).

52. 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).

53. Orozco, M. A theoretical view of protein dynamics. *Chem Soc Rev* **43**, 5051-5066 (2014).

54. Herscovics, A. & Orlean, P. Glycoprotein-biosynthesis in Yeast. *Faseb J* **7**, 540-550 (1993).

55. Defaus, S., Gupta, P., Andreu, D. & Gutierrez-Gallego, R. Mammalian protein glycosylation - Structure *versus* function. *Analyst* **139**, 2944-2967 (2014).

56. Hsu, K.L., Pilobello, K.T. & Mahal, L.K. Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nat. Chem. Biol.* **2**, 153-157 (2006).

57. An, H.J., Kronewitter, S.R., de Leoz, M.L.A. & Lebrilla, C.B. Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **13**, 601-607 (2009).

58. Brooks, S.A. et al. Altered glycosylation of proteins in cancer: What is the potential for new anti-tumour strategies. *Anti-Cancer Agents Med. Chem.* **8**, 2-21 (2008).

59. Freeze, H.H., Chong, J.X., Bamshad, M.J. & Ng, B.G. Solving glycosylation disorders: fundamental approaches reveal complicated pathways. *Am. J. Hum. Genet.* **94**, 161-175 (2014).

60. Fuster, M.M. & Esko, J.D. The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nature Reviews Cancer* **5**, 526-542 (2005).

61. Ju, T.Z., Otto, V.I. & Cummings, R.D. The Tn antigen - structural simplicity and biological complexity. *Angew. Chem.-Int. Edit.* **50**, 1770-1791 (2011).

62. Ruhaak, L.R., Miyamoto, S. & Lebrilla, C.B. Developments in the identification of glycan biomarkers for the detection of cancer. *Mol. Cell. Proteomics* **12**, 846-855 (2013).

63. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Bba-Gen Subjects* **1426**, 239-257 (1999).

64. Kukuruzinska, M.A., Bergh, M.L.E. & Jackson, B.J. Protein glycosylation in yeast. *Annu Rev Biochem* **56**, 915-944 (1987).

65. Wildt, S. & Gerngross, T.U. The humanization of N-glycosylation pathways in yeast. *Nat Rev Microbiol* **3**, 119-128 (2005).

66. Gemmill, T.R. & Trimble, R.B. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Bba-Gen Subjects* **1426**, 227-237 (1999).

67. Lehle, L. & Tanner, W. Specific site of tunicamycin inhibition in formation of dolichol-bound N-acetylglucosamine derivatives. *Febs Lett* **71**, 167-170 (1976).

68. Takatsuk, A., Arima, K. & Tamura, G. Tunicamycin, a new antibiotic .1. Isolation and characterization of tunicamycin. *J Antibiot* **24**, 215-223 (1971).

69. Takatsuk, A. & Tamura, G. Tunicamycin, a New Antibiotic .2. Some biological properties of antiviral activity of tunicamycin. *J Antibiot* **24**, 224-231 (1971).

70. Altelaar, A.F.M., Munoz, J. & Heck, A.J.R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14**, 35-48 (2013).

71. Wang, L.N. et al. Time-resolved proteomic visualization of dendrimer cellular entry and trafficking. *J. Am. Chem. Soc.* **137**, 12772-12775 (2015).

72. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* **20**, 301-305 (2002).

73. Pankow, S. et al. Delta F508 CFTR interactome remodelling promotes rescue of cystic fibrosis. *Nature* **528**, 510-516 (2015).

74. Loo, R.R.O. & Loo, J.A. Matrix-assisted laser desorption/ionization-mass spectrometry of hydrophobic proteins in mixtures using formic acid, perfluorooctanoic acid, and sorbitol. *Anal. Chem.* **79**, 1115-1125 (2007).

75. Wuhrer, M., Catalina, M.I., Deelder, A.M. & Hokke, C.H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B* **849**, 115-128 (2007).

76. Lee, A.E., Castaneda, C.A., Wang, Y., Fushman, D. & Fenselau, C. Preparing to read the ubiquitin code: a middle-out strategy for characterization of all lysine-linked diubiquitins. *J. Mass Spectrom.* **49**, 1272-1278 (2014).

77. Tran, J.C. et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254-U141 (2011).

78. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chemical Science*, **7**, 1393-1400 (2016).

79. Abelin, J.G. et al. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat. Protoc.* **10**, 1308-1318 (2015).

80. Yang, C.X., Zhong, X.F. & Li, L.J. Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics. *Electrophoresis* **35**, 3418-3429 (2014).

81. Hirschey, M.D. & Zhao, Y.M. Metabolic Regulation by lysine malonylation, succinylation, and glutarylation. *Mol. Cell. Proteomics* **14**, 2308-2315 (2015).

82. Richards, A.L., Merrill, A.E. & Coon, J.J. Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* **24**, 11-17 (2015).

83. Xiao, H.P., Tang, G.X. & Wu, R.H. Site-specific quantification of surface N-glycoproteins in statin-treated liver cells. *Analytical Chemistry* **88**, DOI: 10.1021/acs.analchem.1025b04871, in press (2016).

84. Smeekens, J.M., Chen, W.X. & Wu, R.H. Mass spectrometric analysis of the cell surface N-glycoproteome by combining metabolic labeling and click chemistry. *J. Am. Soc. Mass Spectrom.* **26**, 604-614 (2015).

85. Xiao, H.P., Chen, W.X., Tang, G.X., Smeekens, J.M. & Wu, R.H. Systematic investigation of cellular response and pleiotropic effects in atorvastatin-treated liver cells by MS-based proteomics. *J. Proteome Res.* **14**, 1600-1611 (2015).

86. 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).

87. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chemical Science* **6**, 4681-4689 (2015).

88. Kim, W. et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **44**, 325-340 (2011).

89. Hammond, C., Braakman, I. & Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality-control. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 913-917 (1994).

90. Sun, S.S. & Zhang, H. Identification and validation of atypical N-glycosylation sites. *Anal. Chem.* **87**, 11948-11951 (2015).

91. de Godoy, L.M.F. et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1254 (2008).

92. Wu, R.H. et al. Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. *Mol. Cell. Proteomics* **10**, 10.1074/mcp.M1111.009654 (2011).

93. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

94. Hotamisligil, G.S. Endoplasmic reticulum stress and atherosclerosis. *Nat Med* **16**, 396-399 (2010).

95. Steffen, K.K. et al. Ribosome deficiency protects against er stress in saccharomyces cerevisiae. *Genetics* **191**, 107-118 (2012).

96. Horigome, C., Okada, T., Matsuki, K. & Mizuta, K. A ribosome assembly factor Ebp2p, the yeast homolog of EBNA1-binding protein 2, is involved in the secretory response. *Biosci Biotech Bioch* **72**, 1080-1086 (2008).

97. Chen, T. et al. Multigenerational cortical inheritance of the Rax2 protein in orienting polarity and division in yeast. *Science* **290**, 1975-1978 (2000).

98. Kang, P.J., Angerman, E., Nakashima, K., Pringle, J.R. & Park, H.O. Interactions among Rax1p, Rax2p, marking cortical sites for bipolar yeast. *Mol Biol Cell* **15**, 5145-5157 (2004).

99. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature* **425**, 737-741 (2003).

CHAPTER 4. IDENTIFICATION AND QUANTIFICATION OF THE CELL-SURFACE N-GLYCOPROTEINS

Partially adapted with permission from American Chemical Society

Xiao, H. P., Tang, G. X., and Wu, R. H. Site-Specific Quantification of Surface N-Glycoproteins in Statin-Treated Liver Cells. *Analytical Chemistry,* 2016, 88, 3324-3332.

Partially adapted with permission from The Royal Society of Chemistry

Xiao, H. P., and Wu, R. H. Quantitative Investigation of Human Cell Surface N-Glycoprotein Dynamics. *Chemical Science,* 2017, 8, 268-277.

4.1 Analysis of Cell-Surface N-Glycoproteome and Site-specific Quantification of Surface N-glycoproteins in Statin-treated Liver Cells

4.1.1 Introduction

Glycosylation is one of the most important protein modifications and is essential for cell survival.[1] There are two major types of protein glycosylation: N-linked glycosylation and O-linked glycosylation. For O-glycosylation, glycans are bound to the side chains of serine and threonine, while N-glycosylation involves glycans covalently attached to the side chain of asparagine. In eukaryotic cells, there is machinery in the endoplasmic reticulum (ER) responsible for attaching the initial glycan block ($(GlcNAc)_2(Mannose)_9(Glucose)_3$) to nascent peptides. It is well known that N-glycosylation plays determinant roles in protein folding and

trafficking, and N-glycosylated proteins are especially important in regulating extracellular activities, including cell-cell communication and cell-matrix interactions.[2] Aberrant protein N-glycosylation is frequently related to human disease,[3] including Alzheimer's disease (AD),[4] cancer,[5] and infectious diseases.[6]

Half a century ago, early mammalian cell morphology studies discovered abundant carbohydrates on the external surface of the cell membrane.[7, 8] To date, numerous research results have indicated that the majority of cell surface proteins are glycosylated.[9, 10] Despite the number of glycoproteins located on the cell surface and their importance in biological functions, the global analysis of surface glycoproteins is extraordinarily challenging.[11, 12] Modern mass spectrometry (MS)-based proteomics techniques provide the capacity to comprehensively analyze protein modifications.[13-23] These methods can be employed to systematically identify modified proteins, localize the modification sites, and quantify their abundance changes.[24-28] However, the heterogeneity of glycans and low abundance of many glycoproteins make the global analysis of glycoproteins extraordinarily difficult.[29] It is even more challenging to specifically and comprehensively analyze N-glycoproteins only on the cell surface because it requires the selective separation and enrichment of surface N-glycoproteins.

Statins have been widely used to lower cholesterol levels in patients by inhibiting 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCR),[30] an enzyme in the upstream portion of the mevalonate pathway. Besides cholesterol, the synthesis of many intermediate and end products in this pathway, including ubiquinone and dolichol, are significantly affected by the inhibition of this enzyme.[31] Dolichol plays an essential role in protein *N*-glycosylation, and functions as a membrane anchor for the formation of a precursor oligosaccharide.[32] The effect of statin on surface protein glycosylation is still unknown which may contribute to the pleiotropic effects of statins. The systematic and quantitative analysis of surface glycoproteins in statin-treated cells will potentially shed light on the molecular mechanisms behind the

pleiotropic effects of statins, which will allow patients to receive the full benefits of this medicine.

In this work, we systematically evaluated metabolic labeling with three sugar analogs, *i.e.* GalNAz, GlcNAz and ManNAz, for the identification of cell surface N-glycoproteins by combining copper-free click chemistry and MS-based proteomics. The parallel experiments showed that GalNAz labeling resulted in the greatest number of protein N-glycosylation sites identified, while GlcNAz resulted in the smallest number of protein N-glycosylation sites. Thus, GalNAz labeling was employed for the global quantification of surface glycoproteins in HepG2 liver cells treated with statin. Systematic and quantitative analysis of surface proteins in statin-treated cells clearly demonstrated that many glycosylation sites were down-regulated compared to untreated cells. This method offers a means to globally, site-specifically and quantitatively study protein *N*-glycosylation on the cell surface.

4.1.2 Experimental section

4.1.2.1 Cell culture and metabolic labeling

HepG2 (C3A) cells (Hep G2 [HEPG2] (ATCC® HB-8065™)) were grown in a humidified incubator at 37 °C and 5.0% CO_2 in Dulbecco's modified eagle's medium (DMEM) (Sigma-Aldrich) containing low glucose and 10% fetal bovine serum (FBS) (Thermo). For the glycoprotein identification experiments, when cells reached about 40% confluency, 100 μM GalNAz, GlcNAz or ManNAz (Click Chemistry Tools) was added to the media. Cells were further cultured for 24 h. In each case, duplicate biological experiments were performed.

For the quantification experiment, “heavy” and “light” stable isotope labeling by amino acids in cell culture (SILAC) (Sigma-Aldrich) media were freshly prepared by adding 0.146 g/L $^{13}C_6{}^{15}N_2$ L-lysine (Lys-8) and 0.84 g/L $^{13}C_6$ L-arginine (Arg-6) (Cambridge Isotopes Inc.) or the corresponding non-labeled lysine (Lys-0) and arginine (Arg-0) to DMEM and

supplemented with 10% dialyzed FBS (Corning). Cells were cultured for about six generations before the atorvastatin treatment. 40 mM atorvastatin (Cayman Chemical) stock solution was prepared in DMSO (Sigma-Aldrich). About 2×10^7 cells were treated with 15 μM atorvastatin in serum-free heavy medium for 48 h. A similar number of light cells were treated by DMSO in serum-free light medium as a control. 100 μM GalNAz was added in after 24 h of atorvastatin or DMSO treatment.

4.1.2.2 In-flask copper-free click reaction, cell lysis and protein digestion

Cells were washed twice with phosphate buffered saline (PBS) before 100 μM dibenzocyclooctyne (DBCO)-sulfo-biotin in PBS was added into the cell culture flasks. Surface glycoproteins were tagged with biotin through the specific click reaction between DBCO and the azido group in the sugar analogs under physiological conditions[33-35]. Cells were incubated for 1 h with gentle agitation at 37 °C, then harvested by scraping in PBS. For the quantification experiments, heavy and light cells were equally combined based on the protein ratio of 1:1 from a trial run. The cell mixtures were pelleted by centrifugation at 500 g for 3 minutes and washed twice with cold PBS. Cytosol proteins were removed by incubating in a buffer containing 150 mM NaCl, 50 mM HEPES pH=7.6, 25 μg/mL digitonin, and 1 tablet/ 10 mL protease inhibitor (complete mini, EDTA-free, Roche) on ice for 10 minutes. After incubation, samples were centrifuged at 2000 g for another 10 minutes. Cell pellets were washed with the previous buffer, then lysed through end-over-end rotation at 4 ºC for 45 minutes in lysis buffer (50 mM HEPES pH=7.6, 150 mM NaCl, 0.5% SDC, 10 units/mL benzonase and 1 tablet/10 mL protease inhibitor). Lysates were centrifuged, and the resulting supernatant was transferred to new tubes. Proteins were subjected to disulfide reduction with 5 mM DTT (56 ºC, 25 minutes) and alkylation with 14 mM iodoacetamide (RT, 20 minutes in the dark). Detergent was removed by methanol-chloroform protein precipitation. The purified

proteins were digested with 10 ng/μL Lys-C (Wako) in 50 mM HEPES pH 8.6, 1.6 M urea, 5% ACN at 31 ºC for 16 hours, followed by further digestion with 8 ng/uL Trypsin (Promega) at 37 ºC for 4 hours.

4.1.2.3 Glycopeptide separation and enrichment

Digestion mixtures were acidified by the addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, clarified by centrifugation and desalted using a tC18 Sep-Pak cartridge (Waters). Purified peptides were dried and then enriched with NeutrAvidin beads (Thermo) at 37 °C for 30 min. The samples were transferred to spin columns, followed by thoroughly washing according to the manufacturer's protocol. Peptides were then eluted from the beads three times by 2 min incubations with 200 μL of 8 M guanidine-HCl (pH = 1.5) at 56 °C. Eluates were combined, desalted using tC18 Sep-Pak cartridge and lyophilized overnight. Completely dry peptides were deglycosylated with eight units of peptide-*N*-glycosidase F (PNGase F, Sigma-Aldrich) in 40 μL buffer containing 50 mM NH_4HCO_3 (pH=9) in heavy-oxygen water ($H_2{}^{18}O$) for 3 h at 37 °C. The reaction was quenched by adding formic acid (FA) to a final concentration of 1%. Peptides were further purified via stage tip and separated into 3 fractions using 20%, 50% and 80% ACN containing 1% HOAc.

4.1.2.4 LC-MS/MS analysis

Purified and dried peptide samples were each dissolved in a 10 μL of 5% ACN and 4% FA, and 4 μL of the resulting solutions were loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 μm, 200 Å, 100 μm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). Peptides were separated by reversed-phase chromatography using an

UltiMate 3000 binary pump with a 110 min gradient of 3-25%, 8-38%, 10-50% ACN (with 0.125% FA) for the three fractions, respectively. Peptides were detected with a data-dependent Top20 method[36] in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 20 MS/MS in the LTQ for the most intense ions. The selected ions were excluded from further analysis for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

4.1.2.5 Database searching and data filtering

All MS^2 spectra were converted into an mzXML format, and then searched using the SEQUEST algorithm (version 28).[37] Spectra were matched against a database containing sequences of all proteins in the UniProt Human (*Homo sapiens*) database (downloaded in February 2014). The following parameters were used during the search: 20 ppm precursor mass tolerance; 1.0 Da product ion mass tolerance; fully digested with trypsin; up to three missed cleavages; fixed modifications: carbamidomethylation of cysteine (+57.0214); variable modifications: oxidation of methionine (+15.9949), O^{18} tag of asparagine (+2.9883), heavy lysine (+8.0142) and heavy arginine (+6.0201). False discovery rates (FDR) of peptide and protein identifications were evaluated and controlled by the target-decoy method.[38] Each protein sequence was listed in both forward and reversed orders. Linear discriminant analysis (LDA), which is similar to other methods in the literature,[39] was used to control the quality of peptide identifications using parameters such as XCorr, precursor mass error, and charge state.[40] Peptides fewer than seven amino acid residues in length were deleted. Furthermore, peptide spectral matches were filtered to a 1% FDR. The dataset was restricted to glycopeptides when determining FDRs for glycopeptide identification.[41] Furthermore, an additional protein-

level filter was applied in each dataset to reduce the protein-level FDRs (<1%) for glycoproteins. Consequently the FDRs at the peptide level were much less than 1%.

4.1.2.6 Glycosylation site localization and peptide quantification

We assigned and measured the confidence of glycosylation site localizations by calculating their ModScores, which applies a probabilistic algorithm[41] that considers all possible glycosylation sites in a peptide and uses the presence of experimental fragment ions unique to each site. Sites with a ModScore > 13 ($P < 0.05$) were considered confidently localized. For peptide quantification, we required an S/N value >3 for both heavy and light peptides. If the S/N value of a certain heavy peptide was less than 3, then that of the corresponding light peptide was required to be greater than 5, and *vice versa*. If the same glycopeptide was quantified several times, the median value was used as the glycopeptide abundance change. Glycosylation site quantification had the following criteria: first, the quantified glycopeptide contain only a single glycosylation site; second, the site be well-localized with a ModScore >13. If multiple unique singly glycosylated peptides containing the same glycosylation site were identified, the ratio of the glycosylation sites was the median value of these glycopeptide ratios.

4.1.3 Results and discussion

4.1.3.1 Metabolic labeling, surface glycoprotein enrichment and MS analysis

In recent years, unnatural sugars containing a bio-orthogonal group have been used to label glycosylated proteins.[9, 42] Glycoproteins bearing the biologically inert azido or alkyne group can be bound to a fluorescence probe to visualize them. This metabolic labeling can also be used to selectively enrich glycoproteins based on the unique bio-orthogonal group. Several sugar analogs, including GlcNAz, GalNAz and ManNaz (structures are in Figure 4.1b), have been reported to label cells.[9, 43, 44] Here, in parallel experiments, we labeled cells using each of

the three sugar analogs,, and evaluated their effectiveness for the global and site-specific analysis of N-glycoproteins on the cell surface in combination with MS-based proteomics.

After cells were cultured in low glucose media containing each of these three sugar analogs, surface glycoproteins containing the azido functional group on living cells were selectively bound to DBCO-sulfo-biotin *via* in-flask copper-free click chemistry under physiological conditions for 1 h, as shown in Figure 4.1a. Because the hydrophilic DBCO-sulfo-biotin cannot penetrate the cell plasma membrane, only glycoproteins located on the cell surface were tagged under mild conditions. After cell lysis and protein digestion, surface glycoproteins tagged with a biotin group allowed further enrichment to be performed based on strong and specific interactions between biotin in glycopeptides and NeutrAvidin, which was conjugated to beads. Non-modified peptides were removed by washing the beads several times. Non-specific binding is a drawback of the streptavidin enrichment method, however, the enrichment took place at the peptide level, which increased specificity compared to protein level enrichment.

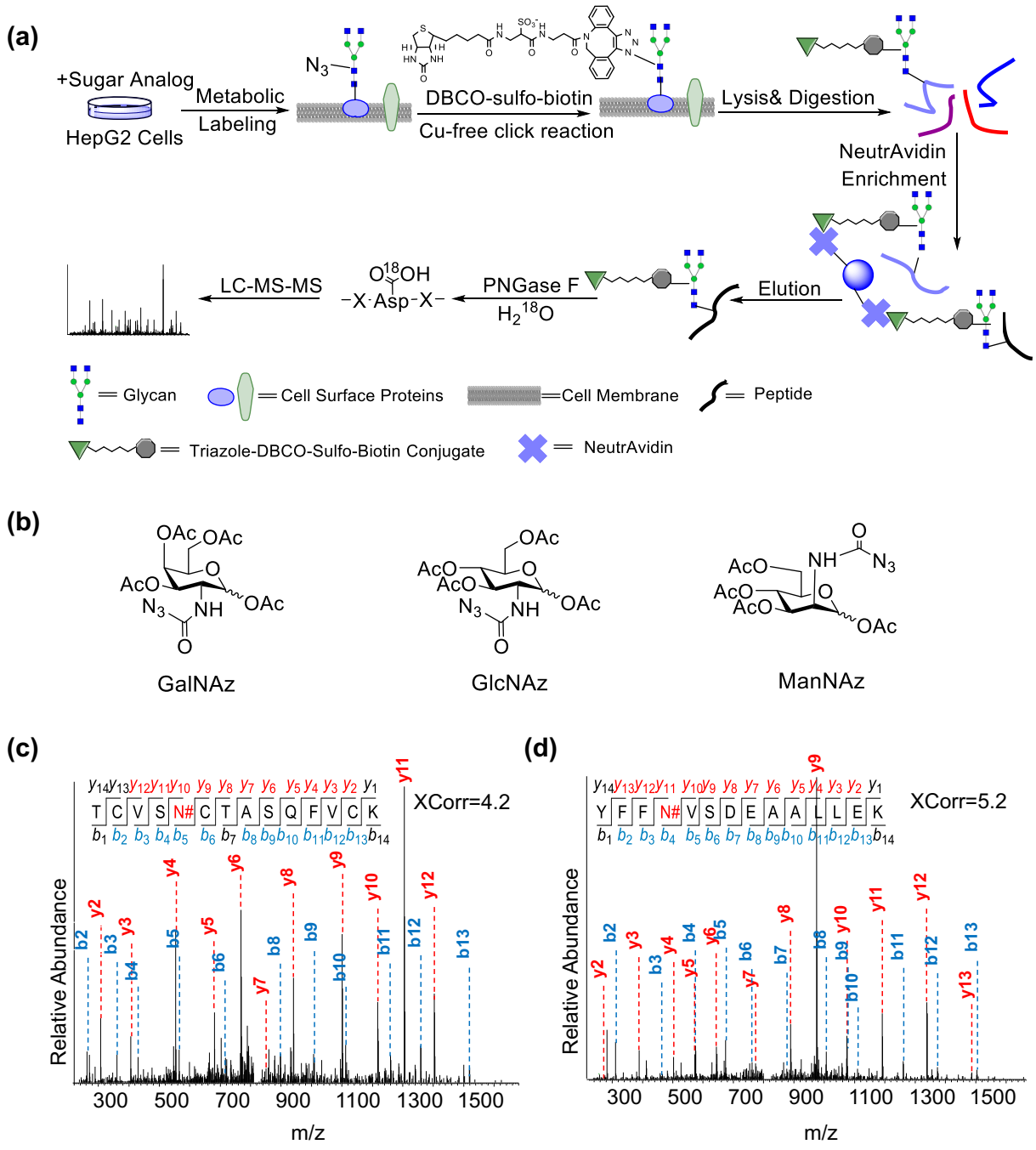


Figure 4.1 (a) Experimental procedure for the global analysis of the N-glycoproteome on the cell surface. (b) The structures of three sugar analogs used: GalNAz, GlcNAz and ManNAz. (c) A sample tandem mass spectrum of the peptide TCVSN#CTASQFVCK from LRP1. (d) Another sample MS^2 of YFFN#VSDEAALLEK from ITGA2 (# denotes the glycosylation site).

In order to generate a common tag for MS analysis, enriched peptides were treated with PNGase F in $H_2{}^{18}O$ to remove N-glycans, which converted asparagine (Asn) to heavy-oxygen aspartic acid (Asp) and created a mass shift of +2.9883 Da for glycosylation site

identification.[45, 46] This strategy was similar to a previous method for proteolytic stable isotope labeling, in which heavy-oxygen water was used to label digested peptides with trypsin and individual proteins from two proteome samples were quantitatively analyzed.[47] In this case, heavy oxygen on Asp enabled us to distinguish authentic N-glycosylation sites from those caused by deamidation on non-glycosylated asparagines *in vitro* and *in vivo*. Deamidation on non-glycosylated asparagines could also occur during PNGase F treatment, which may result in false positive identifications, which is why we ran the reaction for 3 h to minimize false positive identification. Control experiments showed that the effect of uncontrolled deamidation within the three-hour PNGase F treatment was nearly negligible, which is described in more detail below. In addition, after glycopeptide enrichment, the presence of non-glycosylated peptides was significantly decreased, therefore the chance of any deamidation from non-glycosylated peptides was dramatically reduced. Overall, the PNGase F treatment in heavy oxygen water increased the confidence of glycopeptide identification.

An Orbitrap mass spectrometer with high resolution and mass accuracy provides the capability to confidently identify glycopeptides. For example, two tandem mass spectra for two glycopeptides in cluster of differentiation (CD) proteins, which are very important for differentiation and classification of cells, are shown in Figure 4.1c and d. The glycopeptide TCVSN#CTASQFVCK (# represents the glycosylation site) from LRP1 (CD91), prolow-density lipoprotein receptor-related protein 1, was identified with an XCorr of 4.2 and a mass accuracy of -0.47 ppm. LRP1 is a single-pass type I membrane protein and is involved in endocytosis and in phagocytosis of apoptotic cells. It may modulate cellular events, such as kinase-dependent intracellular signaling, neuronal calcium signaling, and neurotransmission. As shown in Figure 4.1d, YFFN#VSDEAALLEK was also very confidently identified with an XCorr of 4.6 and a mass accuracy of 0.13 ppm. This peptide is from the protein ITGA2, integrin alpha-2, which is a receptor for laminin, collagen, collagen C-propeptides, fibronectin and E-

cadherin. It is an extremely important surface protein that regulates cell adhesion, cell-matrix interactions and host-virus interactions. In this work, we have confidently identified several glycosylation sites, *i.e.*, N105, N112, N343, N432, N1057, and N1074 in ITGA2.

4.1.3.2 Evaluation of glycopeptides and glycosylation Sites Identified in cells labeled with different sugar analogs

The procedure for our parallel experiments differed only in the sugar analogs, which allowed us to objectively evaluate the effectiveness of three sugar analogs for global surface glycoprotein analysis. The number of identified glycosylation sites and glycoproteins and their overlap between biological duplicate experiments using each of the three sugar analogs are displayed in Figure 4.2. Theoretically, each sugar analog labels a different group of glycoproteins based on glycan structure and the enzymes responsible for glycan synthesis, therefore the results of from these three labeling experiments are expected to be different. Overall, 590 glycosylation sites were identified on 274 proteins in the GalNAz labeling experiments (listed in a table online at doi.org/10.1021/acs.analchem.5b04871), including 261 proteins (95.3%) which were either secreted, located on the cell membrane or exported by extracellular vesicular exosomes based on the information from Uniprot (www.uniprot.org). Meanwhile, 446 glycosylation sites on 219 proteins, and 117 sites on 91 proteins were identified in the duplicate ManNAz (listed in a table online at doi.org/10.1021/acs.analchem.5b04871) and GlcNAz (listed in a table online at 10.1021/acs.analchem.5b04871) labeling experiments, respectively.

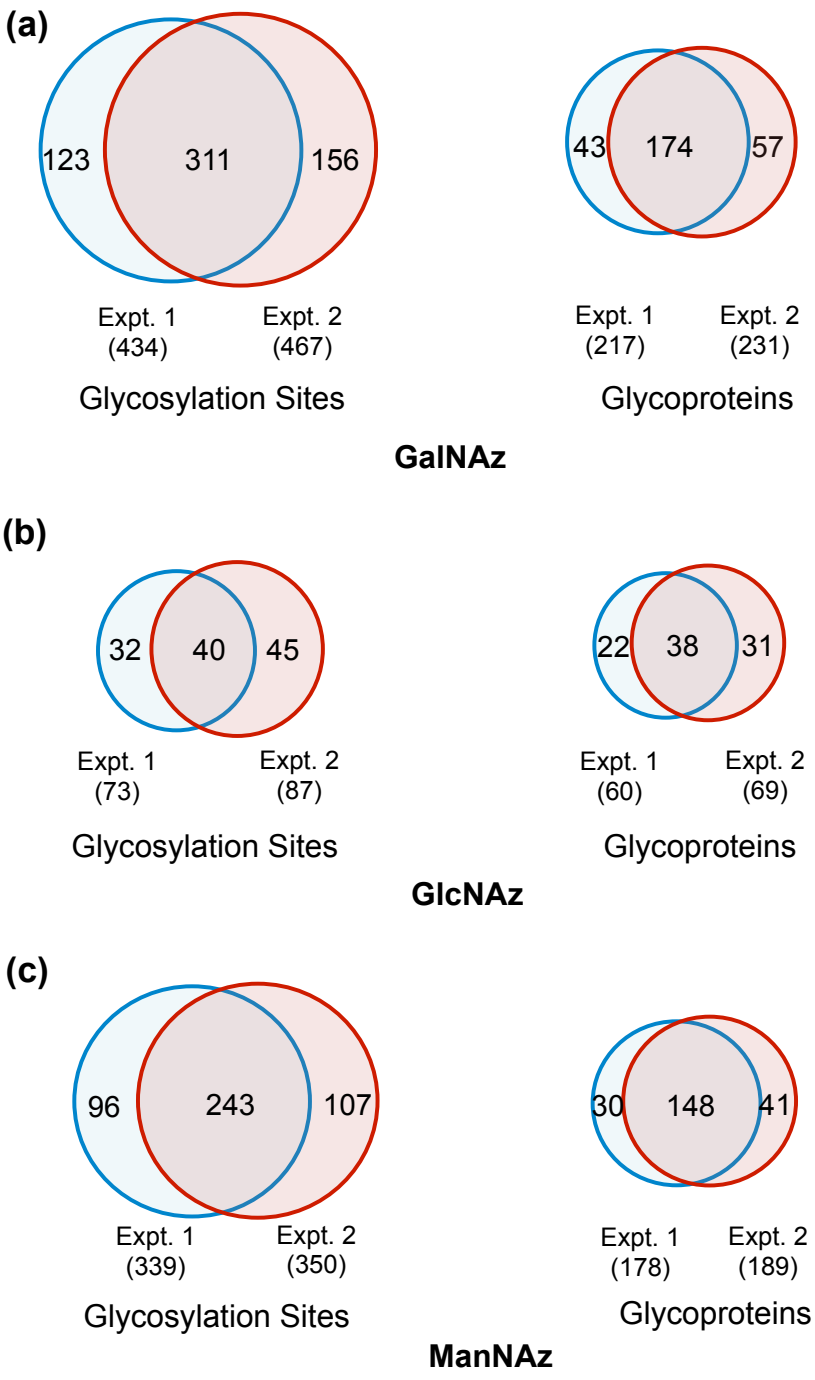


Figure 4.2 Reproducibility assessment in duplicate labeling experiments of (a) GalNAz, (b) GlcNAz, and (c) ManNAz.

The GlcNAz labeling covered the fewest number of glycosylation sites and glycoproteins, which corresponds very well to previous work showing better incorporation of GalNAz or ManNAz over GlcNAz.[48] Among 434 and 467 N-glycosylation sites identified in each GalNAz labeling experiment, 311 sites were common to both experiments. At the

glycoprotein level, as expected, the overlap was even higher. In the two experiments, 217 and 231 glycoproteins were identified, and the number of overlapped proteins was 174. Considering the large-scale analysis, this level of overlap is within a reasonable range. Since we ran biological duplicate experiments for each sugar analog, the inconsistencies between duplicates could be due to the sample differences, the dynamic nature of protein glycosylation, sample preparation (sample loss) or false positive identifications. The comparison of surface glycosylation sites and glycoproteins identified in the three parallel experiments using different sugar analogs is displayed in Figure 4.3. The majority of the glycosylation sites and glycoproteins identified in ManNAz labeling experiments (335 of 446 sites, 176 of 219 proteins) and GlcNAz labeling experiments (90 of 119 sites, 70 of 91 proteins) were also identified in the GalNAz experiments, demonstrating the highest coverage of surface glycosylation sites and glycoproteins. Based on these results, GalNAz was employed for the quantification of surface proteins in statin-treated cells in order to obtain higher glycoprotein coverage, as described below. Overall, 725 cell-surface glycosylation sites on 337 glycoproteins were identified combining all these experimental results.

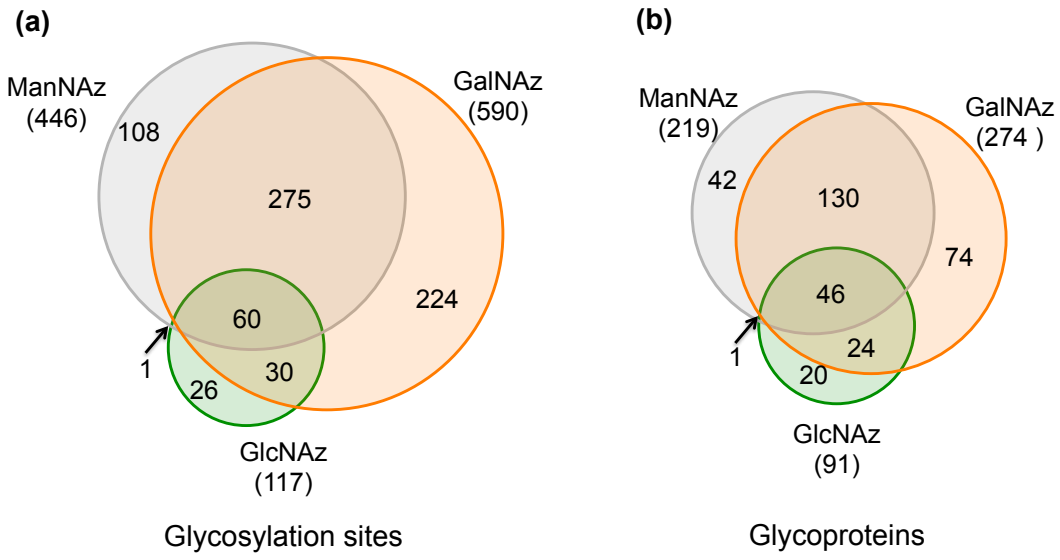


Figure 4.3 Comparison of (a) surface N-glycosylation sites, and (b) N-glycoproteins identified in GalNAz, GlcNAz and ManNAz labeling experiments.

In order to verify that the results were reliable, we designed two controls to run in parallel with the GalNAz labeling experiment. The experimental procedure for the first control was consistent with the GalNAz labeled experiment, except the click reaction was omitted; the second control omitted the PNGase F deglycosylation reaction. We identified 886 unique glycopeptides in the GalNAz labeling experiment, and only 20 glycopeptides in the first control. These 20 unique glycopeptides may have resulted from non-specific binding of the NeutrAvidin enrichment, non-glycosylated Asn deamidation in heavy-oxygen water during PNGase F treatment, or false positive identification of glycopeptides. However, any non-glycosylated Asn deamidation before or after the PNGase F treatment would result in a mass difference of 0.9840 Da, not 2.9883 Da, so they would be easily distinguishable during data analysis. Only non-glycosylated Asn deamidation within the three-hour PNGase F treatment in heavy-oxygen water would contribute to false positive identification. In the second control experiment without PNGase F, only 7 unique glycopeptides were identified, likely due to the deamination of free Asn. This is less than 1% compared to the 886 unique glycopeptides identified in the parallel GalNAz labeling experiment, which indicates that the effect of non-glycosylated Asn deamidation within the three hour reaction was nearly negligible. These control experiments clearly verified the reliability of the current results.

4.1.3.3 Clustering of surface N-glycoproteins identified in GalNAz labeling experiments

Most of the identified glycoproteins contain a single glycosylation site. There were also some proteins with more than ten sites; for example, 21 N-glycosylation sites were identified on LRP1. The clinical importance of LRP1 in Alzheimer's disease and cardiovascular disease also brings extensive attention to this protein. Glycosylation may stabilize this receptor-related protein, and also differentiates the protein's functions in different tissues.[49]

In order to further evaluate the specificity of our method, the identified glycoproteins in the GalNAz labeling experiments were clustered according to molecular function and biological process using the Database for Annotation, Visualization, and Integrated Discovery 6.7 (DAVID 6.7).[50] We investigated the molecular functions of the identified glycoproteins and the biological processes they are involved in. The molecular functions with the highest level of enrichment were receptor activity, signal transducer activity, and binding, with remarkably low *P* values (Figure 4.4). Among biological processes, cell adhesion was prominently enriched with a *P* value of 3.6E-20 and 48 proteins involved. Integrin-mediated signaling pathway and cell motion were also notably enriched with *P* values of 1.8E-10 and 4.1E-10, respectively. These are consistent with the well-known molecular functions and biological processes of surface proteins, which demonstrated that the current method for surface glycoprotein identification is effective.

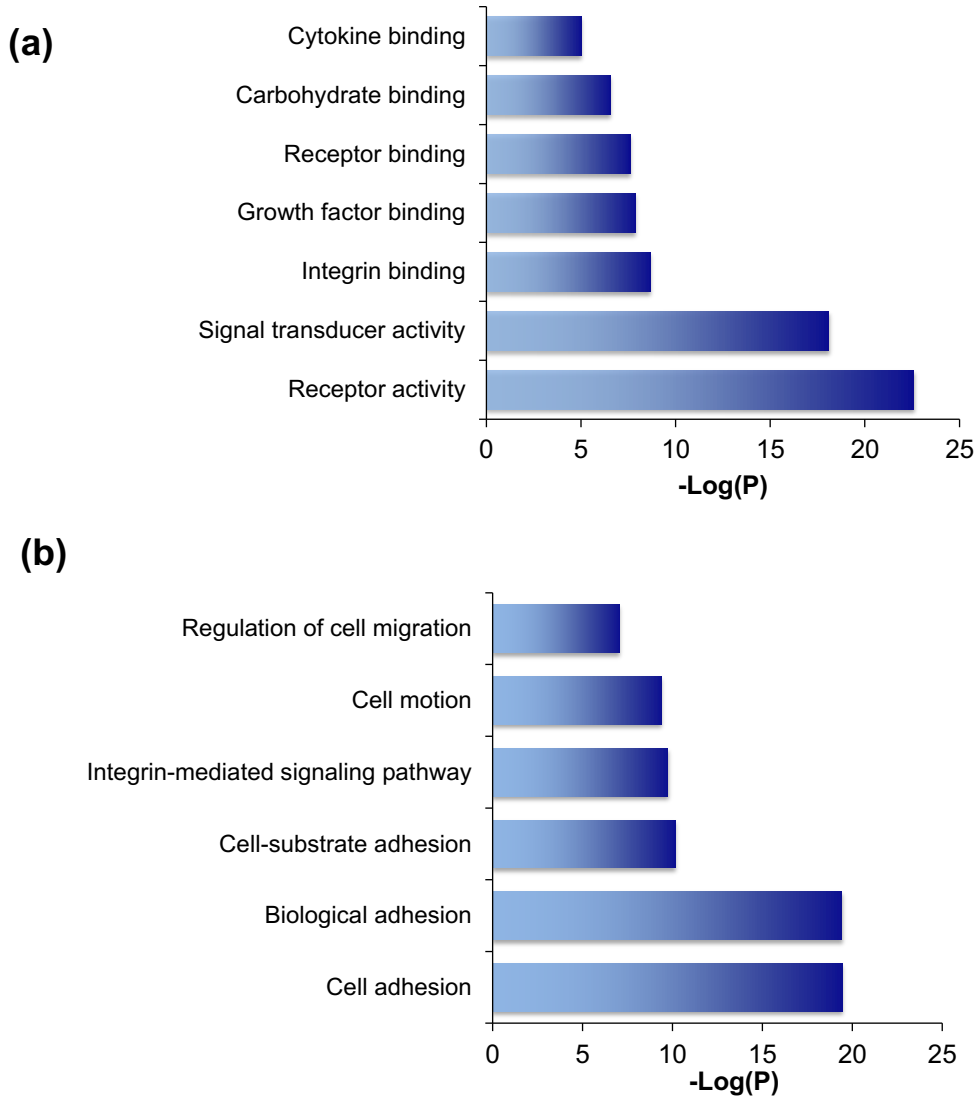


Figure 4.4 Clustering of identified surface N-glycoproteins based on (a) molecular functions and (b) biological processes.

4.1.3.4 Quantification of surface protein N-glycosylation changes in atorvastatin-treated HepG2 cells

Surface glycoproteins play critical roles in cell-cell and cell-matrix interactions. Systematic and quantitative analysis of surface proteins can help us better understand surface protein functions and cellular activities, which will lead to a better understanding of the molecular mechanisms of disease, the discovery of biomarkers, and elucidating the side effects

of drugs. Statins are the most popular and effective drugs for lowering patients' cholesterol. As effective cholesterol-lowering HMGCR inhibitors, statins inhibit the rate-limiting step of the cholesterol biosynthesis pathway, known as the mevalonate pathway. It has been extensively documented that these drugs have pleiotropic effects,[51] but their molecular mechanisms remain to be explored. The inhibition of HMGCR also prevents the synthesis of other products in this pathway, including ubiquinone, dolichol and farnesyl-pyrophosphate (farnesyl-PP). Dolichol is essential to protein N-glycosylation in the form of dolichyl phosphate (Dol-P). Dol-P serves as the carrier in pyrophosphate-linked oligosaccharide assembly as well as acting as the acceptor in the synthesis of the sugar donors Dol-P-Man and Dol-P-Glc from GDP-Man and UDP-Glc, respectively. Upon the inhibition of dolichol, protein N-glycosylation is expected to be dramatically impacted due to the inability to process lipid-linked oligosaccharide biosynthesis and transportation. However, systematic and quantitative analysis of surface N-glycoproteins in statin-treated cells has yet to be reported.

Based on the optimized sugar analog labeling method discussed above, surface protein N-glycosylation changes in atorvastatin-treated cells were analyzed by combining GalNAz labeling and a quantitative proteomics method. Since the primary organ target of statins is the liver, HepG2, a human liver carcinoma cell line, was used in this work. Stable isotope labeling by amino acids in cell culture (SILAC)[52] was employed to evaluate the surface N-glycoprotein changes between statin-treated and untreated cells. Cells were treated by atorvastatin for 24 h to inhibit dolichol synthesis, then labeled with GalNAz for another 24 h in the presence of atorvastatin. An in-flask copper-free click reaction with DBCO-sulfo-biotin was then performed to specifically tag surface N-glycoproteins (Figure 4.5a). Subsequent cell lysis, protein digestion, and enrichment of surface glycopeptides with NeutrAvidin beads were performed as described above. The selectively enriched surface N-glycopeptide samples were analyzed by LC-MS. An example of peptide quantification is shown in Figure 4.5, and the full

MS and elution profile of a peptide (VASVININPN*TTHSTGSCR, where * is the glycosylation site) from LAMP2 (CD107) are shown in Figure 4.5b and c, respectively. LAMP2 is a single-pass type I membrane protein that regulates cell adhesion and inter/intracellular signal transduction when expressed on the cell surface. Based on the elution profiles, we can very accurately quantify the abundance changes of this peptide in statin-treated cells vs. untreated cells, *i.e*. the ratio of the areas under the curves for heavy and light versions of the peptide (H/L = 0.38).

The combination of GalNAz labeling and SILAC led to the quantification of 360 unique N-glycopeptides from 178 cell-surface glycoproteins (listed in a table online at doi.org/10.1021/acs.analchem.5b04871). Among quantified unique glycopeptides, the majority only contained a single N-glycosylation site, while only 20 contained two sites, as shown in Figure 4.6a. The distribution of 360 quantified unique glycopeptides is shown in Figure 4.6b. Based on the two criteria described above, 280 singly glycosylated sites (listed in a table online at doi.org/10.1021/acs.analchem.5b04871) were quantified with a similar distribution, as shown in Figure 4.7.

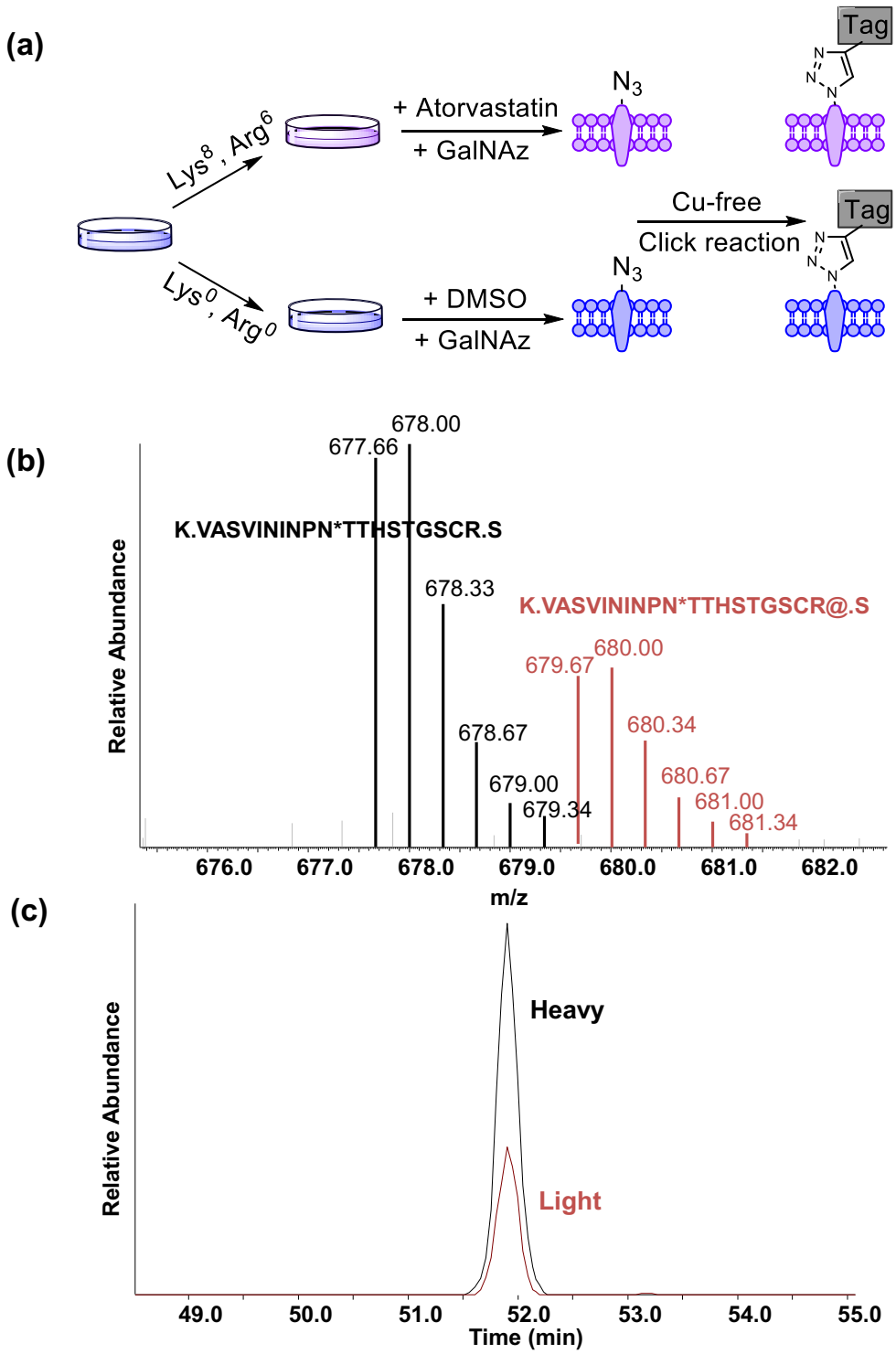


Figure 4.5 (a) Overview of labeling and tagging workflow in quantification experiments, and (b, c) the quantification of the heavy and light versions of an example glycopeptide from LAMP2: (b) full MS (* represents glycosylation site and @ represents heavy arginine) and (c) extracted elution profiles for both versions of the peptides.

In this quantification experiment, we identified significantly more down-regulated glycopeptides or glycosylation sites (103 sites) than up-regulated glycopeptides or glycosylation sites (37 sites) in atorvastatin-treated cells. Although dolichol biosynthesis was inhibited by atorvastatin for one day before GalNAz labeling, dolichol can be recycled in cells after sugar transportation is completed.[53] Therefore, statin treatment for a short time can impact but not entirely prevent protein *N*-glycosylation. Another possible explanation for site up-regulation could be due to the up-regulation of the corresponding parent protein. Namely, if a protein is dramatically up-regulated in treated cells while N-glycosylation sites from this protein are largely unaffected or even slightly down-regulated, we could still find these sites up-regulated. For instance, the abundance of the N1523 site on APOB changed by 1.6 fold in the treated cells, whereas the protein ratio was found to be up-regulated 2.0 fold, as reported previously.[54] Similar effects have been found in protein phosphorylation studies reported in the literature.[55]

4.1.3.5 Analysis of Down-regulated Surface N-glycosylation Sites in Atorvastatin-treated Cells

Among 84 surface proteins bearing 103 down-regulated glycosylation sites, we performed protein clustering using DAVID 6.7.[50] Glycoproteins in the Alzheimer's disease pathway were highly enriched, and glycoproteins and sites in this pathway are listed in Table 4.1. Previous studies have shown inconsistent effects of statin use in AD.[56] Some studies found beneficial effects,[57] but others did not.[58] Glycosylation defects in amyloid precursor protein (APP), tau, nicastrin and other proteins in AD were reported previously,[59] and defective glycosylation may be important in AD pathogenesis, A prior study found that N-glycosylation of human nicastrin was required to interact with lectins, including calnexin and ERGIC-53.[60] In this study, N417 on nicastrin was quantified to be down-regulated by 5.4 fold. Furthermore,

proteins participating in immune response processes, such as response to wounding, external stimulus, etc. were also enriched.

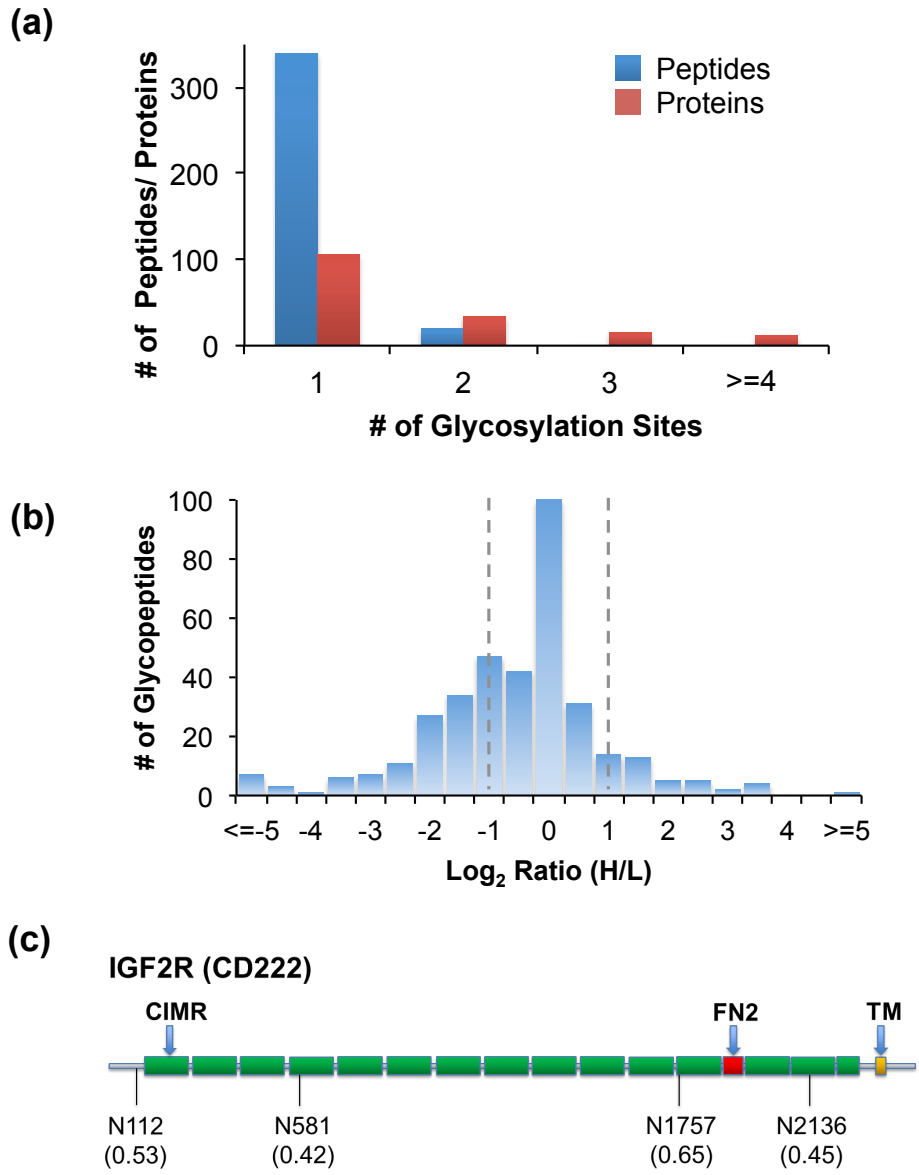


Figure 4.6 (a) Distribution of the quantified glycosylation sites on each peptide and protein, (b) ratio distribution of quantified unique glycopeptides, and (c) domain analysis of IGF2R and quantified N-glycosylated sites (ratio is shown below each site).

We also performed domain analysis to correlate the localization of glycosylation sites and functional domains of proteins.[61] Domains on proteins carry out a wide variety of functions or interactions. Investigating glycosylation site regulations within domains may provide useful information in biological events. For example, CD222, also called IGF2R, is a transporter of phosphorylated lysosomal enzymes from the Golgi complex and the cell surface to the lysosome, and has 15 repeating cation-independent mannose-6-phosphate receptor domains (CIMR) (shown in green in Figure 4.6c). These domains specifically bind the phosphomannosyl residues on lysosomal enzymes. IGF2R also has a fibronectin type II domain (FN2) (shown in red) which serves as the binding site for collagens. All these domains are located in the extracellular space, as shown on the left of the transmembrane domain (TM), which is integrated into the cell plasma membrane. We have quantified four glycosylation sites (N112, N581, N1757, and N2136), among which three are localized within the CIMR domains with the other located at the N-terminal tail. These site abundances decreased to 53%, 42%, 65%, and 45%, respectively, under the statin treatment, which may affect the interactions between this CD and its interactors.

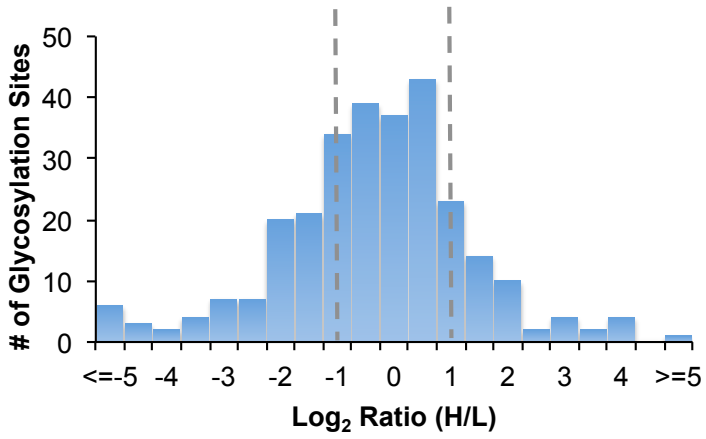


Figure 4.7 Distribution of quantified unique glycosylation sites in atorvastatin-treated cells vs. untreated cells.

By combining GalNAz labeling, click chemistry tagging, and MS-based proteomics, we found that many glycosylation sites on surface proteins were down-regulated in atorvastatin-treated HepG2 cells. Patients are typically prescribed the drug long-term (months to years). Here, we found that many surface protein glycosylation sites were down-regulated when cells were treated for only two days. Further studies, including time-course experiments and animal model experiments, will help us better understand the protein glycosylation changes caused by statin and the molecular mechanisms of its pleiotropic effects.

Table 4.1 Down-regulated glycosylation sites quantified from proteins in the Alzheimer's disease pathway (*P*=0.027)

Gene Symbol	Peptide	PPM	XCorr	Glycosylation Site	Mod Score	Site Ratio	Annotation
ADAM 17	EQQLESCACN*ETDNSCK	1.89	4.69	594	76.7	0.03	Disintegrin and metalloproteinase domain-containing protein 17
	KCQEAIN*ATCK	-0.99	2.41	539	1000	0.21	
NCSTN	RPN*QSQPLPPSSLQR	4.69	2.61	417	1000	0.18	Nicastrin, Essential subunit of the gamma-secretase complex
LRP1	CIPEHWTCDGDNDCGDYSDETHAN*CTNQATRPPGGCHTDEFQCR	2.01	3.17	1050	5.3	0.24	Prolow-density lipoprotein receptor-related protein 1
	QSGDVTCN*CTDGR	-2.61	1.43	4364	1000	0.27	
	CTQQVCAGYCAN*NSTCTVNQGNQPQCR	0.55	5.40	4278	6.6	0.32	
	CTQQVCAGYCANN*STCTVNQGNQPQCR@	-1.68	4.82	4279	6.6	0.32	
FAS	CKPNFFCN*STVCEHCDPCTK	0.1	4.27	136	28.5	0.19	Tumor necrosis factor receptor superfamily member 6
ITPR1	VESGEN*CSSPAPR@	0.13	2.37	2512	1000	0.13	Inositol 1,4,5-trisphosphate receptor type 1

*- glycosylation site; @-heavy arginine

4.1.4 Conclusions

Glycosylation changes on cell-surface proteins are hallmarks of many diseases, but global and site-specific analysis of cell-surface N-glycoproteins is extraordinarily challenging. In-depth analyses of the surface glycoproteome changes will potentially lead to clinical

applications, such as the identification of diagnostic and therapeutic targets. In this work, we compared labeling with three sugar analogs (GalNAz, ManNAz and GlcNAz) for the global analysis of surface glycoproteins, in combination with click chemistry tagging, selective enrichment, and MS analysis. The results clearly demonstrated that more protein glycosylation sites on the cell surface were identified with GalNAz labeling compared to GlcNAz or ManNAz. By using GalNAz labeling, surface protein N-glycosylation changes between statin-treated and untreated cells were comprehensively and site-specifically analyzed in combination with quantitative proteomics. Many glycopeptides were down-regulated in statin-treated HepG2 cells compared to untreated cells because statin prevents the synthesis of dolichol, which is essential for the formation of dolichol-linked precursor oligosaccharides. Several N-glycosylation sites on surface proteins related to Alzheimer's disease were found to be down-regulated. Site-specific information regarding surface proteins will provide insight into protein functions and also lead to a better understanding of the molecular mechanisms of statin's pleiotropic effects.

4.2 Quantitative Investigation of Human Cell Surface N-Glycoprotein Dynamics

4.2.1 Introduction

Nearly all proteins on the cell surface are glycosylated, and surface glycoproteins are essential for cell survival.[1] Protein glycosylation play crucial roles in a wide variety of extracellular activities, including antibody recognition, cell adhesion, microorganism binding, facilitating ligand binding and affecting receptor multimerization.[62-68] Aberrant surface protein glycosylation impacts cellular properties, such as cell solubility and mobility, which is related to human disease,[3, 69, 70] including cancer,[5, 71] congenital disorders and infectious diseases.[6, 72] It has been long understood that the covalent attachment between glycans and proteins is extremely complicated because of the heterogeneity of glycan structures, which makes the comprehensive analysis of protein glycosylation challenging.[14, 16, 45, 73-78] It is even more difficult to analyze glycoproteins only located on the cell surface. The elegant and pioneering work of using sugar analogs to engineer cell surface glycans and glycoproteins has opened a new avenue to study cell surface glycoproteins.[9, 79]

Surface glycoproteins are dynamic for cells to adapt the ever-changing extracellular environment. The presence of glycans on proteins not only facilitates protein folding and trafficking, but also protects proteins from degradation.[80-83] Glycans create a steric hindrance around the peptide backbone, which mechanically prevents proteases from properly binding to proteins. In addition, protein glycosylation also protects the protein backbone from being damaged or degraded through oxidation, chemical crosslinking, precipitation, and denaturation.[84, 85] However, systematic study of glycoprotein dynamics and half-lives has yet to be reported, including the dynamics of crucial cell surface glycoproteins, due to the lack of effective methods. In recent years, MS-based proteomics enable the global analysis of proteins and protein modifications, including glycosylation.[55, 86-95] Due to the complexity of biological

samples, effective separation and enrichment are required to comprehensively analyze every type of protein modification.[71, 96, 97] In order to analyze glycoproteins located only on the cell surface, it is essential to selectively separate and enrich them from high abundance intracellular proteins prior to MS analysis.

In this work, we have designed a method to target surface N-glycoproteins and quantify their half-lives by combining pulse-chase metabolic labelling, click chemistry, and multiplexed proteomics. A sugar analog, N-azidoacetylgalactosamine (GalNAz), was employed to label cells to generate a chemical handle for further surface glycoprotein tagging via copper-free click chemistry under mild physiological conditions. Pulse-chase labelling allowed us to track the abundance changes of cell surface glycoproteins while avoiding contribution from newly synthesized glycoproteins during the cell growth because they were not labelled with the functional azido group. After enrichment of tagged glycopeptides, six-plexed Tandem Mass Tag (TMT) reagents[98] were used to label enriched glycopeptides at six different time points for quantification with MS-based proteomics. Eventually the glycoprotein abundance changes as a function of time were measured, and their half-lives were globally determined. This integrated method specifically targeting surface glycoproteins can be extensively applied to biological and biomedical research.

4.2.2 Experimental section

3.2.2.1 Cell culture, metabolic labeling, and copper-free click reaction

MCF-7 cells (from American type culture collection (ATCC)) were equally seeded into twelve T175 cell culture flasks (Thermo) with Dulbecco's Modified Eagle's Mmedium (DMEM) (Sigma-Aldrich) containing 10% fetal bovine serum (FBS) (Thermo). Cells were grown in a humidified incubator with 5.0% CO_2 at 37 °C. When cells reached 50% confluency, 100 μM GalNAz (Click Chemistry Tools) was added to the media and cells were cultured for

another 24 h. Click reaction was then performed for all flasks. Briefly, cells were gently washed twice with phosphate buffered saline (PBS), then 100 μM dibenzocyclooctyne (DBCO)-sulfo-biotin in DMEM was added into the cell culture flasks. Cells were incubated for 1 h at 37 °C, and then washed twice using PBS. The media were then switched to normal DMEM with 10% FBS and different flasks were further cultured for 0, 2, 4, 6, 8, and 10 hours. Cells were pelleted by centrifugation at 300 g for 5 minutes, and washed twice with cold PBS. Cells were then incubated in a buffer containing 150 mM NaCl, 50 mM HEPES pH=7.6, 25 μg/mL digitonin, and 1 tablet/ 10 mL protease inhibitor (Complete mini, EDTA-free, Roche) on ice for 10 minutes. Cytosolic proteins were removed by centrifuging the samples at 2500 g for another 10 minutes and discarding the supernatant. Cell pellets were lysed through end-over-end rotation at 4 ºC for 45 minutes in lysis buffer (50 mM HEPES pH=7.6, 150 mM NaCl, 0.5% SDC, 10 units/mL benzonase and 1 tablet/ 10 mL protease inhibitor). Lysates were centrifuged, and the resulting supernatant was transferred to new tubes. Proteins were subjected to disulfide reduction with 5 mM DTT (56 ºC, 25 minutes) and alkylation with 14 mM iodoacetamide (RT, 20 minutes in the dark). Detergent was removed by methanol-chloroform protein precipitation. The purified proteins were digested with 10 ng/μL Lys-C (Wako) in 50 mM HEPES pH=8.6, 1.6 M urea, 5% ACN at 31 ºC for 16 hours, followed by further digestion with 8 ng/uL Trypsin (Promega) at 37 ºC for 4 hours.

4.2.2.2 Glycopeptide separation and enrichment

Digestion mixtures were acidified by addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, clarified by centrifugation and desalted using a tC18 Sep-Pak cartridge (Waters). Purified peptides were dried and then enriched with NeutrAvidin beads (Thermo) at 37 °C for 30 minutes. The samples were transferred to spin columns, followed by thoroughly washing according to the manufacturer's protocol. Peptides were then eluted from the beads

three times by 2-min incubations with 200 µL of 8 M guanidine-HCl (pH = 1.5) at 56 °C. Eluates were combined, desalted using tC18 Sep-Pak cartridge, and lyophilized.

4.2.2.3 TMT labelling and PNGase F cleavage

Purified peptides from each of the six time points were labelled with one of the sixplex TMT reagents (Thermo) following the manufacturer's protocol. Briefly, purified and lyophilized peptides were dissolved in 100 µL of 100 mM triethylammonium bicarbonate (TEAB) buffer, pH= 8.5. Each tube of TMT reagents was dissolved in 41 µL of anhydrous DMSO and transferred into the peptide tube. The reaction lasted for 1 h at room temperature, and then was quenched by adding 8 µL of 5% hydroxylamine. Peptides from all six tubes were then mixed, desalted again using a tC18 Sep-Pak cartridge, and lyophilized overnight. Completely dried peptides were deglycosylated with eight units of peptide-*N*-glycosidase F (PNGase F, Sigma-Aldrich) in 40 µL buffer containing 50 mM NH_4HCO_3 (pH=9) in heavy-oxygen water ($H_2{}^{18}O$) for 3 h at 37 °C. The reaction was quenched by adding formic acid (FA) to a final concentration of 1%. Peptides were further purified via stage tip and separated into three fractions using 20%, 50% and 80% ACN containing 1% HOAc, respectively.

4.2.2.4 LC-MS/MS analysis

Purified and dried peptide samples were dissolved in 13 µL of solvent containing 5% ACN and 4% FA, and 4 µL of dissolved sample were loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 µm, 200 Å, 100 µm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 112-minute gradient of 1-12%, 3-14%, or 3-24% ACN (with 0.125% FA) for the three fractions. Peptides were detected with a data-dependent

Top15 method [99] in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 15 MS/MS for the most intense ions. The selected ions were excluded from further analysis for 90 seconds. Ions with single or unassigned charge were discarded. MS^2 scans were performed in the orbitrap cell by activating with high energy collision dissociation (HCD) at 40% normalized collision energy with 1.2 m/z isolation width.

4.2.2.5 Database search and data filtering

All MS^2 spectra were converted into an mzXML format, and then searched using the SEQUEST algorithm (version 28).[37] Spectra were matched against a database containing sequences of all proteins in the UniProt Human (*Homo sapiens*) database (downloaded in February 2014). The following parameters were used during the search: 10 ppm mass tolerance; fully digested with trypsin; up to 2 missed cleava ges; fixed modifications: carbamidomethylation of cysteine (+57.0214), TMT modification of lysine (+229.1629) and N-terminus (+229.1629); variable modifications: oxidation of methionine (+15.9949), ^{18}O tag on asparagine (+2.9883). False discovery rates (FDR) of peptide and protein identifications were evaluated and controlled by the target-decoy method.[38] Each protein sequence was listed in both forward and reversed orders. Linear discriminant analysis (LDA), which is similar to other methods in the literature,[39] was used to control the quality of peptide identifications using parameters such as Xcorr, precursor mass error, and charge state.[40] Peptides fewer than seven amino acid residues in length were deleted. Furthermore, peptide spectral matches were filtered to <1% FDR. The dataset was restricted to glycopeptides when determining FDRs for glycopeptide identification.[41] Furthermore, an additional protein-level filter was applied in

each dataset to reduce the protein-level FDRs (<1%) for glycoproteins. Consequently the FDRs at the glycopeptide level were much less than 1%.

4.2.2.6 Glycosylation site localization, glycopeptide quantification, and bioinformatics analysis

The confidence associated with each glycosylation site localization was represented by their ModScore, which is calculated from a probabilistic algorithm.[41] ModScore is similar to AScore,[41] and it considers all possible modification sites in a modified peptide, and matches the fragments with theoretical fragments from the peptide with potential modification sites. If the ModScore for a residue is relatively high, then the probability of modification occurred on that site is also high. Conversely, there may be a low score for potential sites, which means that there are not sufficient fragments to confidently locate the modification site. Sites with ModScore > 13 (P <0.05) were considered as confidently localized. The TMT reporter ion intensities obtained in MS^2 were recorded and calibrated prior to performing glycopeptide quantification. If the same glycopeptide was quantified several times, the median value was used as the glycopeptide abundance change. The protein ratio was calculated based on the median ratios of all unique glycopeptides. Protein annotations were extracted from the UniProt database (http://www.uniprot.org). The Database for Annotation, Visualization and Integrated discovery (DAVID) v6.7 (http://david.abcc.ncifcrf.gov/home.jsp)[100] was employed to perform functional analysis. All raw files and annotated spectra are accessible in the following public accessible website (http://www.peptideatlas.org/PASS/PASS00913, Password: BE6745wv).

4.2.3 Results

4.2.3.1 The principle of surface glycoprotein enrichment and identification

The incorporation of bio-orthogonal groups into proteins or modified proteins has recently been demonstrated to be very effective to study proteins in complex biological systems.[9, 33, 101-108] In this work glycoproteins were labelled with a sugar analog containing a biologically inert but chemically functional azido group, and the labelled surface glycoproteins in living cells were specifically tagged with biotin via copper-free click chemistry (Figure 4.8). Here we performed click reaction prior to medium switch, which can eliminate potential negative effects from cells using stored GalNAz and protein internalization on protein half-life quantification. After cell lysis and protein digestion, only biotin-tagged glycopeptides were selectively enriched with NeutrAvidin beads through specific biotin-avidin interactions. Enriched and purified samples were analyzed by an online LC-MS system, and both full MS and MS^2 were recorded in the Orbitrap cell with high resolution and high mass accuracy.

The TMT method enables the identification and quantitation of glycopeptides and glycoproteins in different samples in combination with tandem mass spectrometry (MS). The tags contain four regions, namely a mass reporter region, a cleavable linker region, a mass normalization region and a protein/peptide reactive group. In this case, the reactive group of N-hydroxysuccinimide (NHS) can react quickly with the amine group at the N-terminus and the side chain of the lysine residue for every peptide. Each of six samples was tagged with one channel of TMT reagents, then mixed. For the same peptide in six samples, they all carry isobaric tags, and have the same elution time and m/z in the full mass spectra. When peptides are fragmented, the reporter ion generated from tagged peptide will have an intensity proportional to the peptide amount in each sample. Eventually peptide backbone fragments allow us to identify peptides and the reporter ion intensities enable us to quantify the peptide abundance changes across the six samples.

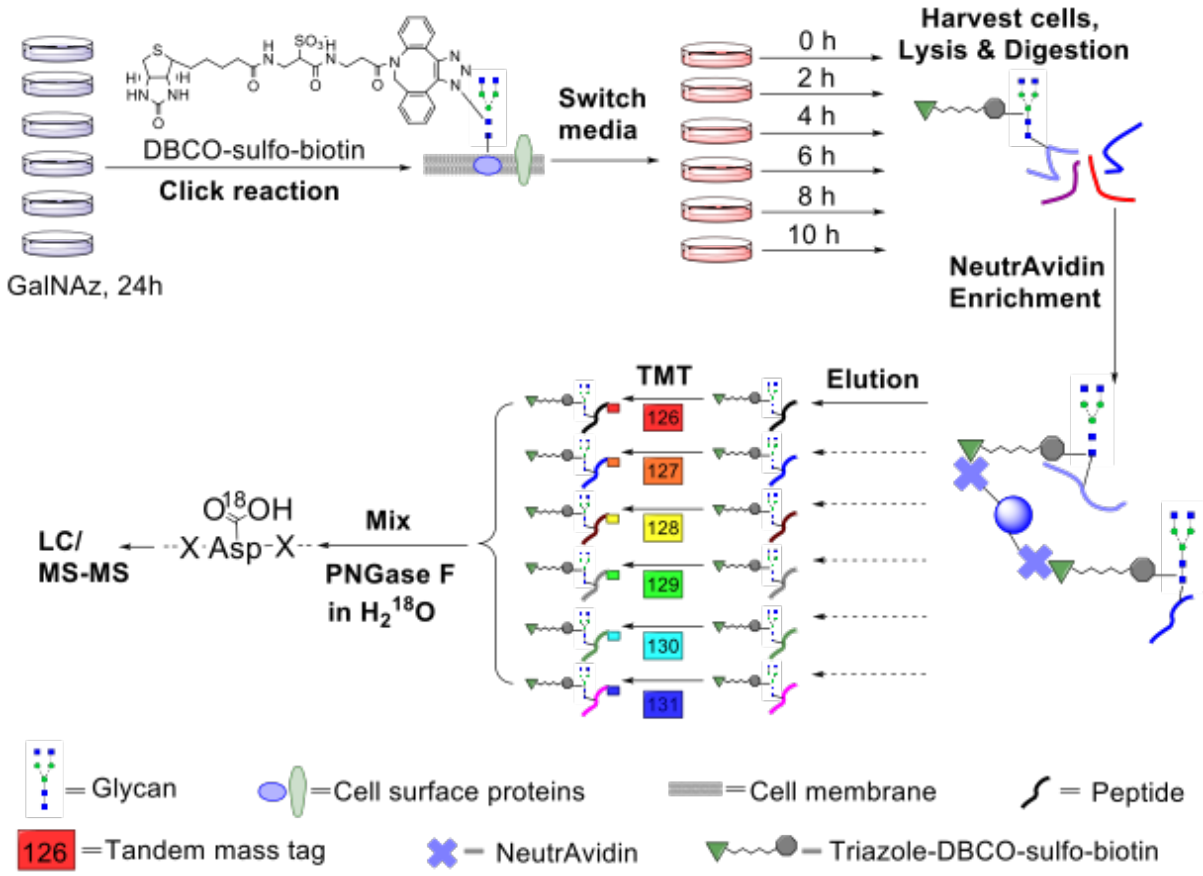


Figure 4.8 Experimental procedure for studying surface glycoprotein dynamics and measuring their half-lives.

An example of peptide identification and quantification is displayed in Figure 4.9(a). The peptide N#VSVAEGK (# denotes the glycosylation site) was confidently identified with an XCorr of 3.2. The XCorr value is the cross-correlation value from the SEQUEST search, which reflects how good the match between theoretical and experimental tandem mass spectra is. XCorr values are usually higher for well-matched, large peptides, and lower for smaller peptides. Considering the short length of this peptide, this XCorr value can allow us to confidently identify this glycopeptide, and as shown in Figure 4.9(a), nearly all *y* and *b* ions were detected. The ModScore for the glycosylation site (N286) is 1000 because there is only one possible site localization. This peptide is from PTGFRN, which is a well-known receptor regulator located on the cell surface.[109] The reporter ion intensities enabled us to accurately

quantify the glycopeptide abundance changes as a function of time (Figure 4.9(a), left insert). Correspondingly we were able to calculate the half-life of 15.5 h based on the abundance changes (Figure 4.9(a), right insert).

Here in duplicate experiments, we identified a total of 545 unique glycosylation sites on 265 glycoproteins (listed in tables online at doi.org/10.1039/C6SC01814A), and most of them (480 sites) were well localized with a ModScore >13. The overlap of unique glycosylation sites identified between two replicates is around 80% across all identification and quantification results (Figure 4.9(b) and (c)), which demonstrated that the current method is highly reproducible. The majority of unique glycopeptides contained a single glycosylation site, and there are a small group of proteins bearing more than five sites, including IGF1R, ECE1, LAMP1, CELSR2, PLXNB2, CEACAM5, ITGB1, and PTPRJ. For example, IGF1R, a receptor tyrosine kinase which mediates actions of insulin-like growth factor 1 (IGF1) located on plasma membrane. Here we identified eleven glycosylation sites: N244, N314, N607, N622, N638, N640, N747, N756, N764, N900, and N913. All these sites exist in the extracellular space, which is further discussed below.

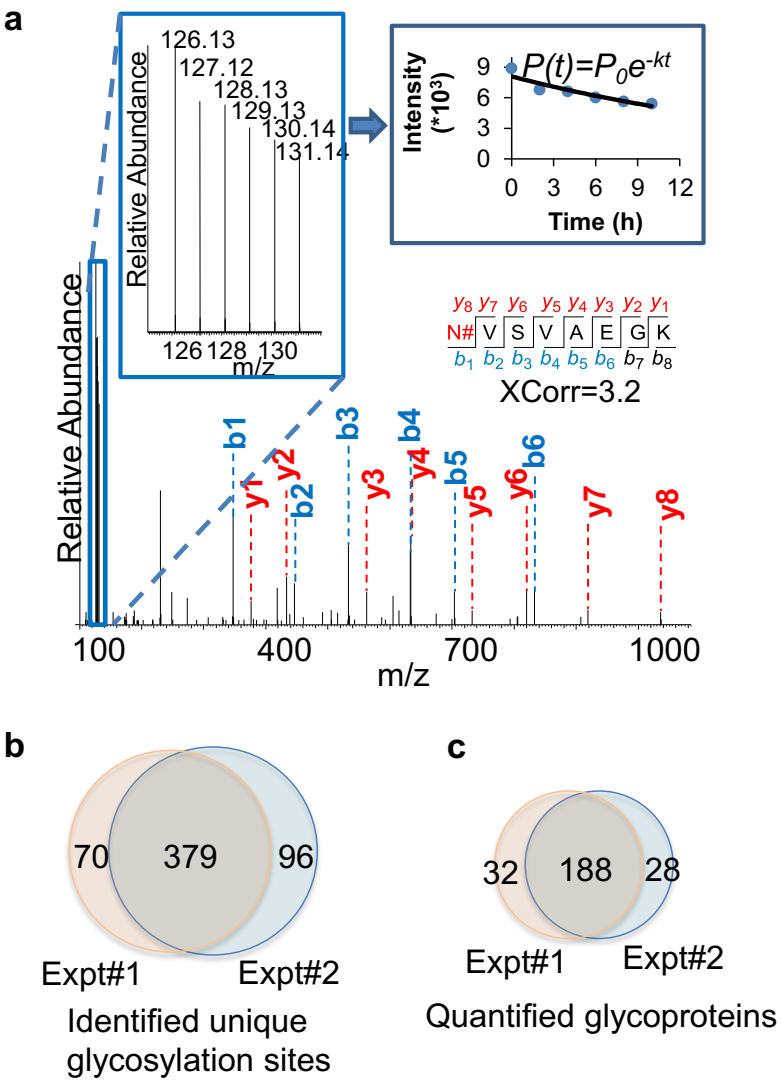


Figure 4.9 An example of glycopeptide identification and quantification and the comparison of identified unique glycosylation sites and quantified surface glycoproteins. (a) Example MS showing peptide identification and quantification. Based on the fragments, we were able to confidently identify the glycopeptide N#VSVAEGK (# denotes the glycosylation site) from the protein PTGFRN, and based on the reporter ion intensities, the half-life of this glycopeptide was 15.5 hours. (b) Comparison of the unique surface protein glycosylation sites identified in two parallel experiments. (c) Comparison of the quantified surface glycoproteins in duplicate experiments.

We clustered the identified glycoproteins according to cellular compartment and pathway using the Database for Annotation, Visualization, and Integrated Discovery 6.7 (DAVID 6.7) (Figure 4.10). [50] For cellular compartments, membrane-related categories were

highly enriched, including intrinsic to membrane (P=1.80*10^{-81}), plasma membrane (8.50*10^{-49}), cell surface (6.20*10^{-20}), external side of the plasma membrane (5.50*10^{-14}), and receptor complex (6.20*10^{-11}). Among the pathways, the ECM-receptor interactions (5.60*10^{-11}) and cell adhesion molecules (CAMs) (1.30*10^{-10}) pathways were prominently enriched. CAMs are cell-surface proteins involved in binding with the extracellular matrix (ECM) or with other cells during cell adhesion. These enriched categories are consistent with the expected functions of cell-surface glycoproteins.

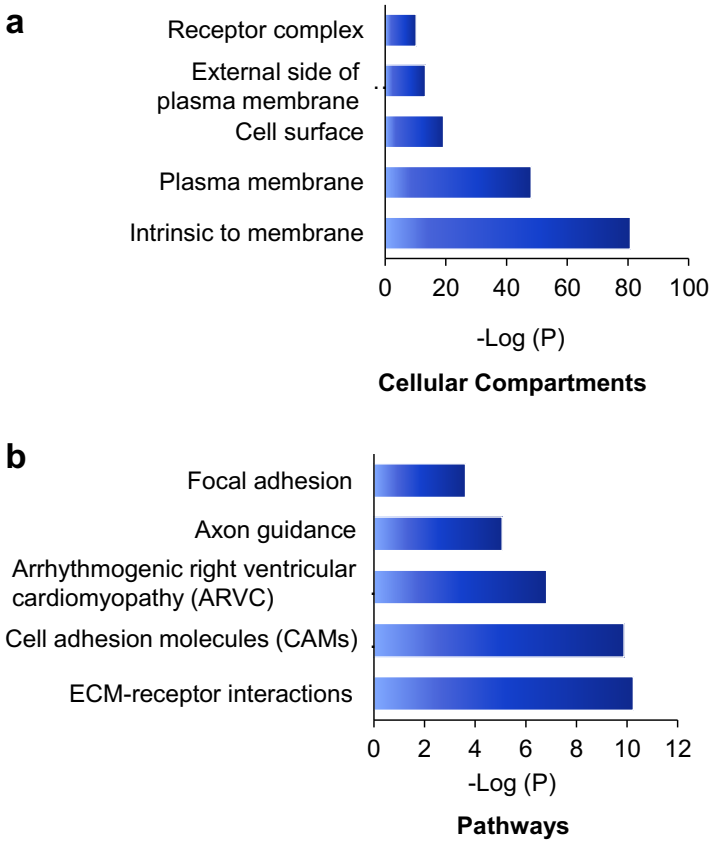


Figure 4.10 Clustering of surface glycoproteins identified in this work. (a) Cellular compartments, and (b) pathways.

4.2.3.2 Site location of type I and II glycoproteins based on the transmembrane domain

The site-specific virtue of our method allowed us to localize each glycosylation site in this experiment. In Figure 4.11, we illustrated the site localization on type I and II

transmembrane glycoproteins identified in this experiment. Type I transmembrane proteins have their N-termini located in the extracellular space while type II transmembrane proteins have their C-termini located in the extracellular space. As shown in Figure 4.11, the *x*-axis represents the transmembrane (TM) domain of any proteins, and the *y*-axis denotes the number of amino acid residues away from the transmembrane domain. The space above the *x*-axis is the extracellular space and below is the intracellular space. Each line depicts a protein, and the yellow dots represent the glycosylation sites.

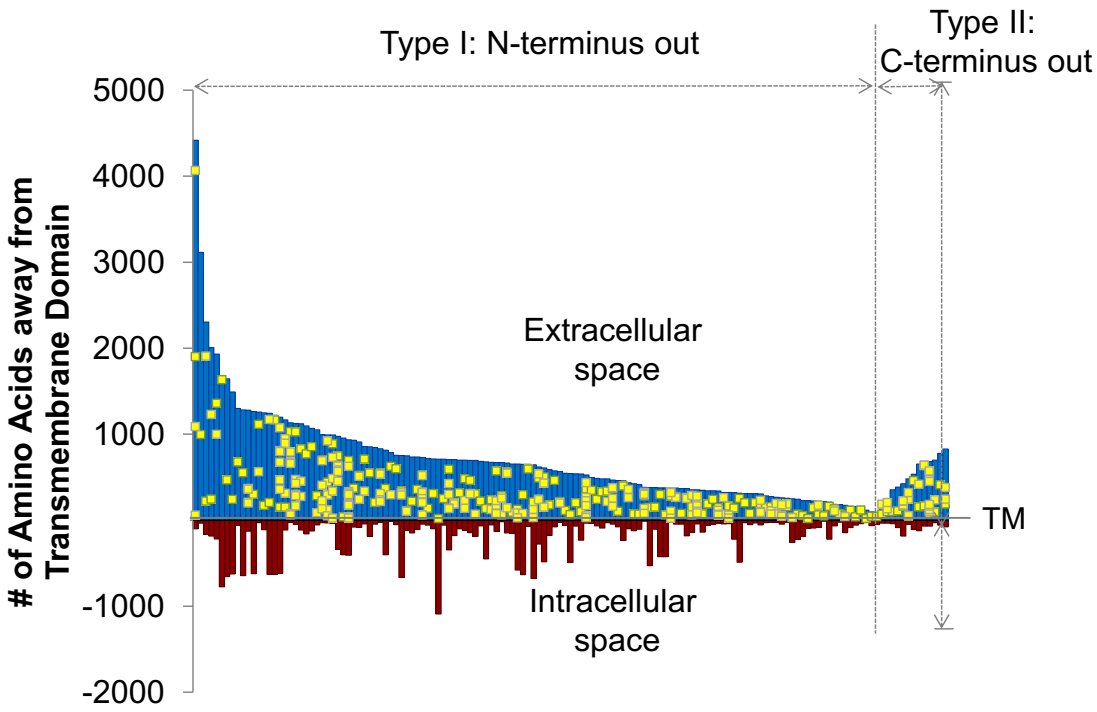


Figure 4.11 Site location of the type I and II N-glycoproteins based on the transmembrane domain (TM). We aligned each glycoprotein according to their transmembrane domain, which is known to be anchored in the plasma membrane, and each yellow dot refers to one glycosylation site.

All glycosylation sites are clearly located in the extracellular space, which is in agreement with the experimental design and the common belief that glycans on surface proteins are located outside of the cell. We identified many more type I transmembrane proteins than

type II, which corresponds well to the ratio of type I and II transmembrane proteins in UniProt (www.uniprot.org).

4.2.3.3 Quantification of surface glycoprotein abundance changes

Enriched peptides in each sample were labelled with one of six TMT reagents. TMT labelling allowed us to quantify multiple samples at once. Here we measured six samples from six time points simultaneously. This can dramatically increase the experimental throughput and reduce potential quantification errors. The starting amount of labelled surface glycoproteins was similar for each sample before the medium was switched. Based on this, we then quantified these surface glycoprotein abundance changes as a function of time. The six groups of TMT-labelled glycopeptides were mixed and subjected to PNGase F cleavage in heavy-oxygen water ($H_2{}^{18}O$) to generate a common tag (+2.9883 Da) for MS analysis.[45] This enabled us to distinguish authentic N-glycosylation sites from those caused by the naturally occurred deamidation of Asn. Finally, the peptide mixture was purified and loaded into an online LC-MS system for further analysis.

The TMT reporter ion intensities in the MS^2 provided us an opportunity to accurately measure the abundance changes of glycopeptides from different time points. Potential interferences from TMT labelling were likely avoided in this experiment because these samples were much simpler than whole cell lysates since surface glycoproteins only represent a very small portion of the whole proteome, and we also further fractionated the mixed sample into three fractions. Furthermore, long LC gradients were used to separate each fraction. Because the abundances of the same glycopeptide from six samples can be measured in one MS^2 spectrum, this dramatically lowered the measurement error. In some cases, for instance, if one of the six TMT channels has abnormal signal intensity, then it will be dropped and the half-life will be calculated based on the signal intensities from the other five channels.

4.2.3.4 Measurement of surface glycoprotein half-lives

Based on the abundance changes of glycopeptides at six time points, their half-lives were simulated by the following exponential decay equation, as performed previously:[99, 110]

$$P(t) = P_0 * \exp(-kt)$$

where P_0 is the intensity of the reporter ion at the first time point, $P(t)$ is the intensity of the reporter ion at each subsequent time point, k is the degradation rate constant and t is time. In duplicate experiments, we quantified 522 unique glycopeptides(ModScore>13); the vast majority of them (484 glycopeptides) contained a single glycosylation site.

In the duplicate experiments, we quantified 386 glycosylation sites (listed in tables online at doi.org/10.1039/C6SC01814A) based on two criteria: glycopeptides were singly glycosylated and the ModScore was larger than 13. If a glycoprotein contained two or more unique glycosylation sites, the half-life refers to the median half-life of the mixed different glycoforms. The half-life values for the 248 glycoproteins were determined, and are listed in tables online at doi.org/10.1039/C6SC01814A. We plotted the half-lives of glycosylation sites in replicate 1 against those in replicate 2 (Figure 4.12(a)), and good linear simulation and a high R^2 value were obtained. The reproducibility is much better when the half-lives are relatively low, which is discussed below.

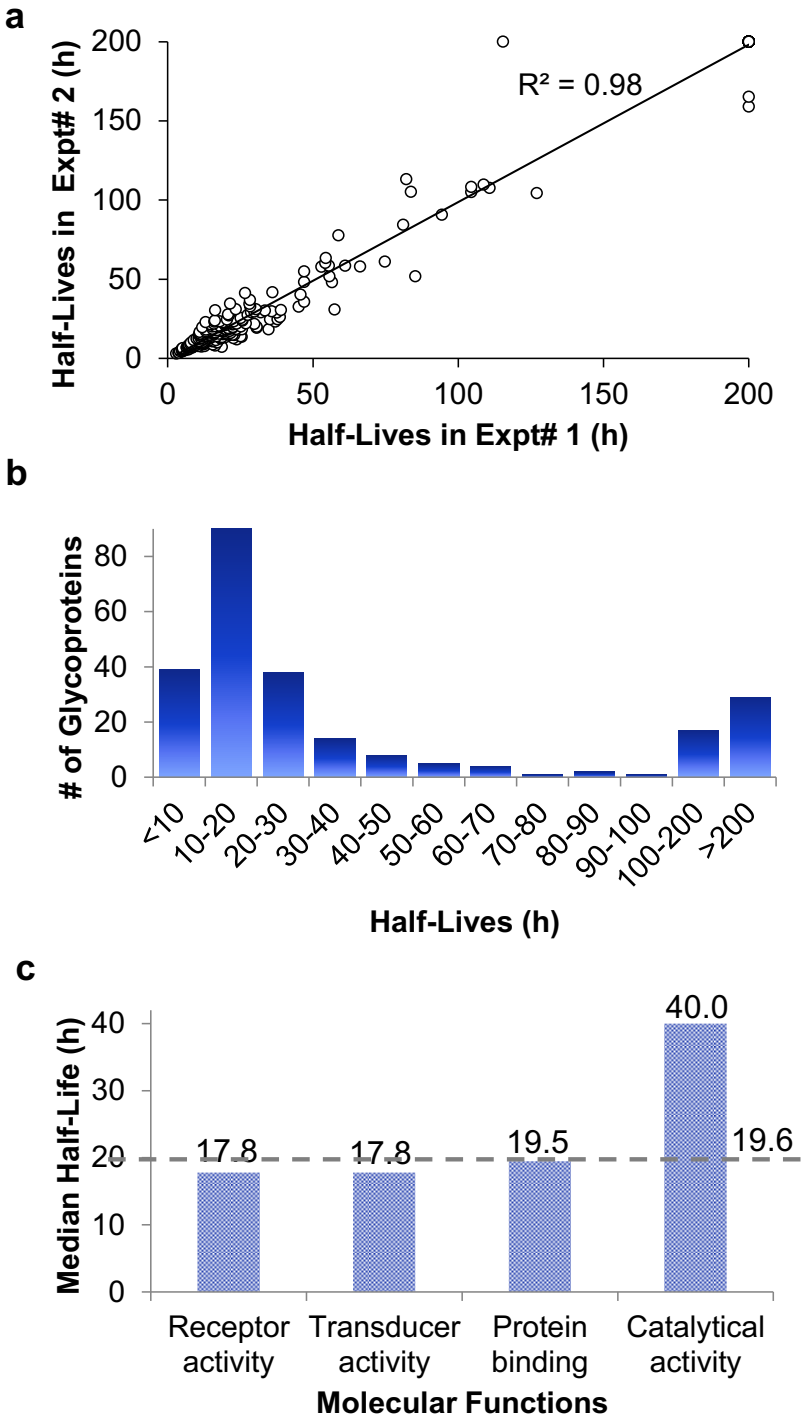


Figure 4.12 (a) Distribution of the half-lives of surface glycoproteins. (b) Comparison of the half-lives of surface protein glycosylation sites measured in the duplicate experiments. (c) The median half-lives of glycoproteins with different molecular functions. Proteins with receptor and transducer activities have the shortest median half-life (17.8 h), while proteins with catalytic activity have a longer median half-life (40.0 h).

The distribution of the half-lives of surface glycoproteins is shown in Figure 4.12(b). Most proteins have a half-life between 10-30 h. A total of 39 glycoproteins have a half-life of less than 10 h, while about one fifth of glycoproteins (46) have a half-life of longer than 100 h. The median half-life of all glycoproteins quantified in our experiment was 19.6 h, which is much longer than the half-life of 8.7 h for over 800 newly synthesized proteins in our previous work,[99] and also longer than a half-life of 8.2 h for 100 proteins measured with a MS-independent method.[111] This is consistent with the assumption that glycans can stabilize proteins by preventing them from being degraded.

The functions associated with relatively long- or short-lived proteins were also investigated. Proteins with a half-life longer than 100 h or shorter than 10 h were clustered according to biological processes (Figure 4.13). While cell adhesion is enriched in both categories, notably, positive regulation of catalytic activity is enriched among long-lived proteins.

The median half-lives for proteins with various molecular functions were examined. As shown in Figure 4.12(c), the median half-life of proteins with receptor activity (17.8 h), molecular transducer activity (17.8 h), and binding activity (19.5) is very similar to the overall protein median half-life (19.6 h), while proteins related to catalytic activity (do not include receptor tyrosine kinases since they only have intracellular catalytic activities) have a longer median half-life of 40.0 h, which suggests that glycan may protect enzymes on the cell surface more effectively.

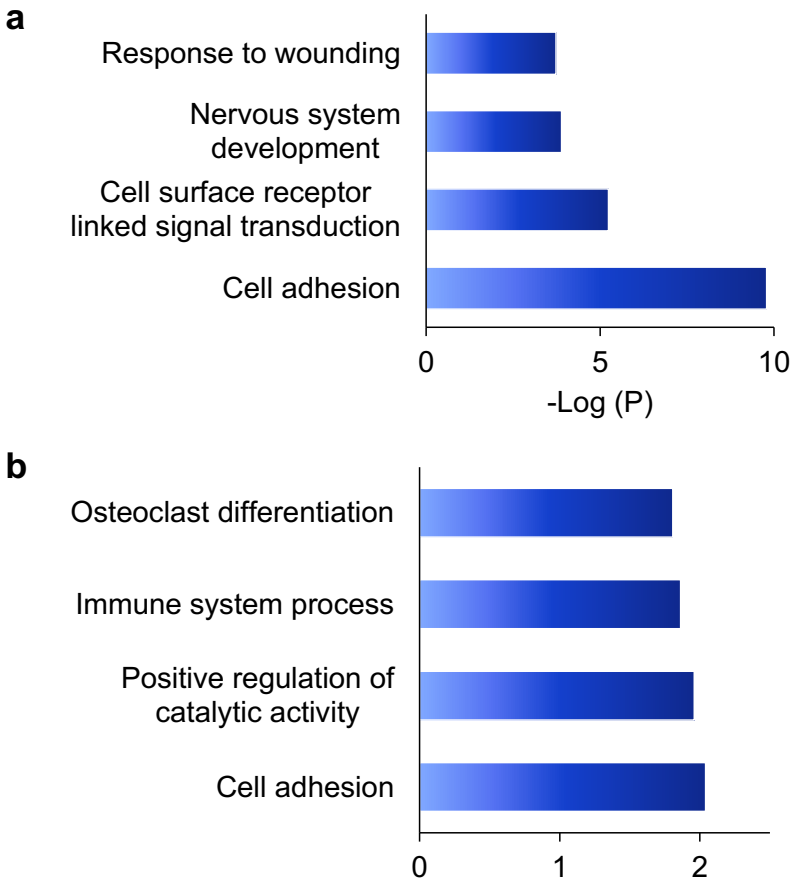


Figure 4.13 (a) Biological processes of relatively short-lived glycoproteins (half-life <10 h). (b) Biological processes of relatively long-lived glycoproteins (half-life >100 h).

4.2.3.5 Half-lives of glycosylation sites within or outside of domains

Among 386 quantified glycosylation sites, nearly half of them (170 sites) were located in different domains based on the domain information on UniProt, while 216 sites were not located in any protein domains. The median half-life for the 216 sites located outside of any domain is 21.5 h, which is 21% longer than that the median half-life of 170 sites located within a specific domain (17.7 h), as shown in Figure 4.14(a). The domains containing the greatest number of quantified glycosylation sites are Ig-like, fibronectin type-III, cadherin, and sema domains, which are shown in Figure 4.14(b), along with their median half-lives. These domains are frequently contained in cell surface proteins, and play crucial roles in regulating cell-matrix interactions and cell surface receptor protein-ligand interactions. 68 sites are located in the Ig-

like domain with a median half-life of 19.0 h. The median half-life of sites located in the cadherin domain is only 11.0 h, which is dramatically shorter than the median half-life of 21.5 h for sites located outside of any domain. These results suggest that glycans located within a domain may play a major role in regulating protein interactions with other molecules, while glycans located outside of any domain are mostly involved in protecting proteins from degradation.

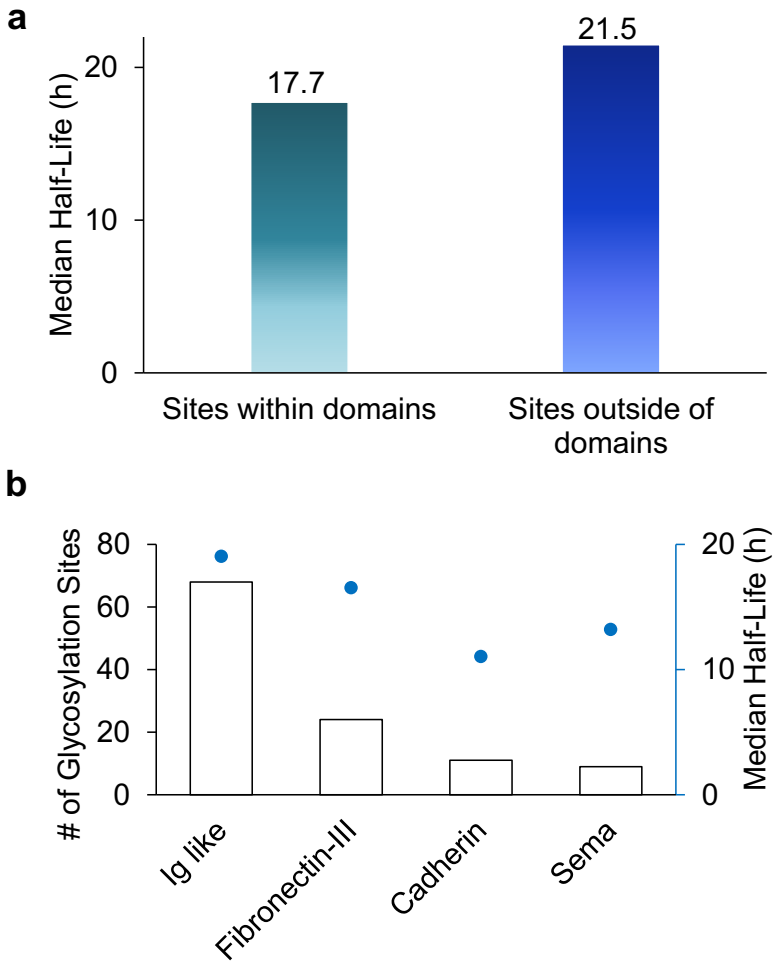


Figure 4.14 (a) Comparison of median half-lives for sites located outside domains and within domains. (b) The number of glycosylation sites located in different domains and their median half-lives

4.2.3.6 Half-lives of CD proteins and receptors

Cluster of differentiation (CD) molecules are of great biomedical significance because they serve as cell markers in immunophenotyping to distinguish and classify cells.[112] CDs refer not only to proteins but can also be assigned to lipid and glycans on the cell surface. Among all of the glycoproteins quantified here, 62 are CD proteins (Table 4.2 and a table online at doi.org/10.1039/C6SC01814A). The site-specific nature of this method provides an avenue to quantify the real glycosylated form of proteins. For example, we identified the glycosylation site N365, N381, and N424 on CD98, which is involved in sodium-independent, high-affinity transport of large neutral amino acids. The half-life of the glycosylated form of CD98 is 27.2 h, which is much longer than the half-life (10.1 h) in the literature.[99] Furthermore, the half-life of CD71 (Transferrin receptor protein 1) is 18.3 h in this work for its glycoform on the cell surface, while the half-life of this protein was reported to be 4.4 h previously.[99] Glycosylated and non-glycosylated forms of a protein coexist at any given time. Traditional gel-based or MS-based methods measure the half-life of the mixed glycosylated and non-glycosylated forms of a protein, but here we were able to measure the half-lives of only the glycosylated form of each protein because only surface glycoproteins were separated and analyzed.

4.2.4 Discussion

Mammalian cell surface is typically covered with sugars, and these sugars may be bound to lipids or proteins. Glycoproteins located on the cell surface regulate nearly every extracellular activity. Systematic and quantitative analysis of surface glycoproteins can aid in a better understanding of protein structure, properties and functions and also cellular activities. Due to the heterogeneity of glycans and low abundance of many glycoproteins, it is extremely challenging to globally identify and quantify glycoproteins in complex biological samples.[97] It is even much more challenging to specifically analyze surface glycoproteins. Fluorescence

experiments have obtained very valuable information about cell surface glycans.[113] However, it is hard to identify which proteins are bound to glycans and the exact glycosylation sites. MS-based proteomics provides the possibility to identify and quantify glycoproteins, but in order to analyze surface glycoproteins, selective enrichment of surface glycoproteins is required prior to MS analysis. It has remained a daunting task to systematically investigate cell surface glycoproteins dynamic, and to date, it has yet to be reported. Integrating pulse-chase metabolic labelling, selective enrichment of surface glycoproteins, and multiplexed proteomics, for the first time, we site-specifically and systematically quantified surface glycoprotein abundance changes as a function of time, and measured their half-lives.

Table 4.2 Half-lives of exemplary CD proteins.

UniProt ID	Gene symbol	CD name	Protein half-life		Annotation
			This work (h)	Previous work (h)[a]	
P02786	TFRC	CD71	18.3	4.4[99]	Transferrin receptor protein 1
P05556	ITGB1	CD29	24.2		Integrin beta-1
P08069	IGF1R	CD221	12.6		Insulin-like growth factor 1 receptor
P08195	SLC3A2	CD98	27.2	10.1[99]	4F2 cell-surface antigen heavy chain
P08962	CD63	CD63	24.2		CD63 antigen
P25445	FAS	CD95	39.1		Tumor necrosis factor receptor superfamily member 6
P26006	ITGA3	CD49c	37.3		Integrin alpha-3
P48960	CD97	CD97	13.2		CD97 antigen
P54709	ATP1B3	CD298	61.1		Sodium/potassium-transporting ATPase subunit beta-3
P78536	ADAM17	CD156b	112.7		Disintegrin and metalloproteinase domain-containing protein 17

[a] Half-lives of corresponding proteins reported in the literature.

Besides protein degradation, other contributions to cell surface glycoprotein dynamics include protein internalization/recycling, and deglycosylation. By tagging surface glycoproteins immediately before the medium switch, the effect of protein internalization on

the measurement of protein half-lives was able to be avoided because even though a protein was internalized, the biotin tag can ensure that it will be eventually analyzed. In other cases, when deglycosylation event happens, the protein will turn into a non-glycoprotein, which does not fit into our experimental subject and thus will not be enriched and analyzed.

One limitation of this method is that proteins with very long half-lives might not be accurately determined because the full length of the time course may only cover the very beginning of the simulation curve, thus a minor variation could result in a large error. We applied a 200 h cut-off value to those long-lived proteins, namely, any protein with a half-life longer than 200 h was included in the >200 h category. Although this category did not present actual half-life values, it still indicates that these glycoproteins are very stable. Since the majority of the glycoproteins have half-lives shorter than 200 h, this category did not affect the calculation of the median half-life nonetheless.

The current experimental results have clearly demonstrated that glycans can more effectively protect enzymes than receptors and binding proteins located on the cell surface from being degraded, because proteins related to catalytic activity have a long median half-life of 40.0 h (quantified surface enzymes are listed in a table online at doi.org/10.1039/C6SC01814A). It is well-known that there are many proteases in the extracellular space, but these quantified surface enzymes were still relatively more stable. For example, for GAPDH (glycerolaldehyde-3-phosphate dehydrogenase), it has both glyceraldehyde-3-phosphate dehydrogenase and nitrosylase activities, thus playing a role in glycolysis and nuclear functions, respectively. In our previous work, its half-life was measured to be 10.4 h.[99] Here its glycoform located on the cell surface are extraordinarily stable with a half-life of more than 200 h. Proteins in the mitochondria or nuclei typically have a longer half-life because proteins located in these compartments may avoid being accessed by many

proteases. Cell surface proteins are exposed to different environments, but glycans on surface proteins may provide one layer of protection, especially for proteins with catalytic activity.

4.2.5 Conclusions

We have designed the first method to target surface glycoproteins, site-specifically study their dynamics and measure their half-lives by incorporating metabolic labelling, click chemistry, and TMT tagging. The current method has several advantages. Firstly, only surface glycoproteins were selectively tagged and enriched for MS analysis. Secondly, site-specific protein glycosylation information was obtained in this work, and only authentic glycosylated forms of proteins were analyzed. Thirdly, multiplexed proteomics enabled to quantify glycoproteins at several time points simultaneously, increasing the accuracy of measuring protein abundance changes and the corresponding half-lives. Furthermore, the high throughput MS-based experiment allowed us to systematically study surface glycoprotein dynamics.

By using this new method, we quantified 248 surface glycoprotein dynamics with the median half-life was 19.6 h, which is over two times longer than that of newly synthesized proteins measured in our recent work (8.7 h).[99] The median half-life of glycopeptides with sites located outside of any domain is longer than that of glycopeptides with sites within different domains. Surface glycoproteins corresponding to catalytic activities were more stable with the median half-life of 40.0 h. Although there are many proteases outside of the cells, glycans can effectively protect surface enzymes from being degraded. Investigation of surface glycoprotein dynamics can aid in better understanding their properties and functions. This method can be extensively applied to investigate surface glycoproteins and their dynamics in biological and biomedical research.

4.3 References

1. Varki, A. et al. Essentials of Glycobiology (2nd Edition). (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; 2008).
2. Gabius, H.J., Andre, S., Kaltner, H. & Siebert, H.C. The sugar code: Functional lectinomics. *Biochim. Biophys. Acta-Gen. Subj.* **1572**, 165-177 (2002).
3. Dennis, J.W., Granovsky, M. & Warren, C.E. Protein glycosylation in development and disease. *Bioessays* **21**, 412-421 (1999).
4. Wang, J.Z., GrundkeIqbal, I. & Iqbal, K. Glycosylation of microtubule-associated protein tau: An abnormal posttranslational modification in Alzheimer's disease. *Nat Med* **2**, 871-875 (1996).
5. Christiansen, M.N. et al. Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525-546 (2014).
6. Vigerust, D.J. Protein glycosylation in infectious disease pathobiology and treatment. *Cent. Eur. J. Biol.* **6**, 802-816 (2011).
7. Rambourg, A., Neutra, M. & Leblond, C.P. Presence of a cell coat rich in carbohydrate at surface of cells in rat. *Anat Rec* **154**, 41-79 (1966).
8. Gahmberg, C.G. & Tolvanen, M. Why mammalian cell surface proteins are glycoproteins. *Trends Biochem Sci* **21**, 308-311 (1996).
9. Mahal, L.K., Yarema, K.J. & Bertozzi, C.R. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **276**, 1125-1128 (1997).
10. Lau, K.S. et al. Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* **129**, 123-134 (2007).
11. Smeekens, J.M., Chen, W.X. & Wu, R.H. Mass spectrometric analysis of the cell surface N-glycoproteome by combining metabolic labeling and click chemistry. *J Am Soc Mass Spectr* **26**, 604-614 (2015).
12. Tian, Y.A., Kelly-Spratt, K.S., Kemp, C.J. & Zhang, H. Mapping tissue-specific expression of extracellular proteins using systematic glycoproteomic analysis of different mouse tissues. *J Proteome Res* **9**, 5837-5847 (2010).
13. Yates, J.R., Eng, J.K., McCormack, A.L. & Schieltz, D. Method to correlate tandem mass-spectra of modified peptides to amino-acid-sequences in the protein database. *Anal. Chem.* **67**, 1426-1436 (1995).
14. Wollscheid, B. et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378-386 (2009).
15. Cannon, J., Nakasone, M., Fushman, D. & Fenselau, C. Proteomic identification and analysis of K63-linked ubiquitin conjugates. *Anal. Chem.* **84**, 10121-10128 (2012).
16. 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).
17. Sidoli, S. et al. Sequential window acquisition of all theoretical mass spectra (SWATH) Analysis for characterization and quantification of histone post-translational modifications. *Mol. Cell. Proteomics* **14**, 2420-2428 (2015).

18. Marino, F. et al. Extended O-GlcNAc on HLA class-I-bound peptides. *J. Am. Chem. Soc.* **137**, 10922-10925 (2015).

19. Olsen, J.V. et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).

20. Burke, M.C., Oei, M.S., Edwards, N.J., Ostrand-Rosenberg, S. & Fenselau, C. Ubiquitinated proteins in exosomes secreted by myeloid-derived suppressor cells. *J Proteome Res* **13**, 5965-5972 (2014).

21. 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).

22. Huang, H., Lin, S., Garcia, B.A. & Zhao, Y.M. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115**, 2376-2418 (2015).

23. Khidekel, N. et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* **3**, 339-348 (2007).

24. Chen, W.X., Smeekens, J.M. & Wu, R.H. A universal chemical enrichment method for mapping the yeast n-glycoproteome by mass spectrometry (MS). *Mol Cell Proteomics* **13**, 1563-1572 (2014).

25. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chemical Science* **6**, 4681-4689 (2015).

26. Woo, E.M., Fenyo, D., Kwok, B.H., Funabiki, H. & Chait, B.T. Efficient identification of phosphorylation by mass spectrometric phosphopeptide fingerprinting. *Anal. Chem.* **80**, 2419-2425 (2008).

27. Sidoli, S., Lin, S., Karch, K.R. & Garcia, B.A. Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative abundance and stoichiometry. *Anal. Chem.* **87**, 3129-3133 (2015).

28. Phanstiel, D. et al. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4093-4098 (2008).

29. Liu, T. et al. Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res* **4**, 2070-2080 (2005).

30. Liao, J.K. Statin therapy: Having the good without the bad. *Hypertension* **43**, 1171-1172 (2004).

31. Forbes, K. et al. Statins inhibit insulin-like growth factor action in first trimester placenta by altering insulin-like growth factor 1 receptor glycosylation. *Mol Hum Reprod* **21**, 105-114 (2015).

32. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta-Gen. Subj.* **1426**, 239-257 (1999).

33. Hong, V., Steinmetz, N.F., Manchester, M. & Finn, M.G. Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconjugate Chem.* **21**, 1912-1916 (2010).

34. Shelbourne, M., Chen, X., Brown, T. & El-Sagheer, A.H. Fast copper-free click DNA ligation by the ring-strain promoted alkyne-azide cycloaddition reaction. *Chem. Commun.* **47**, 6257-6259 (2011).

35. Debets, M.F. et al. Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun.* **46**, 97-99 (2010).

36. 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).

37. 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).

38. 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).

39. 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).

40. Huttlin, E.L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010).

41. 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).

42. Hang, H.C., Yu, C., Kato, D.L. & Bertozzi, C.R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14846-14851 (2003).

43. Mercer, N., Ramakrishnan, B., Boeggeman, E., Verdi, L. & Qasba, P.K. Use of novel mutant galactosyltransferase for the bioconjugation of terminal n-acetylglucosamine (GlcNAc) residues on live cell surface. *Bioconjugate Chem.* **24**, 144-152 (2013).

44. Nandi, A. et al. Global identification of O-GlcNAc-modified proteins. *Anal. Chem.* **78**, 452-458 (2006).

45. Kaji, H. et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* **21**, 667-672 (2003).

46. Asperger, A., Marx, K., Albers, C., Molin, L. & Pinato, O. Low abundant N-linked glycosylation in hen egg white lysozyme is localized at nonconsensus sites. *J Proteome Res* **14**, 2633-2641 (2015).

47. 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).

48. Boyce, M. et al. 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. May, P., Bock, H.H., Nimpf, J. & Herz, J. Differential glycosylation regulates processing of lipoprotein receptors by gamma-secretase. *J Biol Chem* **278**, 37386-37392 (2003).

50. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

51. Liao, J.K. & Laufs, U. Pleiotropic effects of statins. *Annu Rev Pharmacol* **45**, 89-118 (2005).

52. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376-386 (2002).

53. Cantagrel, V. & Lefeber, D.J. From glycosylation disorders to dolichol biosynthesis defects: a new class of metabolic diseases. *J Inherit Metab Dis* **34**, 859-867 (2011).

54. Xiao, H.P., Chen, W.X., Tang, G.X., Smeekens, J.M. & Wu, R.H. Systematic investigation of cellular response and pleiotropic effects in atorvastatin-treated liver cells by MS-based proteomics. *J Proteome Res* **14**, 1600-1611 (2015).

55. Wu, R.H. et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods* **8**, 677-U111 (2011).

56. Serrano-Pozo, A. et al. Effects of simvastatin on cholesterol metabolism and alzheimer disease biomarkers. *Alz Dis Assoc Dis* **24**, 220-226 (2010).

57. Wong, W.B., Lin, V.W., Boudreau, D. & Devine, E.B. Statins in the prevention of dementia and Alzheimer's disease: A meta-analysis of observational studies and an assessment of confounding. *Pharmacoepidem Dr S* **22**, 345-358 (2013).

58. Ellul, J. et al. The effects of commonly prescribed drugs in patients with Alzheimer's disease on the rate of deterioration. *J Neurol Neurosur Ps* **78**, 233-239 (2007).

59. Schedin-Weiss, S., Winblad, B. & Tjernberg, L.O. The role of protein glycosylation in Alzheimer disease. *Febs J* **281**, 46-62 (2014).

60. Morais, V.A. et al. N-glycosylation of human nicastrin is required for interaction with the lectins from the secretory pathway calnexin and ERGIC-53. *Bba-Mol Basis Dis* **1762**, 802-810 (2006).

61. Finn, R.D. et al. Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-D230 (2014).

62. Lis, H. & Sharon, N. Protein glycosylation - structural and functional-aspects. *Eur. J. Biochem.* **218**, 1-27 (1993).

63. Pulsipher, A., Griffin, M.E., Stone, S.E. & Hsieh-Wilson, L.C. Long-lived engineering of glycans to direct stem cell fate. *Angew. Chem.-Int. Edit.* **54**, 1466-1470 (2015).

64. Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A. & Dwek, R.A. Glycosylation and the immune system. *Science* **291**, 2370-2376 (2001).

65. Takeuchi, H. & Haltiwanger, R.S. Role of glycosylation of Notch in development. *Semin. Cell Dev. Biol.* **21**, 638-645 (2010).

66. Bi, S.G. & Baum, L.G. Sialic acids in T cell development and function. *Biochim. Biophys. Acta-Gen. Subj.* **1790**, 1599-1610 (2009).

67. Gabius, H.J., Siebert, H.C., Andre, S., Jimenez-Barbero, J. & Rudiger, H. Chemical biology of the sugar code. *ChemBioChem* **5**, 740-764 (2004).

68. Kuball, J. et al. Increasing functional avidity of TCR-redirected T cells by removing defined N-glycosylation sites in the TCR constant domain. *J. Exp. Med.* **206**, 463-475 (2009).

69. Maverakis, E. et al. Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review. *J. Autoimmun.* **57**, 1-13 (2015).

70. Lehle, L., Strahl, S. & Tanner, W. Protein glycosylation, conserved from yeast to man: A model organism helps elucidate congenital human diseases. *Angew. Chem.-Int. Edit.* **45**, 6802-6818 (2006).

71. Holst, S. et al. N-glycosylation profiling of colorectal cancer cell lines reveals association of fucosylation with differentiation and caudal type homebox 1 (CDX1)/Villin mRNA expression. *Mol. Cell. Proteomics* **15**, 124-140 (2016).

72. Freeze, H.H. & Aebi, M. Altered glycan structures: the molecular basis of congenital disorders of glycosylation. *Curr. Opin. Struct. Biol.* **15**, 490-498 (2005).

73. Dotz, V. et al. Mass spectrometry for glycosylation analysis of biopharmaceuticals. *Trac-Trends Anal. Chem.* **73**, 1-9 (2015).

74. Marino, K., Bones, J., Kattla, J.J. & Rudd, P.M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **6**, 713-723 (2010).

75. Steentoft, C. et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977-982 (2011).

76. Hang, I. et al. Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi. *Glycobiology* **25**, 1335-1349 (2015).

77. Schubert, M., Walczak, M.J., Aebi, M. & Wider, G. Posttranslational modifications of intact proteins detected by NMR spectroscopy: application to glycosylation. *Angew. Chem.-Int. Edit.* **54**, 7096-7100 (2015).

78. Bie, Z.J., Chen, Y., Ye, J., Wang, S.S. & Liu, Z. Boronate-affinity glycan-oriented surface imprinting: a new strategy to mimic lectins for the recognition of an intact glycoprotein and its characteristic fragments. *Angew. Chem.-Int. Edit.* **54**, 10211-10215 (2015).

79. Hubbard, S.C., Boyce, M., McVaugh, C.T., Peehl, D.M. & Bertozzi, C.R. Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorg Med Chem Lett* **21**, 4945-4950 (2011).

80. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**, 2364-2369 (2001).

81. Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8256-8261 (2008).

82. Wormald, M.R. & Dwek, R.A. Glycoproteins: glycan presentation and protein-fold stability. *Struct. Fold. Des.* **7**, R155-R160 (1999).

83. Hanson, S.R. et al. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3131-3136 (2009).

84. Sola, R.J. & Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci-Us* **98**, 1223-1245 (2009).

85. Wang, C.Q., Eufemi, M., Turano, C. & Giartosio, A. Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35**, 7299-7307 (1996).

86. Yates, J.R., Ruse, C.I. & Nakorchevsky, A. in Annual Review of Biomedical Engineering, Vol. 11 49-79 (Annual Reviews, Palo Alto; 2009).

87. Rexach, J.E. et al. Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. *Nat. Chem. Biol.* **6**, 645-651 (2010).

88. Zielinska, D.F., Gnad, F., Schropp, K., Wisniewski, J.R. & Mann, M. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol Cell* **46**, 542-548 (2012).

89. Munoz, J. & Heck, A.J.R. From the human genome to the human proteome. *Angew. Chem.-Int. Edit.* **53**, 10864-10866 (2014).

90. Li, Y., Cross, F.R. & Chait, B.T. Method for identifying phosphorylated substrates of specific cyclin/cyclin-dependent kinase complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11323-11328 (2014).

91. Tan, Z.J. et al. Large-scale identification of core-fucosylated glycopeptide sites in pancreatic cancer serum using mass spectrometry. *J. Proteome Res.* **14**, 1968-1978 (2015).

92. Hwang, L. et al. Specific enrichment of phosphoproteins using functionalized multivalent nanoparticles. *J. Am. Chem. Soc.* **137**, 2432-2435 (2015).

93. Bausch-Fluck, D. et al. A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, 22 (2015).

94. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chemical Science* **6**, 4681–4689 (2015).

95. Ramya, T.N.C., Weerapana, E., Cravatt, B.F. & Paulson, J.C. Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans. *Glycobiology* **23**, 211-221 (2013).

96. Larsen, M.R., Jensen, S.S., Jakobsen, L.A. & Heegaard, N.H.H. Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Mol. Cell. Proteomics* **6**, 1778-1787 (2007).

97. Sun, S.S. et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat Biotechnol* **34**, 84-88 (2016).

98. Thompson, A. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895-1904 (2003).

99. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chemical Science* **7**, 1393-1400 (2016).

100. 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).

101. Kiick, K.L., Saxon, E., Tirrell, D.A. & Bertozzi, C.R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 19-24 (2002).

102. Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A. & Schuman, E.M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9482-9487 (2006).

103. Howden, A.J.M. et al. QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10**, 343-346 (2013).

104. Dieck, S.T. et al. Direct visualization of newly synthesized target proteins in situ. *Nat. Methods* **12**, 411-414 (2015).

105. Tsai, Y.H., Essig, S., James, J.R., Lang, K. & Chin, J.W. Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat. Chem.* **7**, 554-561 (2015).

106. Belardi, B. et al. Imaging the glycosylation state of cell surface glycoproteins by two-photon fluorescence lifetime imaging microscopy. *Angew Chem Int Edit* **52**, 14045-14049 (2013).

107. Grammel, M. & Hang, H.C. Chemical reporters for biological discovery. *Nat. Chem. Biol.* **9**, 475-484 (2013).

108. Yang, Y.Y., Ascano, J.M. & Hang, H.C. Bioorthogonal chemical reporters for monitoring protein acetylation. *J. Am. Chem. Soc.* **132**, 3640-3641 (2010).

109. Teckchandani, A. et al. Quantitative proteomics identifies a Dab2/integrin module regulating cell migration. *J. Cell Biol.* **186**, 98-110 (2009).

110. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

111. Eden, E. et al. Proteome half-life dynamics in living human cells. *Science* **331**, 764-768 (2011).

112. Zola, H. et al. CD molecules 2006 - human cell differentiation molecules. *J. Immunol. Methods* **319**, 1-5 (2007).

113. Seibel, J. et al. Investigating infection processes with a workflow from organic chemistry to biophysics: the combination of metabolic glycoengineering, super-resolution fluorescence imaging and proteomics. *Expert Rev. Proteomics* **10**, 25-31 (2013).

CHAPTER 5. GLOBAL AND SITE-SPECIFIC ANALYSIS REVEALING UNEXPECTED AND EXTENSIVE PROTEIN S-GLCNACYLATION IN HUMAN CELLS

Partially adapted with permission from American Chemical Society

Xiao, H. P., and Wu, R. H. Global and Site-Specific Analysis Revealing Unexpected and Extensive Protein S-GlcNAcylation in Human Cells. *Analytical Chemistry,* 2017, 89, 3656-3663.

5.1 Unexpected Observation of Protein S-GlcNAcylation in Human Cells

5.1.1 Introduction

Glycosylation is one of the most common and diverse protein modifications, and is essential for mammalian cell survival.[1-3] Heterogeneous protein glycosylation contains a wealth of information regarding the cellular developmental and diseased statuses. Aberrant protein glycosylation is directly related to human disease, such as cancer and infectious diseases.[4, 5] However, heterogeneity of glycosylation renders their study much more challenging compared to many other types of protein modification.[6, 7] Investigation of protein glycosylation can aid in a better understanding of protein function, cellular activity, and the molecular mechanisms of disease.

N-acetylglucosamine (GlcNAc) was discovered to be bound to the side chains of serine and threonine over three decades ago,[8] termed O-GlcNAcylation, and was regulated by O-GlcNAc transferase (OGT)[9] and O-GlcNAc amidase (OGA).[10] This modification is involved in many cellular events, from regulation of signal transduction to gene expression.[10, 11] The same glycan (GlcNAc) has recently been found to be attached to the side chain of cysteine in

bacteria,[12] which is named as protein S-GlcNAcylation. Glycocin F, a 43-amino acid bacteriocin from *Lactobacillus plantarum*, contains two beta-linked GlcNAc moieties, attached through side chain linkages to a serine *via* oxygen, and to a cysteine *via* sulfur.[12] This modification has just been reported in rat and mouse samples, and 14 modification sites in 11 proteins were identified.[13] In the same report, recombinant Host Cell Factor 1 isolated from HEK cells was identified to be S-GlcNAcylated.[13] Compared to O-GlcNAcylation, S-GlcNAcylation remains to be explored.

Modern mass spectrometry (MS)-based proteomics provides a unique opportunity to comprehensively and site-specifically analyze protein modifications,[14-26] which is beyond the reach of conventional biochemistry methods. Due to the fact that many modified proteins have low abundance, it is critical to separate and enrich proteins or peptides with the modified group of interest prior to MS analysis.[27-32] In order to globally analyze protein O-GlcNacylation, several methods were reported to enrich O-GlcNAcylated proteins/peptides from complex biological samples. Mild beta-elimination followed by Michael addition with dithiothreitol (BEMAD) was employed for global analysis of protein O-GlcNAcylation.[33] Lectins or antibodies were also used to enrich glycoproteins/glycopeptides for MS analysis. [14, 34, 35] An elegant chemoenzymatic approach was developed by exploiting an engineered galactosyltransferase enzyme to selectively label O-GlcNAc proteins with a ketone-biotin tag, which permits enrichment of low-abundance O-GlcNAc species from complex mixtures.[36, 37] The enzymatic reaction combined with click chemistry was further developed to enrich O-GlcNAcylated peptides for MS analysis.[15, 38] A sugar analog (N-azidoacetylglucosamine (GlcNAz)) was also used to feed cells and label glycoproteins for visualization and MS analysis. [39, 40] Currently bio-orthogonal chemistry is very powerful in investigating proteins and protein modifications.[41-44] In combination with metabolic labeling and MS-based proteomics, global analysis of protein modification may be achieved. In order to analyze protein modification site-

specifically, an effective cleavable linker is required, and the tag after cleavage must be relatively small to be compatible with MS analysis. Site-specific analysis provides not only valuable information about the modification site but also solid experimental evidence of protein modification.

In this work, unexpected S-GlcNAcylation on cysteine residues was demonstrated to extensively exist in human cells through global and site-specific analysis of protein GlcNAcylation. This result was further confirmed by different independent experiments. Motif analysis showed that the modified cysteine sites surrounded with an acidic amino acid residue (D or E) are highly enriched, which strongly suggests that a particular type of enzyme is responsible for this modification and has a preference for the sites surrounded by acidic amino acids. Protein clustering results showed that glycoproteins with well-localized S-GlcNAcylation sites are involved in the regulation of cell-cell interactions and gene expression. For the first time, the global and site-specific analysis unraveled extensive protein S-GlcNAcylation existing in human cells.

5.1.2 Experimental section

5.1.2.1 Cell culturing and metabolic labeling

MCF-7 cells (from American type culture collection (ATCC)) were grown in a humidified incubator at 37 °C and 5.0% CO_2 in Dulbecco's modified eagle's medium (DMEM) (Sigma-Aldrich) containing 10% fetal bovine serum (FBS) (Thermo). When cells reached ~60% confluency, the media was switched to DMEM containing 10% dialyzed FBS with 250 μM N-azidoacetylglucosamine-tetraacylated (GlcNAz) (Click Chemistry Tools). Cells were metabolically labeled in this media for 36 h and then treated with 50 μM O-(2-Acetamido-2-deoxy-D-glucopyranosylidenamino)N-phenylcarbamate (PUGNAc) (Cayman Chemicals) for two hours before harvest.

5.1.2.2 Cell lysis, Copper-catalyzed azide alkyne cycloaddition (CuAAC), and protein digestion

Cells were washed three times with phosphate buffered saline (PBS) to remove GlcNAz and then harvested by scraping. The cell mixtures were pelleted by centrifugation at 300 g for 5 minutes and washed twice with cold PBS. Cell pellets were lysed through end-over-end rotation at 4 ºC for 45 minutes in lysis buffer (50 mM HEPES pH=7.4, 150 mM NaCl, 0.5% sodium deoxycholate (SDC), 25 units/mL benzonase, 100 μM PUGNAc, and 1 tablet/ 10 mL protease inhibitor (Roche)). Lysates were centrifuged, and the resulting supernatant was transferred to new tubes to perform CuAAC. Briefly, bis-N-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl] (DDE)-biotin-alkyne (Click Chemistry Tools) was dissolved in DMSO and added to the cell lysate to a final concentration of 250 μM. For the experiment using the photocleavable (PC) linker, PC-biotin-alkyne (Click Chemistry Tools) was added into the cell lysate to the same final concentration. At the same time, $CuSO_4$ and tris(3-hydroxypropyltriazolylmethyl) amine (THPTA) were added to the lysate to final concentrations of 1 mM and 5 mM, respectively. Finally, sodium ascorbate was freshly prepared and added to the lysis mixture at a concentration of 15 mM to initiate the reaction. The reaction vessel was covered with aluminum foil, and then placed onto an end-over-end rotor at room temperature, and rotated for 2 h. Since DDE-biotin-alkyne is cleavable by strong reducing reagent such as dithiothreitol (DTT), reduction and alkylation of disulfide bonds were not performed. SDC was removed by the methanol-chloroform protein precipitation method. The purified proteins were digested with 10 ng/μL Lys-C (Wako) in 50 mM HEPES pH 8.6, 1.6 M urea, 5% ACN at 31 ºC for 16 hours, followed by further digestion with 8 ng/uL Trypsin (Promega) at 37 ºC for 4 hours.

5.1.2.3 Glycopeptide separation and enrichment

Digestion mixtures were acidified by the addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, clarified by centrifugation, and desalted by using a tC18 Sep-Pak cartridge (Waters). Purified peptides were dried and then enriched with NeutrAvidin beads (Thermo) at 37 °C for 30 min in 100 mM PBS. The samples were transferred to spin columns and washed thoroughly according to the manufacturer's protocol. Peptides were then eluted from the beads through incubation with 2% hydrazine in 100 mM sodium phosphate for 2 hours, and then washed twice with the elution buffer. Eluates were combined and acidified by adding TFA, desalted using tC18 Sep-Pak cartridge, and lyophilized. Dried peptides were dissolved in 1% formic acid (FA), purified further with a stage-tip, and separated into 3 fractions using 20%, 50% and 80% ACN containing 1% HOAc.

5.1.2.4 LC-MS/MS analysis

Purified and dried peptide samples were dissolved in a 9 μL solution of 5% ACN and 4% FA each. 4 μL of the resulting solutions were loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 μm, 200 Å, 100 μm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 110 min gradient of 3-22%, 8-35%, 12-45% ACN (with 0.125% FA), respectively, for three fractions. Peptides were detected with a data-dependent Top20 method in a hybrid dual-cell quadrupole linear ion trap - Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 20 MS/MS in the LTQ for the most intense ions. The selected ions were excluded from further

analysis for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

5.1.2.5 Database search and data filtering

All MS^2 spectra were converted into mzXML files, and then searched using the SEQUEST algorithm (version 28).[45] Spectra were matched with sequences of all proteins in the UniProt Human (*Homo sapiens*) database. The following parameters were used during the search: 10 ppm precursor mass tolerance, 1.0 Da product ion mass tolerance, fully digested with trypsin, up to three missed cleavages. The variable modifications: oxidation of methionine (+15.9949), tag of GlcNAcylation on serine, threonine, and cysteine (+299.12297). False discovery rates (FDR) of peptide and protein identifications were evaluated and controlled by the target-decoy method.[46, 47] Each protein sequence was listed in both forward and reverse orders. Linear discriminant analysis (LDA) was used to control the quality of peptide identifications using the following parameters: XCorr, differential sequence dCN, missed cleavages, ppm, precursor mass error, peptide length, and charge states. Peptides shorter than seven amino acid residues in length were deleted. Furthermore, peptide spectral matches were filtered to 1% FDR. The dataset was restricted to GlcNAcylated peptides when determining FDR for glycopeptide identification. In order to increase the identification confidence, we further filtered the dataset to less than 1% FDR at the glycoprotein level, and in this case, the FDR for glycopeptide identification is much less than 1%.

5.1.2.6 GlcNAcylation site localization and quality control

We assigned and measured the confidence of glycosylation site localizations by calculating the corresponding ModScores, which applies a probabilistic algorithm that considers all possible glycosylation sites in a peptide and uses the presence of experimental fragment ions unique to each site.[48, 49] Sites with ModScore > 13 ($P < 0.05$) were considered

as confidently localized. All sites and peptides were then manually checked; sites on peptides with low mass accuracy (>5 ppm) or low XCorr (<1.5) were manually removed to ensure the quality of our results.

5.1.2.7 Motif Analysis

In the motif analysis, only well-localized S-GlcNAcylation sites were used, and sequences were centered on each site, extended to 13 aa (6 residues on each side of the site), and analyzed with the Motif-X algorithm.[50] The number of occurrences is set to at least 20, and the significance is 0.0001. The *Homo sapiens* protein database was used as a background.

5.1.2.8 Data Availability

All raw files are accessible in the following public accessible website (http://www.peptideatlas.org/PASS/PASS00981, pass- word: MV9545qx).

5.1.3 Results and discussion

5.1.3.1 Principle of the enrichment of GlcNAcylated peptides

Usage of sugar analogs to label cells has recently been proven to be effective for protein glycosylation investigation, which has been well documented.[51, 52] Labeled glycoproteins with sugar analogs can be bound to fluorophoric groups for visualization and also be enriched for further analysis by MS.[53-55] Here, we used a sugar analog (Ac_4GlcNAz) to label glycoproteins in MCF-7 cells, as shown in Figure 5.1. After cell lysis and protein extraction, we incubated proteins with DDE-biotin-alkyne with Cu(I) as the catalyst.[56] GlcNAcylated proteins with the functional azido group were tagged with biotin. After digestion, glycopeptides tagged with biotin were enriched with NeutrAvidin beads. After removing non-glycopeptides, we released the enriched glycopeptides using hydrazine (N_2H_4). The cleavable linker was cleaved and an

amine group was left on glycopeptides for MS analysis. Finally, the mass tag of 299.1230 Da is well-suited for glycopeptide identification and site localization by LC-MS/MS.

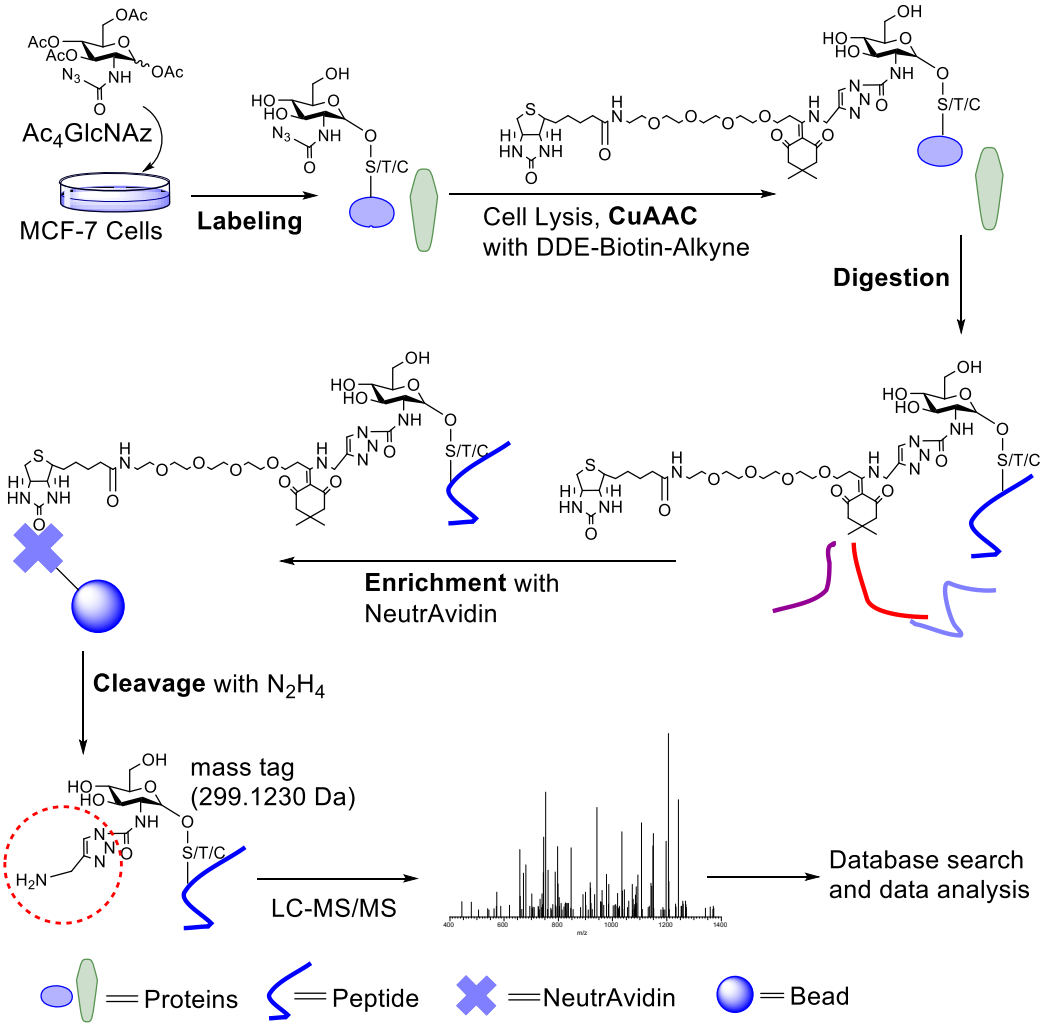


Figure 5.1 Experimental procedure of selectively enriching GlcNAcylated peptides for MS analysis (The curves with different colors represent peptides).

5.1.3.2 Integration of a cleavable linker for site-specific identification of protein GlcNAacylation

To perform site-specific analysis of protein GlcNAcylation, a cleavable linker is required, which is different from the analysis of protein N-glycosylation because N-glycans can be removed with an enzyme, such as PNGase F, to generate a common tag for MS analysis.[27, 57] Ideally, the cleavable linker needs to be compatible with the experimental

conditions, including click reaction conditions. Furthermore, after cleavage, the remaining tag should be small for MS analysis. Cleavable linkers based on disulfide bond are extensively reported in the literature.[58] However, this type of cleavable linker is incompatible with the current experiment because the disulfide bond may not survive under reductive conditions of the Cu(I)-catalyzed alkyne-azide cycloaddition (CuAAC) reaction.

Here a DDE-based cleavable linker was used due to its advantages. First, the linker is stable under reductive CuAAC conditions. Second, after the cleavage, the remaining tag is small with a mass of 95 Da, as marked in a red circle in Figure 5.1. Third, the tag containing an amine group will increase the ionization efficiency of glycopeptides, facilitating the MS analysis of glycopeptides. In addition, with this specific sugar analog and the remaining tag, the final mass (299.1230 Da) can distinguish this modification from any other modifications. This cleavable linker allows us to site-specifically identify protein GlcNAcylation, and the tag on the peptides also provides solid experimental evidence for the protein modification.

5.1.3.3 Identification of protein GlcNAcylation

By using the cleavable linker, a small tag on glycopeptides enabled us to confidently identify glycopeptides and localize the glycosylation sites. Two examples are shown in Figure 5.2. The glycopeptide VS#VCAETYNPDEEEEDTDPR (# - glycosylation site) was identified with a mass accuracy of 0.98 ppm and XCorr of 3.86. This peptide is from protein PRKAR2A, which is a regulatory subunit of the cAMP-dependent kinases.

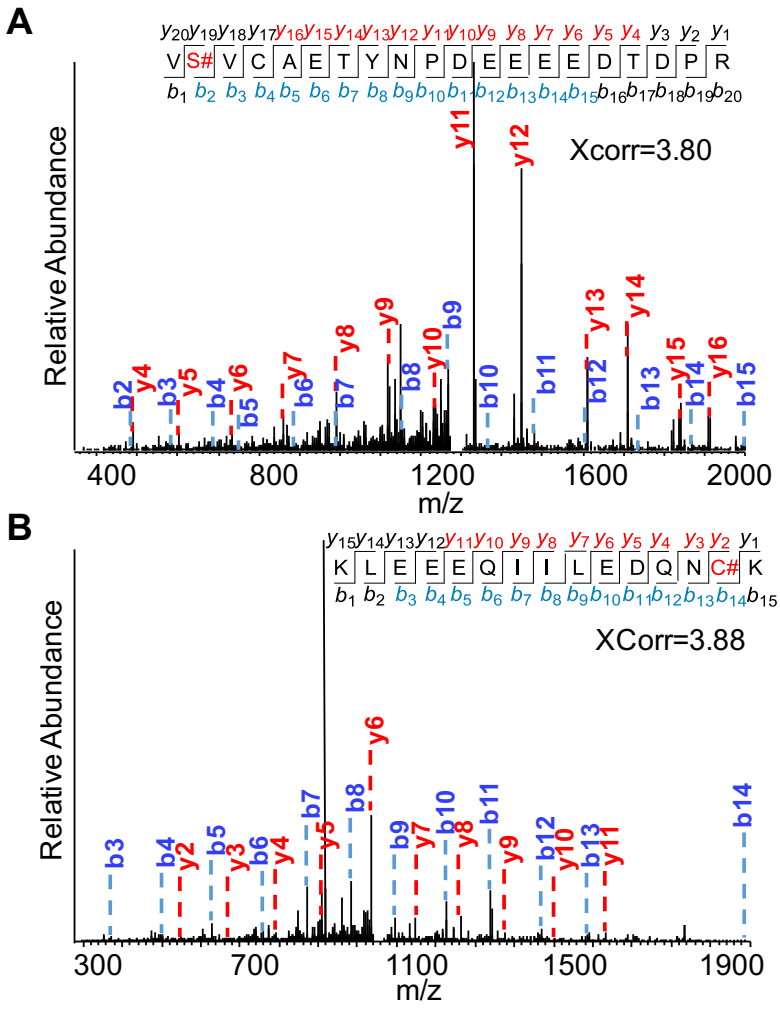


Figure 5.2 Examples of glycopeptide identification. (A) The peptide VS#VCAETYNPDEEEEDTDPR (# - glycosylation site) was identified, which is from protein PRKAR2A. (B) The peptide KLEEEQIILEDQNC#K from Myh9 was confidently identified with an XCorr of 3.88 and mass accuracy of 1.45 ppm, and the site C1002 was bound to the glycan.

Unexpectedly, we identified many sites located on cysteine, which is discussed in details below, and one example is shown in Figure 5.2B. The glycopeptide KLEEEQIILEDQNC#K was confidently identified with an XCorr of 3.88 and mass accuracy of 1.45 ppm. It is from protein MYH9, and GlcNAc is bound to the site of cysteine 1002. Furthermore, the lack of any serine or threonine residue in this peptide excludes the possibility of protein O-GlcNAcylation. From the fragments, the tag neutral loss occurred (the highest

peak in Fig. 4.2B). Nevertheless, there was enough energy left to further fragment the peptide for confident identification.

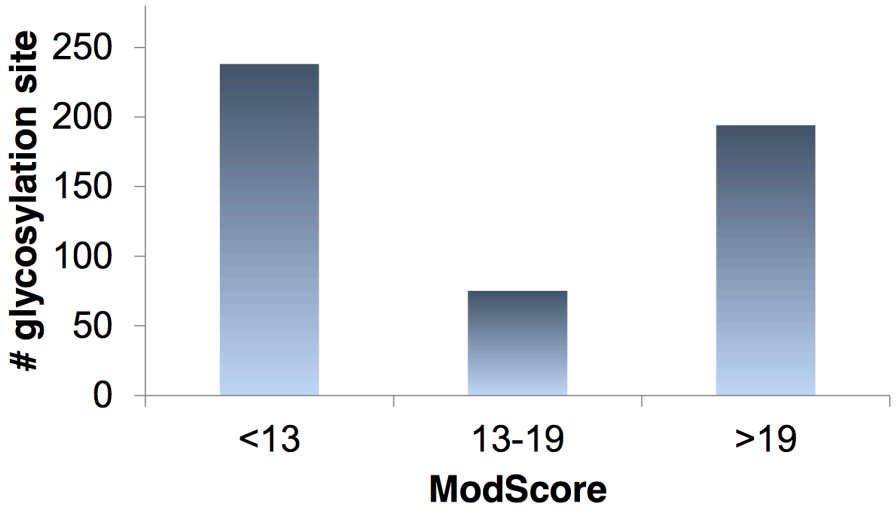


Figure 5.3 ModScore distribution of the GlcNAcylation sites identified in the DDE experiment.

In this experiment, we identified 537 unique glycopeptides, and correspondingly 507 sites were assigned on 367 proteins (listed in a table online at doi.org/10.1021/acs.analchem.6b05064). Among 507 sites, 269 were well-localized with ModScore>13 (Figure 5.3). The portion of sites being well localized is relatively low because of the following reasons. First, the sugar (GlcNAc) was bound to peptides, thus the neutral loss frequently occurred under collision induced dissociation (CID), as shown in Figure 5.2, which results in the relatively low rate of the site localization. Furthermore, there are multiple possible glycosylation sites, i.e. STC, and many peptides contain several of these residues. Sometimes only two fragments can distinguish two possible sites nearby, making the site localization more challenging.

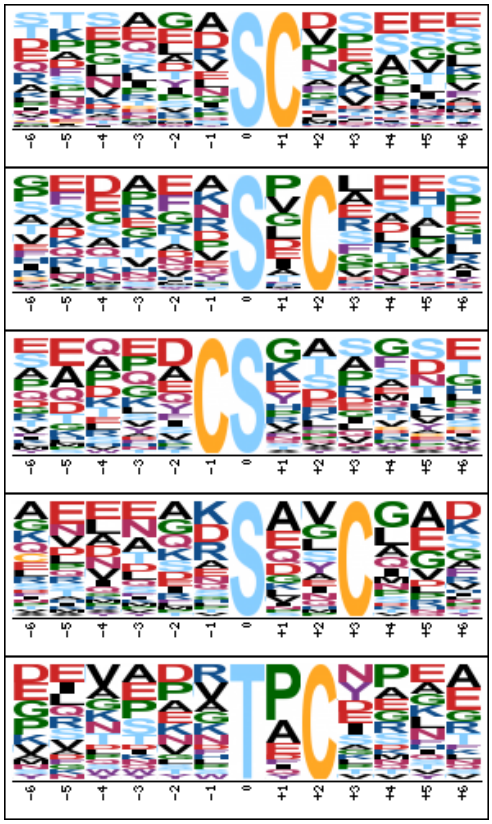


Figure 5.4 Motifs identified from the well-localized protein O-GlcNAcylation sites (ModScore>13), using only ST as possible modification sites to perform SEQUEST search.

4.1.3.4 Confident identification of protein S-GlcNAcylation on cysteine residues

Initially, we used ST as dynamic modification sites for SEQUEST search because it is well-known that protein O-GlcNAcylation occurs on serine and threonine. Subsequently, we observed that cysteine is commonly located around S or T. When we conducted motif analysis, several motifs with cysteine around the sites were highly enriched, as shown in Figure 5.4. This results in a consideration that the cysteine residue may be modified with GlcNAc, and correspondingly, we used STC as dynamic modification sites to perform the search. Surprisingly, we found many sites (even greater than S and T, see below) were located on cysteine. In order to demonstrate the reliability of S-GlcNAcylation, we further performed the following experiments and data analysis.

1). S-GlcNAc is not from possible chemical reaction during sample preparation: The free thiol group of cysteine is reactive and could undergo reactions during sample preparation. To avoid the potential reactions during the sample preparation, we alkylated free thiol groups for the subsequent experiment by adding 1 mM IAA in the solution 20 minutes before and during the two-hour click reaction. Similar results were obtained, including the number of total and unique glycopeptides with S-GlcNAc. The comparison of the results is included below.

2). S-GlcNAc is not produced by the cleavable linker: In order to further confirm that S-GlcNAcylation is authentic, we employed a completely different type of cleavable linker, photocleavable linker (the structure is in Figure 5.5). After enrichment, glycopeptides were cleaved under UV radiation at 350 nm for one hour. Because the remaining tag is the same, it has all the advantages described above. Similarly, we identified many glycopeptides with S-GlcNAc, as compared below.

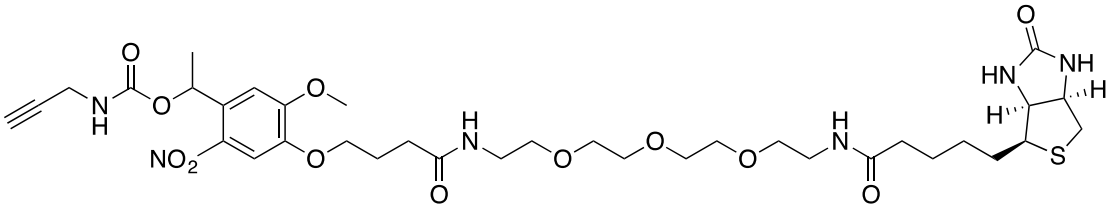


Figure 5.5 The structure of the photocleavable (PC) linker. After enrichment, the linker was cleaved using radiation at 350 nm for one hour, which generates the same tag as the DDE linker.

3). S-GlcNAc is not due to false assignment: To exclude the possibility of wrong assignment, we checked whether there are some identified glycopeptides without any S or T. Among the glycopeptides identified in the DDE experiment, 26 unique glycopeptides did not contain any serine or threonine. Among 269 well-localized glycosylation sites, 199 sites (74%) were localized on cysteine while only 20 sites were located on threonine and 50 sites on serine

(Figure 5.6). These results further demonstrated that S-GlcNAcylation is not due to wrong assignment.

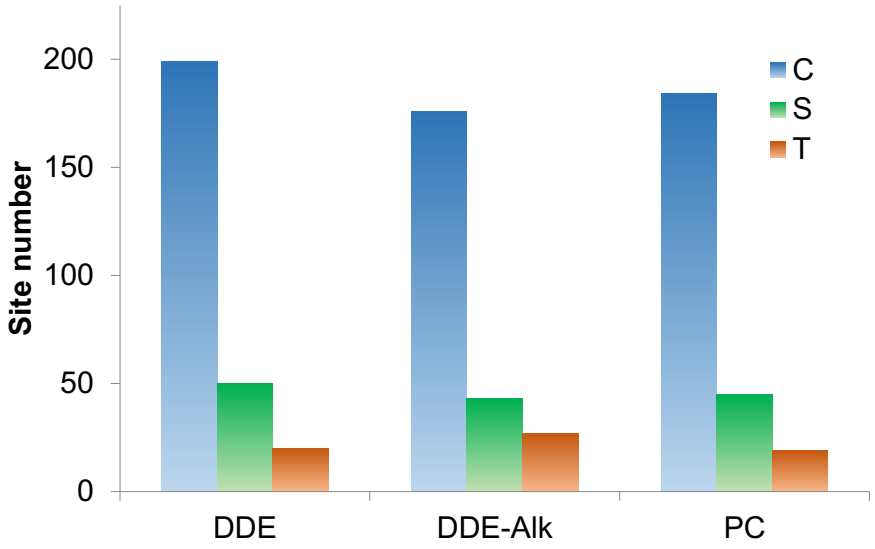


Figure 5.6 Distributions of well-localized sites on cysteine (C), serine (S) and threonine (T) in three independent experiments (DDE, DDE-Alk and PC).

5.1.3.5 Comparison of glycopeptides identified in three independent experiments

Among the three experiments (DDE-biotin-alkyne without alkylation (DDE), with alkylation (DDE-Alk), and photocleavable linker (PC)), we obtained very comparable results. In the DDE experiment, 74% of well-localized sites (199/269) were located on cysteine. With alkylation (DDE-Alk) experiment, 71% sites (174/244) belonged to S-GlcNAcylation (All identified sites are listed in a table online at doi.org/10.1021/acs.analchem.6b05064). In the PC experiment, 247 well-localized sites were identified (listed in a table online at doi.org/10.1021/acs.analchem.6b05064), and most of them (183 sites, 74%) were also located on cysteine (Figure 5.6). From the independent experiments, the number of protein S-GlcNAcylation sites from each experiment was very similar. Well-localized S-GlcNAcylation sites in three independent experiments were listed in a table online at doi.org/10.1021/acs.analchem.6b05064 and compared in Figure 5.8A. In addition,

glycoproteins with well-localized sites were compared in Figure 5.7. Relatively, the overlap was not exceptionally high, but this is also common for large-scale analysis of protein modification considering that the modification may be dynamic and these are biologically independent experiments with different linker molecules and the presence or absence of iodoacetamide.

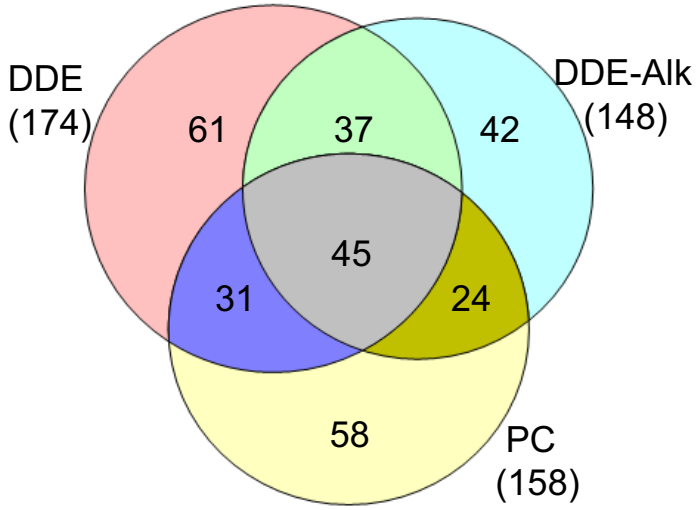


Figure 5.7 Comparison of glycoproteins with well-localized S-GlcNAcylation sites identified in three independent experiments.

5.1.3.6 Motif Analysis of Well-Localized S-GlcNAcylation Sites

Motif analysis was performed only for well-localized S- GlcNAcylation sites (357), among which 45 sites contain the motif of EC, and 40 sites with a motif of CXE (X - any amino acid residue). In addition, there are 26 sites with a cysteine followed by an aspartic acid residue, and 29 sites have a motif of CXXXE. As shown in Fig. 4.8B, the sites surrounded by an acidic amino acid are highly enriched. Protein modification can result from enzymatic or chemical reactions. For example, O-GlcNAcylation is catalyzed by OGT, but protein glycation is caused

by a chemical reaction. It is uncertain whether S-GlcNAcylation is due to chemical or enzymatic reactions. However, the identification of these motifs strongly suggests that some enzymes are responsible for this type of modification, and the enzyme's binding pocket has a preference for acidic residues. It will be of great interest to identify the enzyme(s) responsible for S-GlcNAcylation.

5.1.3.7 Clustering of Proteins Modified with Well-Localized S-GlcNAc

To understand possible functions of protein S-GlcNAcylation, we clustered glycoproteins with well-localized S-GlcNAcylation sites using the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.8).[59] Based on cellular compartment, the most highly enriched categories are cell-cell adherens junction, nuclear part, organelle lumen, and intracellular membrane-bounded organelle (Fig. 4.9A). 40 identified glycoproteins are located in the cell-cell adherens junction with an extremely low *P* value of 6.46E-23. Almost half of the proteins (137) are nuclear proteins with a *P* value of 3.65E-22. There are 27 proteins belonging to the chromosomal part, and 9 proteins are from the SWI/SNF (SWItch/Sucrose Non-Fermentable) superfamily-type complex, listed in Table 5.1. SWI/SNF is a nucleosome remodeling complex that exists in both eukaryotes and prokaryotes. However, the exact action mode of this complex remains to be further explored. Protein S-GlcNAcylation may contribute to its function. Based on the clustering results, two major functions are related to these glycoproteins modified with S-GlcNAc: the regulation of cell-cell interactions and gene expression, which likely are the major functions of protein S-GlcNAcylation. From DAVID, among 298 glycoproteins, most of them (260) are also phosphoproteins, and 164 glycoproteins can be acetylated. This modification may cross-talk with acetylation and phosphorylation to regulate gene expression.

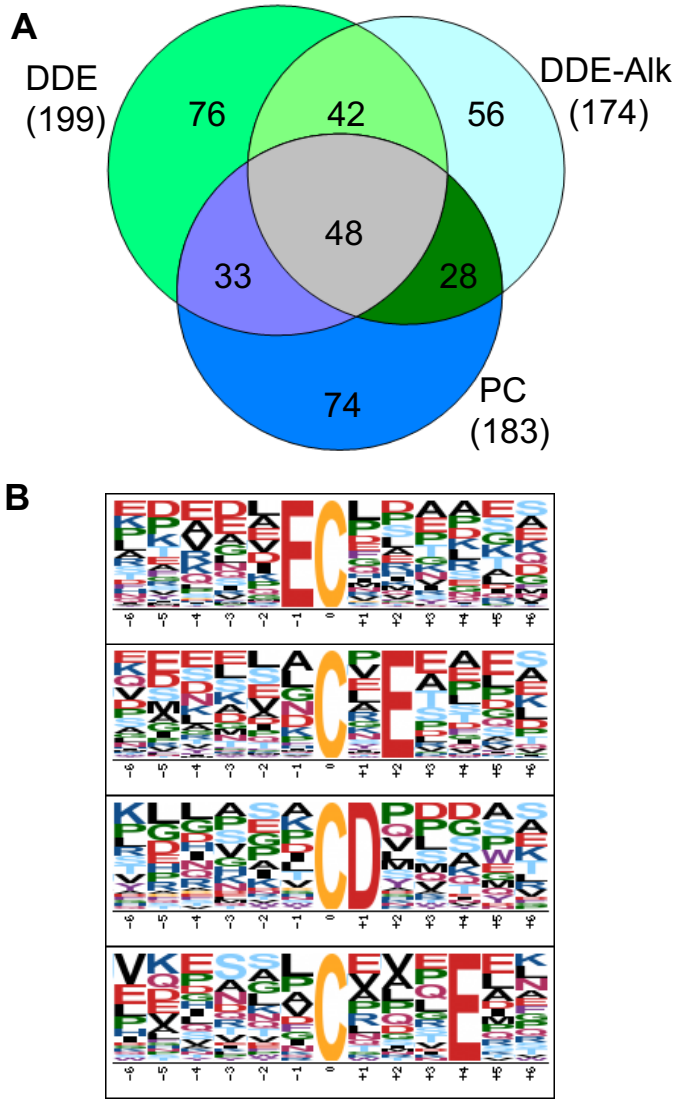


Figure 5.8 Comparison of (A) well-localized S-GlcNAcylation sites; (B) Four motifs were identified among the well-localized S-GlcNAcylation sites.

The clustering results based on molecular function are displayed in Fig. 4.9B. Nearly half of the glycoproteins (131) are nucleic acid binding proteins with a *P* value of 2.51E-15. Furthermore, 16 proteins have transcription cofactor activities, and 8 proteins are bound to histones. 39 proteins belong to the category of binding proteins involved in cell adhesion with a *P* value of 4.78E-22. These results further consolidate the possible functions related to this type of modification: the regulation of cell-cell interactions and gene expression.

Table 5.1 The identification of S-GlcNAcylation sites in glycopeptides from the SWI/SNF superfamily-type complex.

Reference	Gene symbol	Glycopeptide	PPM	XCorr	Site	Mod Score	Annotation
Q9NRG0	CHRAC1	ATELFVQC#LATYSYR	2.58	2.86	55	38.2	Chromatin accessibility complex protein 1
Q15532	SS18	MLDDNNHLIQC#IMDSQNK	0.15	2.72	34	31.3	Appears to function synergistically with RBM14 as a transcriptional coactivator
Q96T23	RSF1	EVVEC#QSTSTVGGQSVK	1.34	3.09	650	51.0	Remodeling and spacing factor 1
Q12830	BPTF	C#APAPPPPPPPPTSGPIGGLR	0.36	2.25	20	85.3	Nucleosome-remodeling factor subunit BPTF
Q9NPI1	BRD7	RKPDGTTTLGLLHPVDPIVGEPGYC#PVR	0.96	4.06	367	80.9	Bromodomain-containing protein 7, acts both as coactivator and as corepressor and may play a role in chromatin remodeling.
		DGTDTSQSGEDGGC#WQR	-0.46	2.68	271	33.7	
O14497	ARID1A	GPADMASQC#WGAAAAAAAAAASGGAQQR	-0.13	3.68	336	14.2	AT-rich interactive domain-containing protein 1A, involved in transcriptional activation and repression of select genes by chromatin remodeling
Q9Y265	RUVBL1	EVYEGEVTELTPC#ETENPMGGYGK	1.15	3.70	141	30.8	RuvB-like 1, possesses single-stranded DNA-stimulated ATPase and ATP-dependent DNA helicase (3' to 5') activity
Q16576	RBBP7	VHIPNDDAQFDASHC#DSDKGEFGGFGSVTGK	-0.15	3.76	97	16.1	Core histone-binding subunit that may target chromatin remodeling factors, histone acetyltransferases and histone deacetylases
Q14839	CHD4	FAEVEC#LAESHQHLSK	-1.34	3.46	1827	31.3	Chromodomain-helicase-DNA-binding protein 4

#-glycosylation site

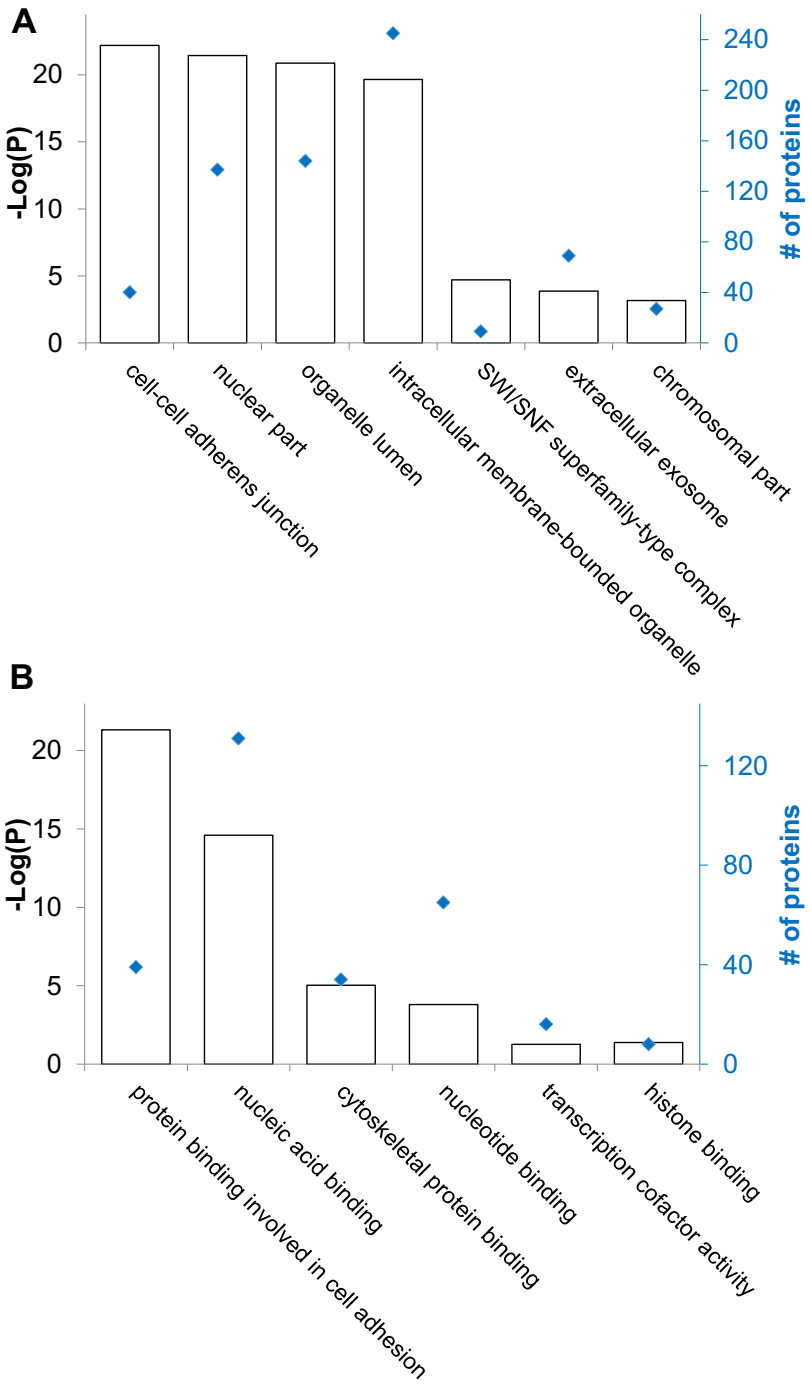


Figure 5.9 Clustering of glycoproteins with well-localized S-GlcNAcylation sites based on cellular compartment (A) and molecular function (B). The right y axis is –Log(P) and the left one is the protein number.

5.1.4 Conclusions

In this work, global and site-specific analysis of protein GlcNAcylation reveals unexpected S-GlcNAcylation on cysteine residues extensively existing in human cells. This result was further confirmed by different independent experiments. Motif analysis showed that the modified cysteine sites surrounded with an acidic amino acid residue (D or E) are highly enriched, which strongly suggests that a particular type of enzyme is responsible for this modification and has a preference for the sites surrounded by acidic amino acids. Protein clustering results demonstrated that glycoproteins with well-localized S-GlcNAcylation sites are involved in the regulation of cell-cell interactions and gene expression. For the first time, we discovered extensive protein S-GlcNAcylation existing in human cells through global and site-specific analysis of protein GlcNAcylation. Further work remains to be performed for a better understanding of protein S-GlcNAcylation, including the identification of possible enzymes responsible for this type of modification and illumination of the functions of protein S-GlcNAcylation.

5.2 Exploring Protein S-GlcNAcylation with Different Sugar Analog Labelling and in Various Types of Human Cells

5.2.1 Introduction

Protein glycosylation is highly diverse because of the heterogeneity of glycans and multiple side chains of amino acid residues being glycosylated.[6, 60, 61] It regulates many cellular events and is essential for mammalian cell survival.[5, 62-64] Aberrant protein glycosylation events are hallmarks of human diseases, such as cancer and infectious diseases.[65-69] Among several types of protein glycosylation, the dynamic posttranslational attachment of β-N-acetylglucosamine (GlcNAc) to the side chains of serine and threonine via O-linkage is termed O-GlcNAcylation,[70] which modifies numerous cytoplasmic and nuclear proteins and regulates their activity, stability, and localization.[71-73] Protein O-GlcNAcylation was discovered by Torres and Hart over three decades ago,[8] and since then many O-GlcNAc-centered studies have been conducted.[37, 38, 74-79]

Compared to O-GlcNAcylation, the attachment of GlcNAc onto the side chain of cysteine, i.e. S-GlcNAcylation, is nearly unexplored with few publications in the literature. In 2011, S-GlcNAc was found to attach to Cys43 of Glycocin F, a bacteriocin from *Lactobacillus plantarum* through S-linkage.[12] Last year, Maynard et al. reported S-GlcNAcylation events in mice samples through performing mass spectrometric analysis using electron transfer dissociation (ETD).[13] Recently synthetic S-GlcNAc was reported to be an enzymatically stable and structurally reasonable surrogate for O-GlcNAc at the peptide and protein levels.[80]

Recent advancements in mass-spectrometry (MS)-based proteomic techniques have provided a unique opportunity to globally and site-specifically investigate protein modifications.[11, 17-21, 25, 81-87] Due to the low abundance of glycoproteins, it is imperative to separate and enrich the modified proteins or peptides prior to MS analysis.[16, 27-32, 55, 88, 89] In

recent years, the combination of metabolic labeling and click chemistry has been proven to be very powerful to enrich modified proteins or peptides for global analysis of protein modifications with MS-based proteomics.[39, 44, 54, 57, 90-93] In our previous work, by combining metabolic labelling, click chemistry and cleavable linkers with MS-based proteomics, we unexpectedly found that S-GlcNAcylation extensively existed in human cells.[94] Several biologically independent experiments were performed to confirm that it was not produced by side reactions of the thiol groups during sample preparation, the cleavable linker utilized in the experiments, or false site assignments.[94] Further investigations will aid in a better understanding of protein S-GlcNAcylation.

In this work, we explored protein S-GlcNAcylation by evaluating the metabolic labelling with two different sugar analogs, namely GalNAz and GlcNAz, for global analysis of protein O- and S-GlcNAcylation. Furthermore, S-GlcNAcylation was systematically investigated in three types of human cells (MCF7, HEK 293T, and HeLa). The percentages of S-GlcNAcylation sites among the total well-localized sites are very similar across all these cells with over 70% of the well-localized sites being located on cysteine. We further performed motif and domain analysis for the well-localized S-GlcNAcylation sites. Together these data provide a systematic view of protein S-GlcNAcylation events in human cells.

5.2.2 Experimental section

5.2.2.1 Cell culturing and metabolic labelling

MCF-7, HEK 293 T, and HeLa cells are from American type culture collection (ATCC) and were seeded in T-75 culturing flasks upon thawing. Cells were grown in a humidified incubator at 37 °C and 5.0% CO_2 in Dulbecco's modified eagle's medium (DMEM) (Sigma-Aldrich) containing 10% fetal bovine serum (FBS) (Thermo). When cells reached ~60% confluency, the media was switched to DMEM containing 10% dialyzed FBS with 100 μM

GalNAz or GlcNAz (Click Chemistry Tools) for metabolic labelling. Cells were labeled for 36 h and then treated with 50 μM O-(2-acetamido-2-deoxy-D-glucopyranosylidenamino)N-phenylcarbamate (PUGNAc) (Cayman Chemicals) for two hours before harvest.

5.2.2.2 Sample preparation, LC-MS/MS analysis and data processing

See section 5.1.2.

5.2.3 Results and discussion

5.2.3.1 Experimental procedure of GlcNAcylation site identification

Here we used the method integrating metabolic labelling, click chemistry and a cleavable linker to enrich GlcNAcylated peptides and generate a small tag on GlcNAcylation sites. Then the enriched glycopeptides were analyzed using MS-based proteomics, as shown in Figure 5.10. Cells were labeled with either GalNAz or GlcNAz, and then lysed. Labeled proteins in the cell lysate were reacted with bis-N-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl] (DDE)-biotin-alkyne through copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) to incorporate a biotin tag for affinity enrichment.[56] Proteins were digested, and the labeled peptides were enriched with NeutrAvidin beads. Non-specific binding peptides were removed with stringent washes and the GlcNAcylated peptides were released through cleaving the DDE group with hydrazine (N_2H_4). This resulted in a mass tag of 299.1230 Da on the GlcNAcylation sites, which can be used for glycopeptide identification and glycosylation site localization. A total of four experiments were performed: GalNAz labelling of MCF7 cells (GAlNAz-MCF7), GlcNAz labelling of MCF7 cells (GlcNAz-MCF7), GlcNAz labelling of HEK 293T cells (GlcNAz-HEK 293T), and GlcNAz labelling of HeLa cells (GlcNAz-HeLa).

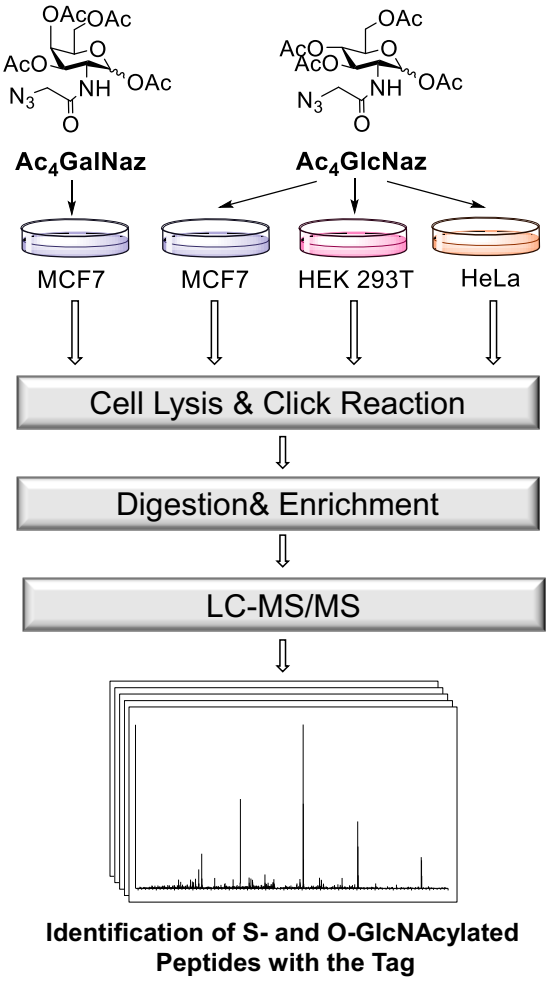


Figure 5.10 Experimental procedure for the chemoproteomic analysis of protein GlcNAcylation.

We previously demonstrated that S-GlcNAcylation was not from side reactions during sample preparation, produced by the cleavable linker, or due to false assignment.[94] Serine and threonine are among the most frequent amino acid residues in proteins, and the percentages of S and T residues in glycopeptides are even higher, rendering the site localization more difficult. Therefore, the ModScores of the GlcNAcylated sites identified in proteomics experiments are generally lower than those for N-glycosylation sites.

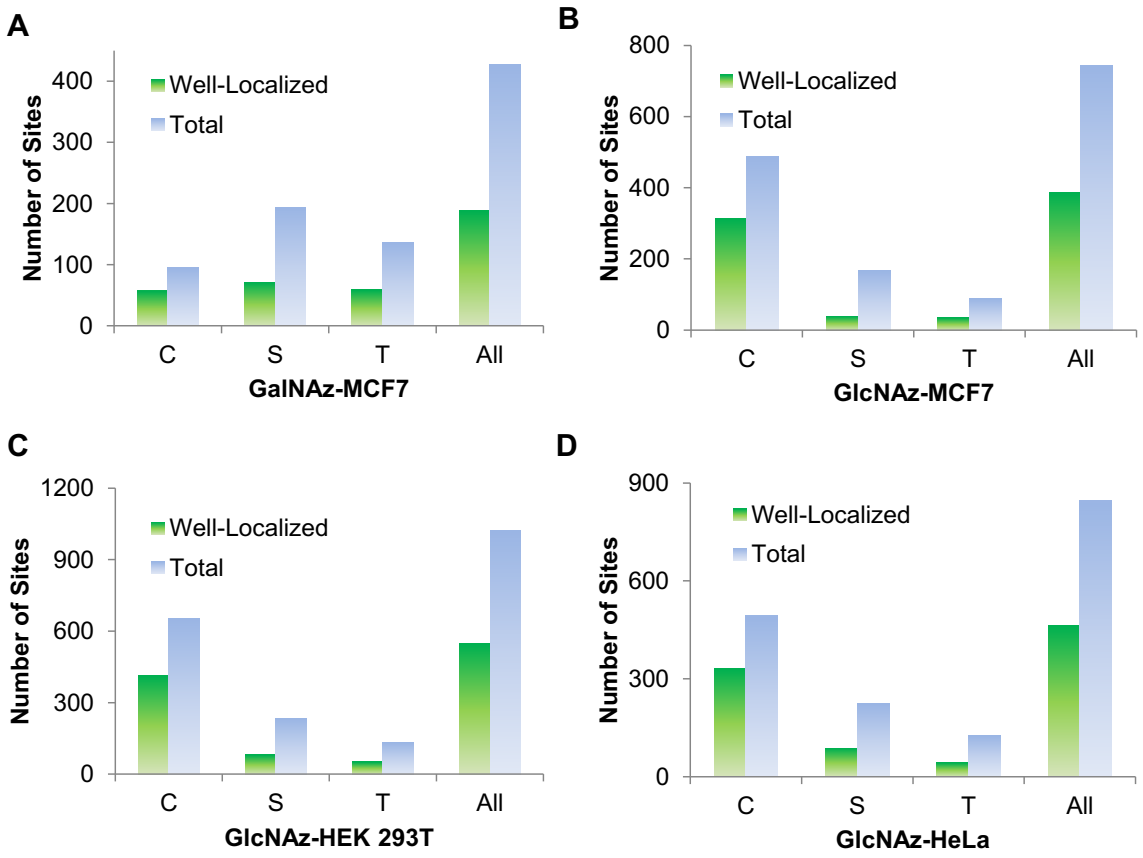


Figure 5.11 The totall and well-localized S- and O-GlcNAcylation sites identified from the four experiments: (A) GalNAz-MCF7; (B) GlcNAz-MCF-7; (C) GlcNAz-HEK 293T; (D) GlcNAz-HeLa.

The total and well-localized O- and S-GlcNAcylation site numbers in the four experiments are compared in Figure 5.11. Interestingly, the site localization confidence for S-GlcNAcylation sites is generally higher partially due to the fact that there are a considerable amount of S-GlcNAc-containing peptides without any S or T residue in the sequence while additional S or T residue(s) nearly always appear in the O-GlcNAcylated peptides.

5.2.3.2 Distinctive labelling performances of GalNAz and GlcNAz

Firstly, we compared labelling MCF7 cells using two different sugar analogs (GalNAz and GlcNAz) for the global analysis of protein GlcNAcylation. In the literature, it was reported that although the sugar analog GlcNAz was structurally more similar to GlcNAc, GalNAz was able to label O-GlcNAcylated proteins in human cells more efficiently.[95] In addition, UDP-GalNAz and UDP-GlcNAc are interconvertible through their salvage pathways.[95, 96] The current results further indicated that GalNAz outperformed GlcNAc for the identification of O-GlcNAcylated proteins, while GlcNAz appeared to be more effective to label S-GlcNAcylated proteins (Figure 5.12A).

In this work, in order to ensure the technical rigor and to avoid the site localization ambiguity caused by false site assignment, we only analyzed the well-localized sites with ModScore > 13. As shown in Figure 5.12A, many more sites were identified in the GlcNAz labelling experiment and the majority of them were S-GlcNAcylation sites, while GalNAz labelling generated more O-GlcNAcylation sites. In the GalNAz labelling experiment, 69.3% of the well-localized sites were O-GlcNAcylation sites while it was only 19.1% for GlcNAz labelling (Figure 5.12A). For S-GlcNAcylation site identification, the labelling with GlcNAz dramatically outperformed GalNAz with 313 well-localized S-GlcNAcylation sites from the GlcNAz labelling *vs.* 58 from the GalNAz labelling. Labelling using these two sugar analogs resulted in a relatively small overlap (Figure 5.12B), and the overlapped sites are mostly located on the cysteine residues (82.6%, Figure 5.12C).

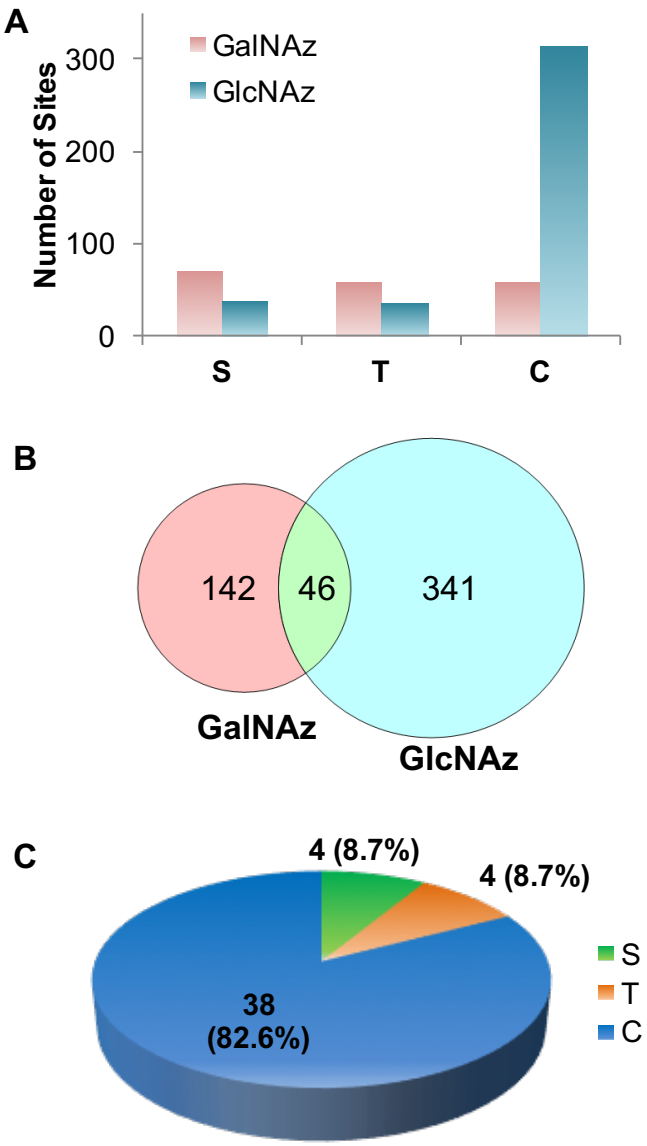


Figure 5.12 (A) Distributions of the well-localized sites on cysteine, serine, and threonine in the GalNAz and GlcNAz labelling experiments; (B) The site overlap between the two experiments; (C) The O- and S-GlcNAcylation site percentages among the overlapped sites.

Although the sites identified from the two experiments are quite different, the GlcNAcylated proteins identified from the two experiments have shown a better overlap with nearly half of the proteins identified in the GalNAz labelling experiment also detected in the GlcNAz labelling experiment (Figure 5.13). We then separately clustered the proteins

identified from each experiment based on cellular compartment (Figure 5.14) and molecular function (Figure 5.15) using the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.8).[59] The proteins identified from the two experiments were mostly intracellular and nuclear proteins with a much smaller portion of them involved in adherens junction, which could potentially locate on cell surface (Figure 5.14). Clustering based on molecular function showed that the glycoproteins from the two experiments participated in several types of binding activities, such as nucleic acid binding and cell adhesion molecule binding (Figure 5.15). The protein clustering results imply that although O- and S-GlcNAc locate on different sites from different proteins, many GlcNAcylated proteins have similar locations and functions.

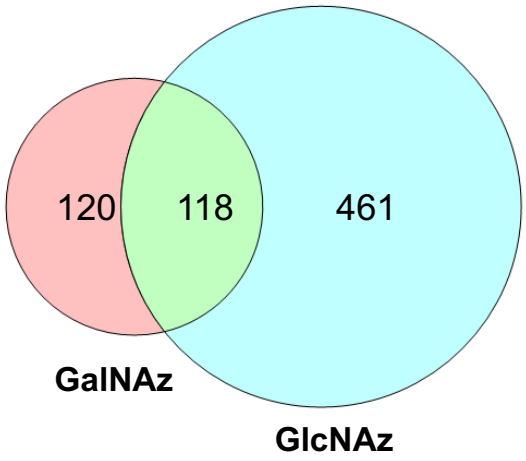


Figure 5.13 The overlap between the GlcNAcylated proteins identified in the GalNAz-MCF7 and GlcNAz-MCF7 experiments.

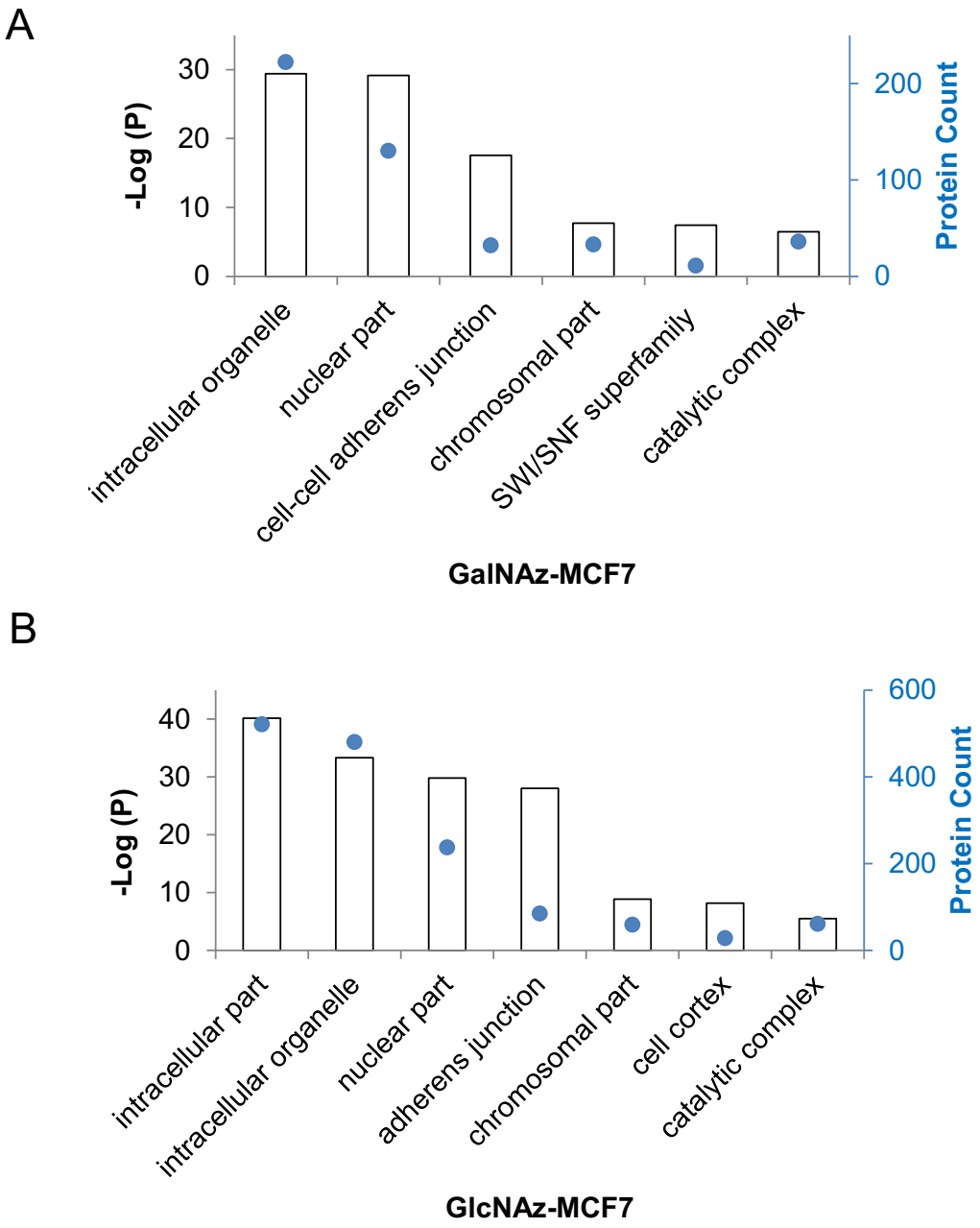


Figure 5.14 Clustering of the GlcNAcylated proteins from (A) the GalNAz-MCF7 experiment and (B) the GlcNAz-MCF7 experiment based on cellular compartment.

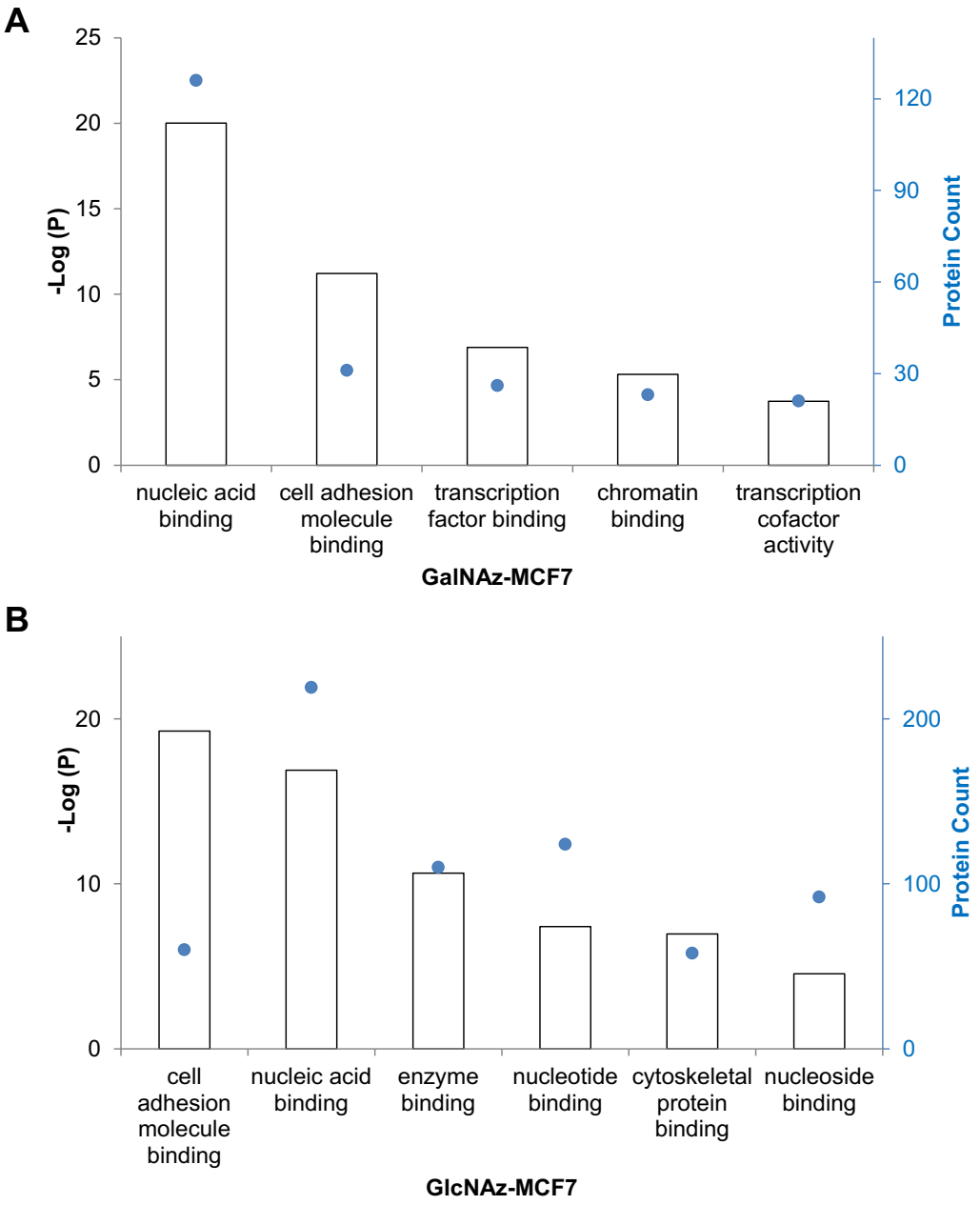


Figure 5.15 Clustering of the GlcNAcylated proteins from (A) the GalNAz-MCF7 experiment and (B) the GlcNAz-MCF7 experiment based on molecular function.

5.2.3.3 Comparison of protein GlcNAcylation in three types of human cells

After identifying GlcNAz as a more robust probe for protein S-GlcNAcylation identification, we then analyzed protein GlcNAcylation in different types of cells using this sugar analog to label cells. The cell lines have different origins with distinctive disease statuses. MCF7 is a breast cancer cell line, HEK 293T is an immortalized but not cancerous cell line originally derived from human embryonic kidney, and HeLa cells were from cervical cancer tissue.

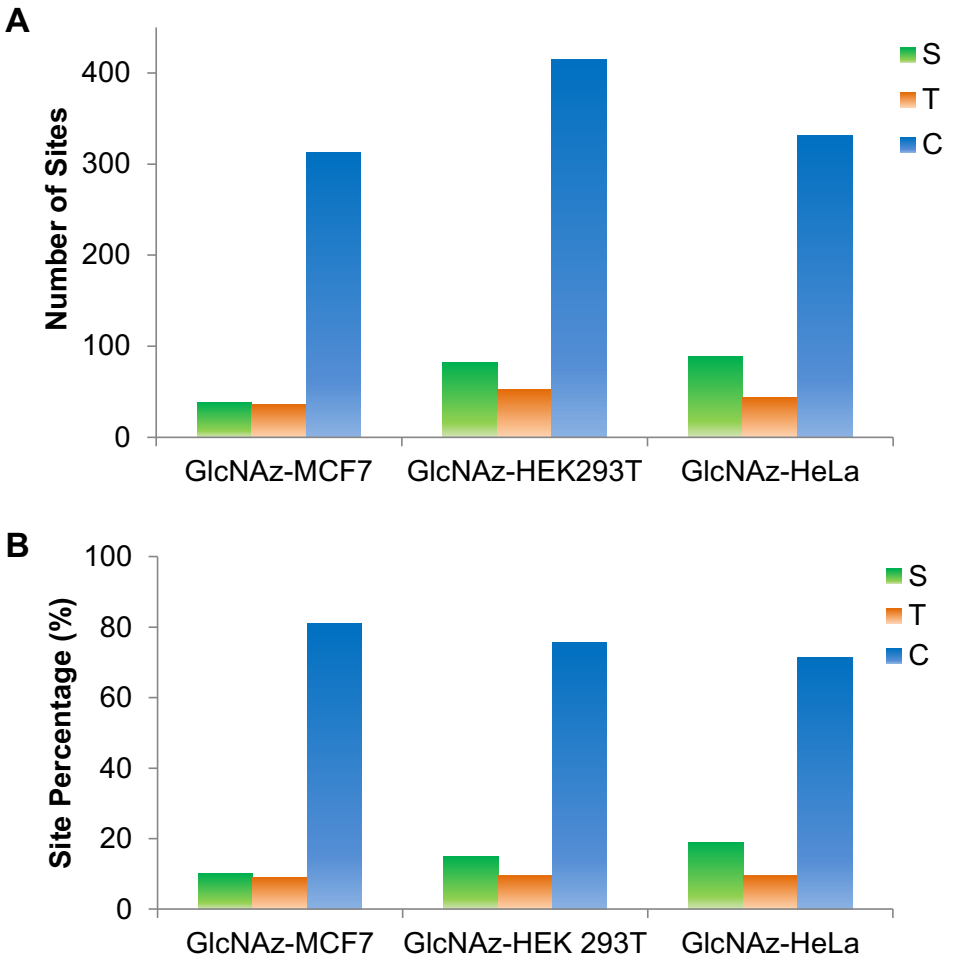


Figure 5.16 The (A) number and (B) percentage distributions of the O- and S-GlcNAcylation sites identified in three types of human cells.

The comparison of the total and well-localized sites from the three experiments is shown in Figure 5.11, and the distributions of the well-localized sites on cysteine, serine, or threonine for three types of cells are displayed in Figure 5.16. The greatest number of S-GlcNAcylation sites were identified from HEK 293T cells (Figure 5.16A). MCF7 has the highest percentage of S-GlcNAcylation sites while HeLa has the highest percentage of O-GlcNAcylation (Figure 5.16B).

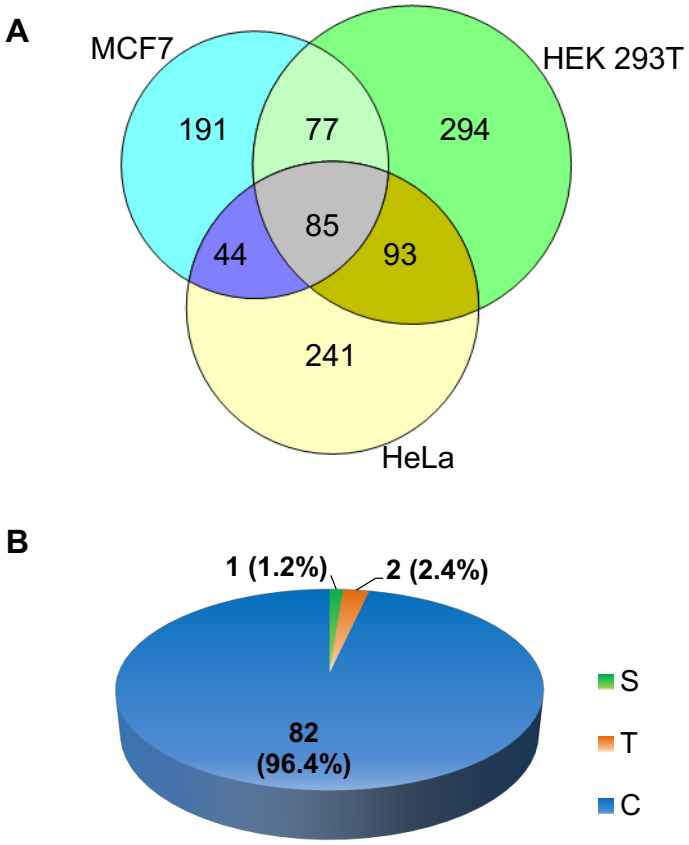


Figure 5.17 (A) the site overlap of the three experiments, and (B) the O- and S-GlcNAcylation site percentages among the overlapped sites.

Overall, the percentages of S-GlcNAcylation sites among the total well-localized sites were similar across the three types of cell, and 80.9%, 75.5%, and 71.5% of GlcNAcylation sites were localized on the cysteine residues in MCF7, HEK 293T, and HeLa cells, respectively.

The overlap is less at the site level than the protein level (Figure 5.17A). Among the overlapped sites, 96.4% of them are located on cysteine (Figure 5.17B), featuring an even higher percentage of S-GlcNAcylation sites than it from the overlap between the GalNAz and GlcNAz labelling experiments. Although the percentages of S-GlcNAcylation sites change only slightly across these different cell lines, the site location was quite different. The possible reasons are that some sites may be cell-specific, and that similar to O-GlcNAcylayion, S-GlcNAcylation may also be dynamic.

The comparison of the GlcNAcylated proteins identified in the three experiments is shown in Figure 5.18A. Compared to labelling MCF7 cells using two different sugar analogs, labelling different types of human cells using one sugar analog has resulted in a better overlap. We clustered the non-overlapping proteins in each type of cells separately using DAVID v6.8 according to their tissue expressions through the UniGene category. As expected, the non-overlapping proteins have shown tissue and disease-specificity as the exact tissue origin and disease status-related category was found enriched in each type of cell (Figure 5.18B).

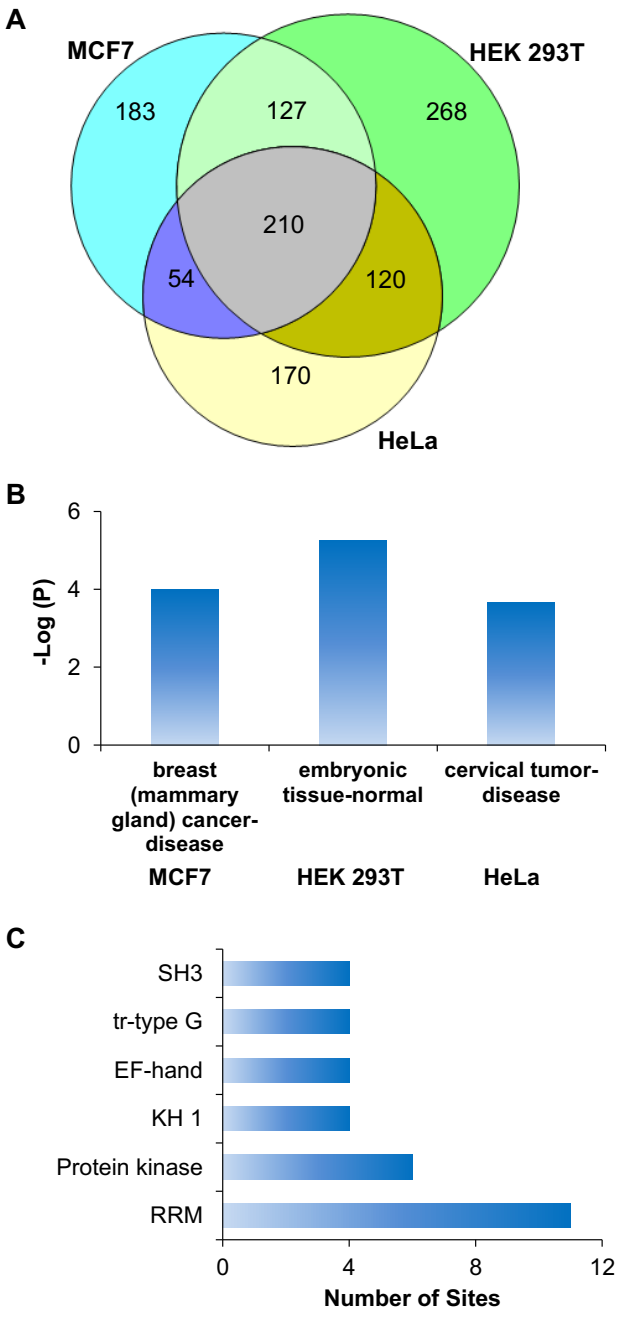


Figure 5.18 (A) GlcNAcylated proteins identified from GlcNAz labelling in three types of human cells; (B) The non-overlapping proteins have shown cell type and disease status specificity; (C) The functional domains found with most C site hits.

5.2.3.4 Analysis of the well-localized S-GlcNAcylation sites

Taken together, we identified a total of 1132 well-localized GlcNAcylation sites from the four experiments, among which 724 are located on cysteine, 240 on serine, and 168 on threonine (Figure 5.19). We clustered the O- and S-GlcNAcylated proteins based on their cellular compartment, molecular function, and biological process and found the highly enriched categories are very similar between the two (Table 5.2).

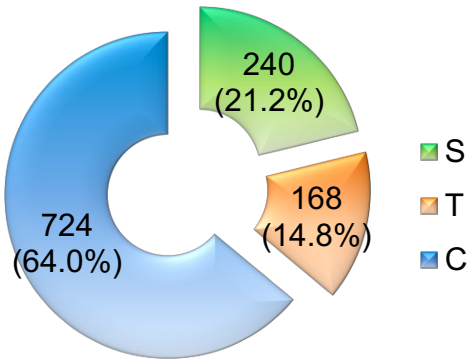


Figure 5.19 The distribution of the well-localized O- and S-GlcNAcylation site identified in all experiments taken together.

After optimizing the experimental conditions and using different types of cells and sugar analogs, we identified dramatically more well-localized sites on cysteine than those in our previous study (724 *vs.* 199).[94] Based on many more well-localized S-GlcNAcylation sites, motif analysis results should be more statistically meaningful. Several motifs with an acidic amino acid near the modification site were once again identified (Figure 5.20), demonstrating that our motif analysis results were highly reproducible. In addition, a new motif with the proline residue locating three amino

acids away from the modified cysteine was found here. Identification of these motifs strongly suggests that some enzymes are responsible for S-GlcNAcylation.

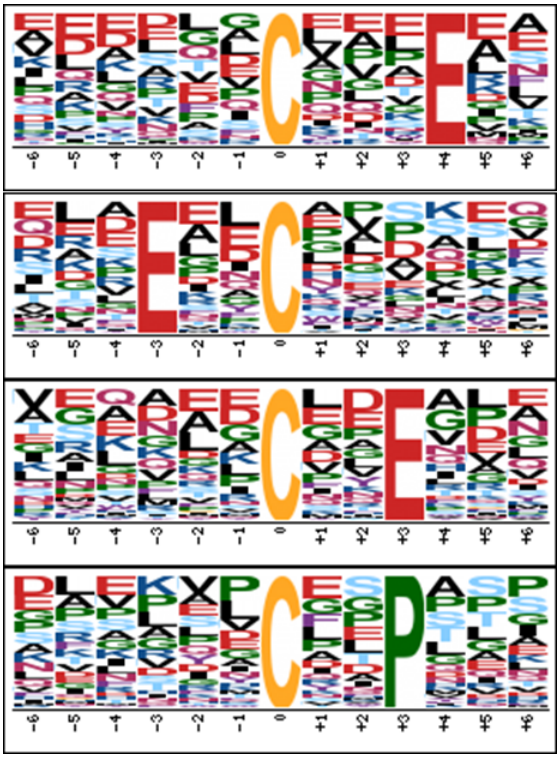


Figure 5.20 The motifs identified from the well-localized S-GlcNAcylation sites.

Moreover, in the literature, researchers found that the location of cysteine-related motifs, such as the CP motif for heme-responsive proteins[97, 98] and CXXC for redox proteins[99, 100] are usually important for their functions. We also studied the S-GlcNAcylation site locations along the protein sequences, but no obvious trend was found (Figure 5.21). The sites seem to locate evenly along their protein sequences despite that there are slightly fewer sites located within 10-20% of the protein length away from the protein N-termini.

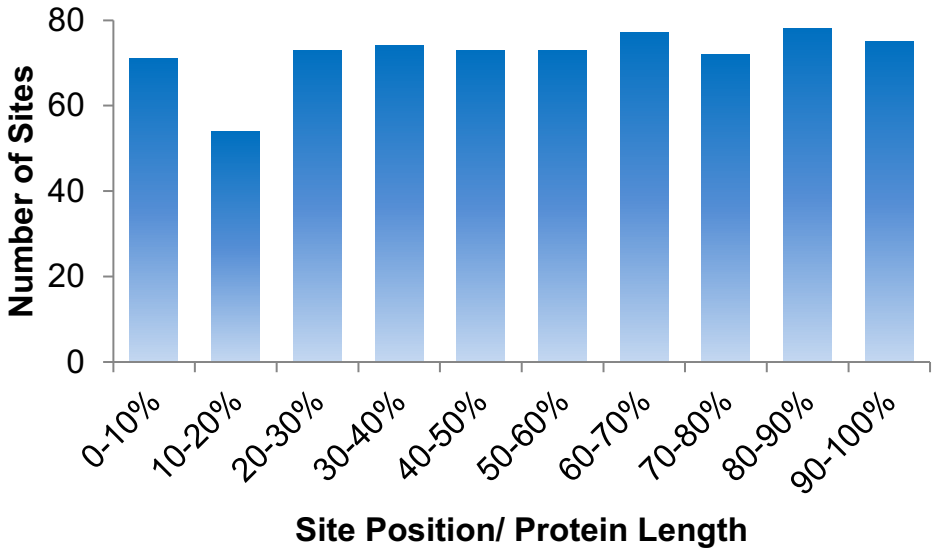


Figure 5.21 The S-GlcNAcylation site location distribution along the protein sequence.

5.2.3.5 Domain analysis of the well-localized S-GlcNAcylation sites

We further performed domain analysis for the well-localized S-GlcNAcylation sites. Ninety-nine sites are located in a variety of functional domains. Figure 5.18C shows the domains with the most site hits. Four domains are involved in binding: RRM is a putative RNA-binding domain that binds single strand RNAs,[101] KH-1 is a nucleic acid binding domain,[102] tr-type G is a translational-type guanine nucleotide-binding domain,[103] and EF-hand domain is responsible for calcium binding.[104] Some sites were located in kinase activity-related domains, such as SH3, which is a tyrosine kinase activity-related domain.[105] These results correspond well with the protein clustering results above and in our previous work, and hence demonstrate that S-GlcNAcylation may be majorly involved in gene expression (especially nucleotides-related binding activities) and signal transduction (protein kinase activity), which are similar to O-GlcNAc. This increases the possibility that protein S-GlcNAcylation has similar functions as O-GlcNAcylation although more research is required to further explore the functions of protein S-GlcNAcylation.

Table 5.2 Clustering of the S- and O-GlcNAcylated proteins according to cellular compartment, molecular function, and biological process using DAVID v6.8.

Category	Term	S-GlcNAcylation		O-GlcNAcylation	
		Protein Count	P-Value	Protein Count	P-Value
Cellular Compartment	Intracellular part	552	4.60E-44	276	1.90E-22
	Nuclear part	270	1.00E-41	147	7.80E-28
	Organelle lumen	280	2.10E-38	147	3.40E-23
Molecular Function	Cell-cell adherens junction	55	1.40E-23	35	3.20E-18
	Nucleic acid binding	262	5.30E-30	137	2.50E-19
	Protein binding involved in cell adhesion	54	7.90E-23	34	1.30E-17
Biological Process	Cellular component organization	310	2.20E-22	156	1.50E-11
	Nitrogen compound metabolic process	335	4.20E-21	169	4.50E-11
	Cellular localization	149	2.70E-13	70	1.20E-05
	Cell cycle process	96	7.00E-13	51	5.10E-08

5.2.4 Conclusions

Protein glycosylation is essential for mammalian cell growth and proliferation, and it is highly diverse. Previously we made an unexpected finding of extensive protein S-GlcNAcylation existing in human cells. Compared to O-GlcNAcylation, this type of glycosylation remains largely unexplored. Here we investigated protein S-GlcNAcylation with the two different sugar analogs and in three types of human cells. The experimental results demonstrated the different percentages of O- and S-GlcNAcylation sites were identified with the GalNAz or GlcNAz labeling. The sugar analog GalNAz outperformed GlcNAz on the identification of protein O-GlcNAcylation, while the latter was proven to be much more effective for the global analysis of protein S-GlcNAcylation. In the comparison of protein GlcNAcylation in three types of human cells: MCF7, HEK 293T and HeLa, the greatest number of GlcNAcylation sites were identified in HEK 293T cells. The percentages of S-GlcNAcylation sites among the total well-localized sites are similar across the three types of human cells, and the majority of the well-localized glycosylation sites were located on the cysteine residues. The results of the S-GlcNAcylation site motif analysis are consistent with our previous finding of several motifs with an acidic amino acid around the site. Compared to other domains, the RNA-binding domains (RRM and KH-1) have the greatest number of sites located within, which is in a good agreement with the clustering results that proteins related to nucleic acid binding and chromatin binding were enriched. Taken together, this work demonstrated the dramatic difference of GlcNAcylated protein labelling with the two sugar analogs, and proteins are S-GlcNAcylated to a similar extent in the three different types of human cells tested. Further investigation of protein glycosylation will advance our understanding of this important and complex modification.

5.3 References

1. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta-Gen. Subj.* **1426**, 239-257 (1999).
2. Haltiwanger, R.S. & Lowe, J.B. Role of glycosylation in development. *Annu. Rev. Biochem.* **73**, 491-537 (2004).
3. Levine, Z.G. & Walker, S. in Annual Review of Biochemistry, Vol 85, Vol. 85. (ed. R.D. Kornberg) 631-657 (Annual Reviews, Palo Alto; 2016).
4. Adamczyk, B., Tharmalingam, T. & Rudd, P.M. Glycans as cancer biomarkers. *Biochim. Biophys. Acta-Gen. Subj.* **1820**, 1347-1353 (2012).
5. Ohtsubo, K. & Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).
6. Spiro, R.G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).
7. Marino, F. et al. Extended O-GIcNAc on HLA class-I-bound peptides. *J. Am. Chem. Soc.* **137**, 10922-10925 (2015).
8. 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.* **259**, 3308-3317 (1984).
9. Lazarus, M.B., Nam, Y.S., Jiang, J.Y., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-U168 (2011).
10. Hart, G.W., Housley, M.P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022 (2007).
11. Wang, X.S. et al. A Novel Quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol Cell Proteomics* **15**, 2462-2475 (2016).
12. Venugopal, H. et al. Structural, dynamic, and chemical characterization of a novel S-glycosylated bacteriocin. *Biochemistry* **50**, 2748-2755 (2011).
13. Maynard, J.C., Burlingame, A.L. & Medzihradszky, K.F. Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), a new post-translational modification in mammals. *Mol. Cell. Proteomics* **15**, 3405-3411 (2016).
14. 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.* **106**, 8894-8899 (2009).
15. Wang, Z.H. et al. 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* **9**, 153-160 (2010).
16. Griffin, M.E. et al. Comprehensive mapping of O-GlcNAc modification sites using a chemically cleavable tag. *Mol. Biosyst.* **12**, 1756-1759 (2016).

17. Wu, R.H. et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods* **8**, 677-U111 (2011).

18. Ludwig, K.R., Sun, L.L., Zhu, G.J., Dovichi, N.J. & Hummon, A.B. Over 2300 Phosphorylated peptide identifications with single-shot capillary zone electrophoresis-tandem mass spectrometry in a 100 min Separation. *Anal. Chem.* **87**, 9532-9537 (2015).

19. Huang, H., Lin, S., Garcia, B.A. & Zhao, Y.M. Quantitative proteomic analysis of histone modifications. *Chem. Rev.* **115**, 2376-2418 (2015).

20. 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).

21. Hong, Q.T. et al. A Method for comprehensive glycosite-mapping and direct quantitation of serum glycoproteins. *J. Proteome Res.* **14**, 5179-5192 (2015).

22. Khatri, K. et al. Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. *J. Proteome Res.* **13**, 4347-4355 (2014).

23. Madsen, J.A. et al. Concurrent automated sequencing of the glycan and peptide portions of O-linked glycopeptide anions by ultraviolet photodissociation mass spectrometry. *Anal. Chem.* **85**, 9253-9261 (2013).

24. 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).

25. Lee, L.Y. et al. Toward automated N-glycopeptide identification in glycoproteomics. *J. Proteome Res.* **15**, 3904-3915 (2016).

26. Chalkley, R.J. & Burlingame, A.L. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *J Am Soc Mass Spectr* **12**, 1106-1113 (2001).

27. 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).

28. Madera, M., Mechref, Y. & Novotny, M.V. Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal. Chem.* **77**, 4081-4090 (2005).

29. Chen, W.X., Smeekens, J.M. & Wu, R.H. A universal chemical enrichment method for mapping the yeast n-glycoproteome by mass spectrometry (MS). *Mol Cell Proteomics* **13**, 1563-1572 (2014).

30. Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* **12**, 561-567 (2015).

31. Zacharias, L.G. et al. HILIC and ERLIC enrichment of glycopeptides derived from breast and brain cancer cells. *J. Proteome Res.* **15**, 3624-3634 (2016).

32. Chandler, K.B. & Costello, C.E. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis* **37**, 1407-1419 (2016).

33. Wells, L. et al. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **1**, 791-804 (2002).

34. Cieniewski-Bernard, C. et al. Identification of O-linked N-acetylglucosamine proteins in rat skeletal muscle using two-dimensional gel electrophoresis and mass spectrometry. *Mol. Cell. Proteomics* **3**, 577-585 (2004).

35. Lee, A. et al. Combined antibody/lectin enrichment identifies extensive changes in the O-GIcNAc sub-proteome upon oxidative stress. *J Proteome Res* **15**, 4318-4336 (2016).

36. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13132-13137 (2004).

37. Khidekel, N. et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* **3**, 339-348 (2007).

38. Alfaro, J.F. et al. Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *P Natl Acad Sci USA* **109**, 7280-7285 (2012).

39. 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. *P Natl Acad Sci USA* **100**, 9116-9121 (2003).

40. Sprung, R. et al. Tagging-via-substrate strategy for probing O-GlcNAc modified proteins. *J Proteome Res* **4**, 950-957 (2005).

41. Agarwal, P., Beahm, B.J., Shieh, P. & Bertozzi, C.R. Systemic fluorescence imaging of zebrafish glycans with bioorthogonal chemistry. *Angew. Chem.-Int. Edit.* **54**, 11504-11510 (2015).

42. McKay, C.S. & Finn, M.G. Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem. Biol.* **21**, 1075-1101 (2014).

43. Smeekens, J.M., Chen, W.X. & Wu, R.H. Mass spectrometric analysis of the cell surface N-glycoproteome by combining metabolic labeling and click chemistry. *J Am Soc Mass Spectr* **26**, 604-614 (2015).

44. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem. Sci.* **7**, 1393-1400 (2016).

45. 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 Spectr* **5**, 976-989 (1994).

46. 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).

47. 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).

48. 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).

49. Huttlin, E.L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010).

50. Schwartz, D. & Gygi, S.P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **23**, 1391-1398 (2005).

51. Mahal, L.K., Yarema, K.J. & Bertozzi, C.R. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **276**, 1125-1128 (1997).

52. 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).

53. Nandi, A. et al. Global identification of O-GlcNAc-modified proteins. *Anal. Chem.* **78**, 452-458 (2006).

54. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chem Sci* **6**, 4681-4689 (2015).

55. Xiao, H.P., Tang, G.X. & Wu, R.H. Site-specific quantification of surface N-glycoproteins in statin-treated liver cells. *Anal Chem* **88**, 3324-3332 (2016).

56. Rodionov, V.O., Presolski, S.I., Diaz, D.D., Fokin, V.V. & Finn, M.G. Ligand-accelerated Cu-catalyzed azide-alkyne cycloaddition: A mechanistic report. *J. Am. Chem. Soc.* **129**, 12705-12712 (2007).

57. Xiao, H.P. & Wu, R.H. Quantitative investigation of human cell surface N-glycoprotein dynamics. *Chem Sci* **8**, 268-277 (2017).

58. Ren, C.H. et al. Disulfide bond as a cleavable linker for molecular self-assembly and hydrogelation. *Chem. Commun.* **47**, 1619-1621 (2011).

59. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

60. Gagneux, P. & Varki, A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **9**, 747-755 (1999).

61. Varki, A. et al. Essentials of Glycobiology (2nd Edition). (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; 2008).

62. Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A. & Dwek, R.A. Glycosylation and the immune system. *Science* **291**, 2370-2376 (2001).

63. Haltiwanger, R.S. & Lowe, J.B. Role of glycosylation indevelopment. *Annu. Rev. Biochem.* **73**, 491-537 (2004).

64. Neelamegham, S. & Mahal, L.K. Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure. *Curr. Opin. Struct. Biol.* **40**, 145-152 (2016).

65. Brownlee, M. Advanced protein glycosylation in diabetes and aging. *Annu Rev Med* **46**, 223-234 (1995).

66. Marquardt, T. & Denecke, J. Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies. *Eur J Pediatr* **162**, 359-379 (2003).

67. Christiansen, M.N. et al. Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525-546 (2014).

68. Freeze, H.H. Update and perspectives on congenital disorders of glycosylation. *Glycobiology* **11**, 129R-143R (2001).

69. Reis, C.A., Osorio, H., Silva, L., Gomes, C. & David, L. Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol.* **63**, 322-329 (2010).

70. 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* **80**, 825-858 (2011).

71. Sayat, R., Leber, B., Grubac, V., Wiltshire, L. & Persad, S. O-GlcNAc-glycosylation of beta-catenin regulates its nuclear localization and transcriptional activity. *Exp Cell Res* **314**, 2774-2787 (2008).

72. Chu, C.S. et al. O-GlcNAcylation regulates EZH2 protein stability and function. *P Natl Acad Sci USA* **111**, 1355-1360 (2014).

73. Lefebvre, T. et al. Evidence of a balance between phosphorylation and O-GlcNAc glycosylation of Tau proteins - a role in nuclear localization. *Bba-Gen Subjects* **1619**, 167-176 (2003).

74. Dong, D.L.Y. & Hart, G.W. Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol. *J Biol Chem* **269**, 19321-19330 (1994).

75. Housley, M.P. et al. O-GlcNAc regulates FoxO activation in response to glucose. *J Biol Chem* **283**, 16283-16292 (2008).

76. Zhao, P. et al. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J Proteome Res* **10**, 4088-4104 (2011).

77. Haynes, P.A. & Aebersold, R. Simultaneous detection and identification of O-GlcNAc-modified glycoproteins using liquid chromatography-tandem mass spectrometry. *Anal Chem* **72**, 5402-5410 (2000).

78. Rexach, J.E., Clark, P.M. & Hsieh-Wilson, L.C. Chemical approaches to understanding O-GlcNAc glycosylation in the brain. *Nat. Chem. Biol.* **4**, 97-106 (2008).

79. Yang, X.Y. & Qian, K.V. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **18**, 452-465 (2017).

80. De Leon, C.A., Levine, P.M., Craven, T.W. & Pratt, M.R. The sulfur-linked analogue of O-GlcNAc (S-GlcNAc) is an enzymatically stable and reasonable structural surrogate for O-GlcNAc at the peptide and protein levels. *Biochemistry* (2017).

81. Zheng, J.N., Xiao, H.P. & Wu, R.H. Specific identification of glycoproteins bearing the Tn antigen in human cells. *Angew Chem Int Edit* **56**, 7107-7111 (2017).

82. Steentoft, C. et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977-982 (2011).

83. Zielinska, D.F., Gnad, F., Wisniewski, J.R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907 (2010).

84. Yang, Y. et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 10 (2016).

85. Wollscheid, B. et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378-386 (2009).

86. Cao, L.W. et al. Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nat. Commun.* **8**, 13 (2017).

87. Reiding, K.R. et al. Human plasma N-glycosylation as analyzed by matrix-assisted laser desorption/ionization-fourier transform ion cyclotron resonance-MS associates with markers of inflammation and metabolic health. *Mol Cell Proteomics* **16**, 228-242 (2017).

88. Calvano, C.D., Zambonin, C.G. & Jensen, O.N. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J. Proteomics* **71**, 304-317 (2008).

89. Palmisano, G. et al. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **5**, 1974-1982 (2010).

90. Hang, H.C., Yu, C., Kato, D.L. & Bertozzi, C.R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *P Natl Acad Sci USA* **100**, 14846-14851 (2003).

91. Hubbard, S.C., Boyce, M., McVaugh, C.T., Peehl, D.M. & Bertozzi, C.R. Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorg Med Chem Lett* **21**, 4945-4950 (2011).

92. Elliott, T.S., Bianco, A., Townsley, F.M., Fried, S.D. & Chin, J.W. Tagging and enriching proteins enables cell-specific proteomics. *Cell Chem. Biol.* **23**, 805-815 (2016).

93. Ciepla, P. et al. New chemical probes targeting cholesterylation of Sonic Hedgehog in human cells and zebrafish. *Chem Sci* **5**, 4249-4259 (2014).

94. Xiao, H.P. & Wu, R.H. Global and site-specific analysis revealing unexpected and extensive protein S-GlcNAcylation in human cells. *Anal Chem* **89**, 3656-3663 (2017).

95. Boyce, M. et al. Metabolic cross-talk allows labeling of O-linked beta-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway. *P Natl Acad Sci USA* **108**, 3141-3146 (2011).

96. 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. *P Natl Acad Sci USA* **108**, 8146-8151 (2011).

97. Igarashi, J. et al. Elucidation of the heme binding site of heme-regulated eukaryotic initiation factor 2 alpha kinase and the role of the regulatory motif in heme sensing by spectroscopic and catalytic studies of mutant proteins. *J Biol Chem* **283**, 18782-18791 (2008).

98. Kuhl, T. et al. Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. *Acs Chem Biol* **8**, 1785-1793 (2013).

99. Fomenko, D.E. & Gladyshev, V.N. Identity and functions of CxxC-derived motifs. *Biochemistry* **42**, 11214-11225 (2003).

100. Conway, M.E., Poole, L.B. & Hutson, S.M. Roles for cysteine residues in the regulatory CXXC motif of human mitochondrial branched chain aminotransferase enzyme. *Biochemistry* **43**, 7356-7364 (2004).

101. Maris, C., Dominguez, C. & Allain, F.H.T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *Febs J* **272**, 2118-2131 (2005).

102. Siomi, H., Choi, M.Y., Siomi, M.C., Nussbaum, R.L. & Dreyfuss, G. Essential role for KH domains in RNA-binding - impaired RNA-binding by a mutation in the KH domain of FMR1 that causes fragile-X syndrome. *Cell* **77**, 33-39 (1994).

103. Leibundgut, M., Frick, C., Thanbichler, M., Bock, A. & Ban, N. Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. *Embo J* **24**, 11-22 (2005).

104. Ikura, M. Calcium binding and conformational response in EF-hand proteins. *Trends Biochem Sci* **21**, 14-17 (1996).

105. Lowenstein, E.J. et al. The SH2 and SH3 domain containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* **70**, 431-442 (1992).

CHAPTER 6. ANALYSIS OF CELLULAR RESPONSES AND PLEIOTROPIC EFFECTS IN STATIN-TREATED LIVER CELLS ON THE PROTEOME, GLYCOPROTEOME, AND PHOSPHOPROTEOME LEVELS

Partially adapted with permission from American Chemical Society

Xiao, H. P., Chen, W. X., Tang, G. X., Smeekens, J. M., and Wu, R. H. Systematic Investigation of Cellular Response and Pleiotropic Effects in Atorvastatin-Treated Liver Cells by MS-Based Proteomics. *Journal of Proteome Research,* 2015, 14, 1600-1611.

Partially adapted with permission from Elsevier B.V.

Xiao, H. P., Hwang, J. E., and Wu, R. Mass Spectrometric Analysis of the N-glycoproteome in Statin-Treated Liver Cells with Two Lectin-Independent Chemical Enrichment Methods. *International Journal of Mass Spectrometry,* DOI: 10.1016/j.ijms.2017.05.010.

6.1 Systematic Investigation of Cellular Response and Pleiotropic Effects in Atorvastatin-treated Liver Cells by MS-based Proteomics

6.1.1 Introduction

Heart and cardiovascular diseases (CVDs) are the leading causes of morbidity and mortality in the United States and around the world.[1, 2] With the increase in life expectancy, these

age-related diseases are becoming increasingly prevalent. Statins, as effective cholesterol-lowering 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCR) inhibitors, are recommended as treatment and preventative drugs by the American Heart Association (AHA).[2, 3] Currently statins are very popular drugs, with ~15 percent of adults in the United States taking statins. In November 2013, the AHA and American College of Cardiology released a new clinical practice guideline for the treatment of blood cholesterol in people at high risk for CVDs caused by atherosclerosis. The guideline identifies four major groups of patients for whom statins have the greatest chance of preventing stroke and heart attack.[4] Under the new guideline, one third of all American adults would meet the threshold to consider taking cholesterol-lowering statin drugs.[5]

The first statin, lovastatin, was approved by the Food and Drug Administration in 1987. In 1996, Pfizer introduced atorvastatin (Lipitor) as a pharmaceutical for the reduction of cholesterol as a preventative measure for heart disease. Atorvastatin quickly became one of the best-selling pharmaceuticals in history, recording sales of $12.4 billion in 2008 alone.[6] Atorvastatin, as well as other statins, were designed to inhibit the rate-limiting step of the cholesterol biosynthesis pathway, known as the mevalonate pathway. The enzyme responsible for catalyzing this rate-limiting step is HMGCR.[7] As more than two-thirds of the body's total cholesterol is synthesized in cells,[8] blocking this pathway to achieve a therapeutic decrease of cholesterol is immensely effective.

HMGCR is an enzyme in the upstream portion of the mevalonate pathway. Besides cholesterol, the synthesis of many intermediate and end products in this pathway, including ubiquinone, dolichol and farnesyl-pyrophosphate (farnesy-PP), are significantly affected by the inhibition of this enzyme. Ubiquinone is a component of the electron transport chain and participates in aerobic cellular respiration, generating energy in the form of adenosine triphosphate

(ATP), which accounts for 95% of the total energy generated in the body.[9, 10] Dolichol plays an essential role in protein *N*-glycosylation, and functions as a membrane anchor for the formation of a precursor oligosaccharide.[11] Farnesyl-PP and geranylgeranyl pyrophosphate (geranylgeranyl-PP) are also important compounds responsible for two other types of protein post-translational modifications, *i.e.* farnesylation and geranylgeranylation, which are involved in protein trafficking and localization, and subsequently regulate cell signaling *via* protein phosphorylation.[12-14] Several end products in the mevalonate pathway play critical roles in cells. Therefore, many other cholesterol-independent effects, so-called pleiotropic effects, have been reported: improving endothelial function,[15] attenuating vascular and myocardial remodeling, inhibiting vascular inflammation and oxidation,[16, 17] and stabilizing atherosclerotic plaques.[7, 18-22]Statins have been shown to reduce the chance of cardiac rejection in cardiac transplants.[23] They have also been reported to inhibit cellular proliferation and induce necrosis, thus allowing them to be potential anticancer agents.[24-29] In addition, statins are associated with a significant reduction in the risk of hip fracture,[30] and could be used as potential anti-Alzheimer's [31-33] and antidiabetic drugs.[34-36] Several adverse effects of statins have also been reported, including rare acute kidney injury, memory loss and confusion.[37] However, the molecular mechanisms of these pleiotropic effects of statins are largely unknown.

With the development of mass spectrometry (MS) instrumentation, genomic science and computer technology, modern liquid chromatography (LC)-MS-based proteomics techniques provide a unique opportunity to globally analyze proteins and protein post-translational modifications in complex biological samples.[38-45] These methods have great advantages for characterizing proteins and protein modifications; they are applicable to identify proteins, locate modification sites and quantify their abundance changes without requiring antibodies.[46-54] Protein

changes in the secretome[55] and lipid rafts[56] in endothelial cells treated by statin were reported recently. Additionally, protein abundance changes in HL-60 (human acute promyelocytic leukemia) cells treated by lovastatin were investigated by MS-based quantitative proteomics techniques.[57] It was found that the abundances of estrogen receptor α and steroid receptor RNA activator 1 in the estrogen receptor signaling pathway were decreased, and glutamate metabolism was altered in treated HL-60 cells.[57] Intracellular signal transduction is mainly carried out by protein phosphorylation, but global analysis of protein phosphorylation in cells treated by statin has yet to be reported. Considering that the majority of cholesterol (~85%) in the human body is synthesized in the liver,[58] investigating the liver cell response to atorvastatin is of great interest. Comprehensive and systematic investigation of proteins and protein phosphorylation in liver cells will provide a better understanding of liver cell responses to statin and the underlying molecular mechanisms of the corresponding pleiotropic effects. This information will be beneficial in expanding the use of statins to other noncardiac vascular diseases and minimizing potential side effects.

In this work, we have globally and quantitatively studied protein and protein phosphorylation changes in cells treated by atorvastatin. Since the majority of cholesterol is produced in the liver, a liver carcinoma cell line (HepG2) was used in this study. After cells were treated by atorvastatin for 24 hours, proteins and phosphoproteins were globally identified and quantified by MS-based proteomics techniques. As expected, many lipid-related proteins were up-regulated in cells treated by atorvastatin, including HMGCR, FDFT, SQLE and LDLR. Phosphopeptides on a group of G-protein modulators were up-regulated, which strongly suggests that cell signal rewiring was a result of statin treatment on lipidated proteins. Several basic motifs were enriched among down-regulated phosphorylation sites, which indicates that kinases with

preference for these motifs, such as PKA and PKC, have attenuated activities. The current work represents the first global analysis of proteins and phosphoproteins in liver cells treated by atorvastatin, which can provide a better understanding of the mode of action of statins and the molecular mechanisms of their pleiotropic effects.

6.1.2 Materials and methods

6.1.2.1 Cell culture, SILAC labeling and atorvastatin treatment

HepG2 (C3A) cells (from American type culture collection (ATCC)) were grown in Dulbecco's modified eagle's medium (DMEM) (Sigma-Aldrich) containing 1000 mg/L glucose and 10% fetal bovine serum (FBS) (Thermo). “Heavy” and “light” SILAC (stable isotope labeling by amino acids in cell culture) (Sigma-Aldrich) media were freshly prepared by adding 0.146 g/L $^{13}C_6{}^{15}N_2$ L-lysine (Lys-8) (Cambridge Isotopes) or the corresponding non-labeled lysine, $^{12}C_6{}^{14}N_2$ L-lysine (Lys-0) and supplemented with 10% dialyzed FBS (Corning). Cells were cultured for about seven generations under 37 ºC, 5% CO_2 and humidified atmosphere before treatment. 40 mM atorvastatin (Cayman Chemical) stock solution was prepared by dissolving in DMSO (Sigma-Aldrich). About 9×10^7 heavy HepG2 cells were treated with 15 μM atorvastatin in serum-free heavy medium for 24 h. A similar number of light cells were treated by DMSO in serum-free light medium as a control.

6.1.2.2. Cell lysis, protein extraction and digestion

Heavy and light SILAC-labeled HepG2 cells were harvested by scraping in phosphate buffered saline (PBS), and equally combined based on the protein ratio from a trial run. The cell mixtures were pelleted by centrifugation at 500 g for 3 min, washed with cold PBS and lysed

through end-over-end rotation at 4 ºC for 45 minutes in lysis buffer (50 mM HEPES pH=7.6, 150 mM NaCl, 50 mM NaF, 50 mM β-glycerophosphate, 1 mM sodium orthovanadate, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium pyrophosphate, 0.5% SDC, 10 units/mL benzonase and one protease inhibitor cocktail tablet (complete mini, EDTA-free, Roche) per 10 mL). Lysates were centrifuged, and the resulting supernatant was transferred to new tubes. Proteins were subjected to disulfide reduction with 5 mM DTT (56 ºC, 25 min) and alkylation with 14 mM iodoacetamide (RT, 20 min in the dark). Detergent was removed by methanol chloroform precipitation. The purified proteins were digested with 10 ng/μL Lys-C (Wako) in 50 mM HEPES pH 8.6, 1.6 M urea, 5% ACN at 31ºC for 16 h.

6.1.2.3 Peptide separation for protein analysis

Digestion mixtures were acidified by addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, clarified by centrifugation and desalted using tC18 SepPak cartridge (Waters). For protein expression analysis, the peptide mixture was separated by high pH reversed-phase high-performance liquid chromatography (HPLC) into 21 fractions with a 40-min gradient of 5-55% ACN in 10 mM ammonium acetate (pH=10).

6.1.2.4 Phosphopeptide enrichment

Effective enrichment is critical to globally identify and quantify protein phosphorylation, and here two-step enrichment was employed. The first separation of phosphopeptides from non-phosphopeptides was carried out using strong cation change (SCX) chromatography as described.[59] At pH~3, phosphate groups carrying negative charges make phosphopeptides elute earlier than their counterparts. The sample was separated into 12 fractions. Phosphopeptides in

each fraction were further enriched by TiO_2 particles (GL Science Inc., Japan). Enriched phosphopeptide samples were purified by tC18 SepPak cartridge prior to MS analysis.

6.1.2.5 LC-MS/MS analyses

Purified and dried peptide samples were dissolved in a solution of 5% ACN and 4% formic acid (FA), and 2 µL of the resulting solutions were loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 µm, 200 Å, 100 µm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 90 min gradient of 4-30% ACN (in 0.125% FA). For phosphorylation samples, a 110 min gradient was used. Peptides were detected with a data-dependent Top20 method[60] in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 20 MS/MS in the LTQ for the most intense ions. The selected ions were excluded from further analysis for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

6.1.2.6 Database searches and data filtering

Raw data files from the mass spectrometer were converted into mzXML format. Individual precursors selected for MS^2 sequencing were checked for incorrect monoisotopic peak assignments while refining precursor ion mass measurements.[50] All MS^2 spectra were then searched using the SEQUEST algorithm (version 28),[61] and spectra were matched 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 reversed orders to control the false discovery rate (FDR) of peptide and protein identification. Data from protein expression experiments were searched using the following parameters: 20 ppm precursor mass tolerance; 1.0 Da product ion mass tolerance; fully digested with Lys-C; up to three missed cleavages; variable modifications: oxidation of methionine (+15.9949); fixed modifications: carbamidomethylation of cysteine (+57.0214). Phosphopeptide samples were searched using the same parameters, with the addition of a variable modification of serine, threonine, and tyrosine (+79.9663).

The target-decoy method was employed to estimate and control FDRs at the peptide and protein levels.[62, 63] Data for either protein or protein phosphorylation analysis were processed separately. Linear discriminant analysis (LDA) was used to distinguish correct and incorrect peptide identifications using numerous parameters such as Xcorr, ΔCn, precursor mass error, and charge state.[50] Separate linear discriminant models were trained for each LC-MS analysis using forward and reversed peptide sequences to provide positive and negative training data. This approach is similar to other methods in the literature which employed different features or alternative classifiers.[64-66] After scoring, only peptides with at least seven amino acid residues in length were kept, and peptide spectral matches were filtered to a 1% FDR based on the number of decoy sequences in the remaining data set. Since phosphorylated and non-phosphorylated peptides have different score distributions, the dataset was restricted to phosphopeptides when determining FDRs for phosphopeptide identification.[67]

When large proteomics datasets are assembled, the protein-level FDR is often dramatically accumulated with the increased sample numbers, despite keeping the peptide-level FDR at a

constant 1% for each run. Therefore, we applied an additional protein-level filter to each dataset to reduce the protein-level FDRs (<1%) for proteins and phosphoproteins. Consequently the FDRs at the peptide level were markedly reduced.

6.1.2.7 Phosphorylation site localization and peptide quantification

To assign phosphorylation site localizations and measure the assignment confidence, we applied a probabilistic algorithm[67] that considers all phospho-forms of a peptide and uses the presence or absence of experimental fragment ions unique to each form to calculate an ambiguity score (Ascore). The Ascore indicates the likelihood that the best site match is correct when compared with the next best match. We considered sites with Ascore ≥13 ($P \leq 0.05$) to be confidently localized. For peptide quantification, we required an S/N value >3 for both heavy and light species. If the S/N value of one member of a pair was less than 3, the partner was required to be greater than 5.

6.1.2.8 Motif analysis

In the motif analysis, only quantified sites were used, *i.e.* singly phosphorylated peptides with well-localized phosphorylation sites (Ascore≥13) were considered. For up- and down-regulated phosphorylation sites, sequences were centered on each phosphorylation site and extended to 13 aa (6 residues on each side of the site) and analyzed with the Motif-X algorithm.[68] The *Homo sapiens* protein database was used as a background.

6.1.3 Results and discussion

6.1.3.1 Protein identification and quantification

HepG2 cells were treated by atorvastatin for 24 hours before harvest and subsequent protein and protein phosphorylation analysis, as shown in Figure 6.1. For protein identification and quantification, proteins were digested by Lys-C, purified and then fractionated into 21 samples by HPLC. Each fraction was then measured by LC-MS/MS. Overall, 157,118 total peptides corresponding to 78,316 unique peptides were identified, and 6,316 proteins (based on gene symbols) were identified with 0.99% FDR at the protein level and 0.12% at the peptide level.

Examples of both a mass spectrum (MS) and tandem mass spectrum (MS^2) are displayed in Figure 6.2. The MS shows a pair of the peptide isotope profiles; the heavy peptide is from cells treated by atorvastatin, which has stronger intensity, while the light peptide is from untreated cells. The peptide DQEVLLQTFLDDASPGDK# (# refers to the heavy lysine) was confidently identified in the MS^2 with an XCorr of 6.6 and a mass accuracy of -0.05 ppm (Figure 6.2b). This peptide is from the protein APOB (apolipoprotein B), which is a very large protein that is a major protein constituent of chylomicrons, low-density lipoproteins (LDL) and very low-density lipoproteins (VLDL).

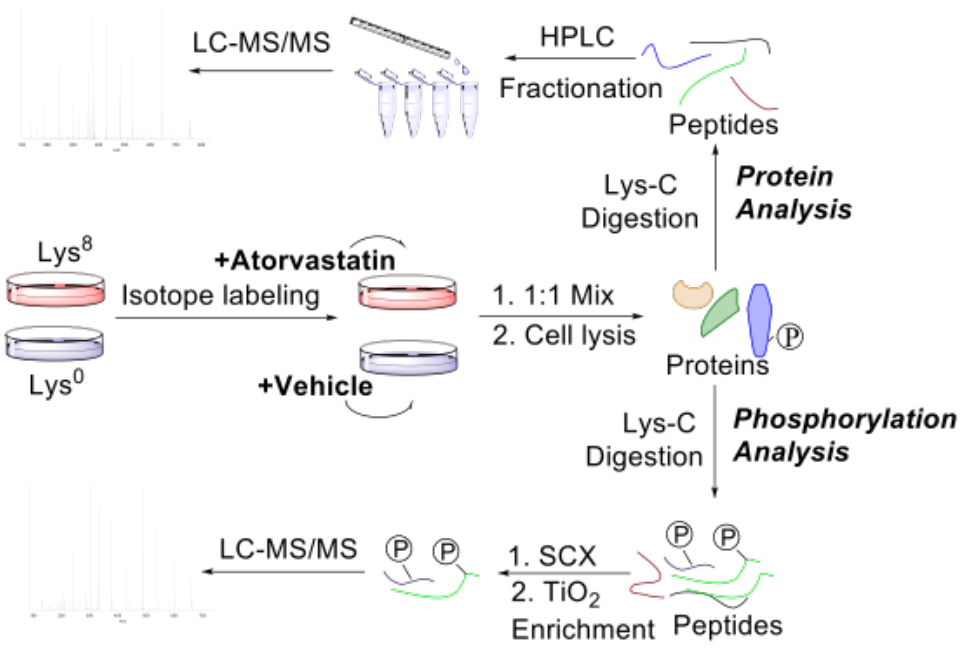


Figure 6.1 Experimental procedure of the global analysis of proteins and protein phosphorylation.

We identified 270 total and 194 unique peptides from this protein. The elution profiles of the light and heavy versions of the peptide are shown in Figure 6.2c. The ratio of the areas under the curves provides highly accurate protein abundance changes. From over 100 quantified unique peptides, we obtained an overall abundance change of 2.02 between treated and untreated cells.

Based on the criteria described in the methods section, we were able to quantify 6,181 proteins (listed in a table online at doi.org/10.1021/pr501277g), with their abundance distribution displayed in Figure 6.3a. At least two peptides were quantified in 5,907 proteins. A total of 104 proteins were down-regulated while 81 proteins were up-regulated by over 2-fold in cells treated by atorvastatin. As expected, few proteins were regulated because atorvastatin is mild enough that it can be taken for many years without significant side effects, and cells were only treated for one day in these experiments.

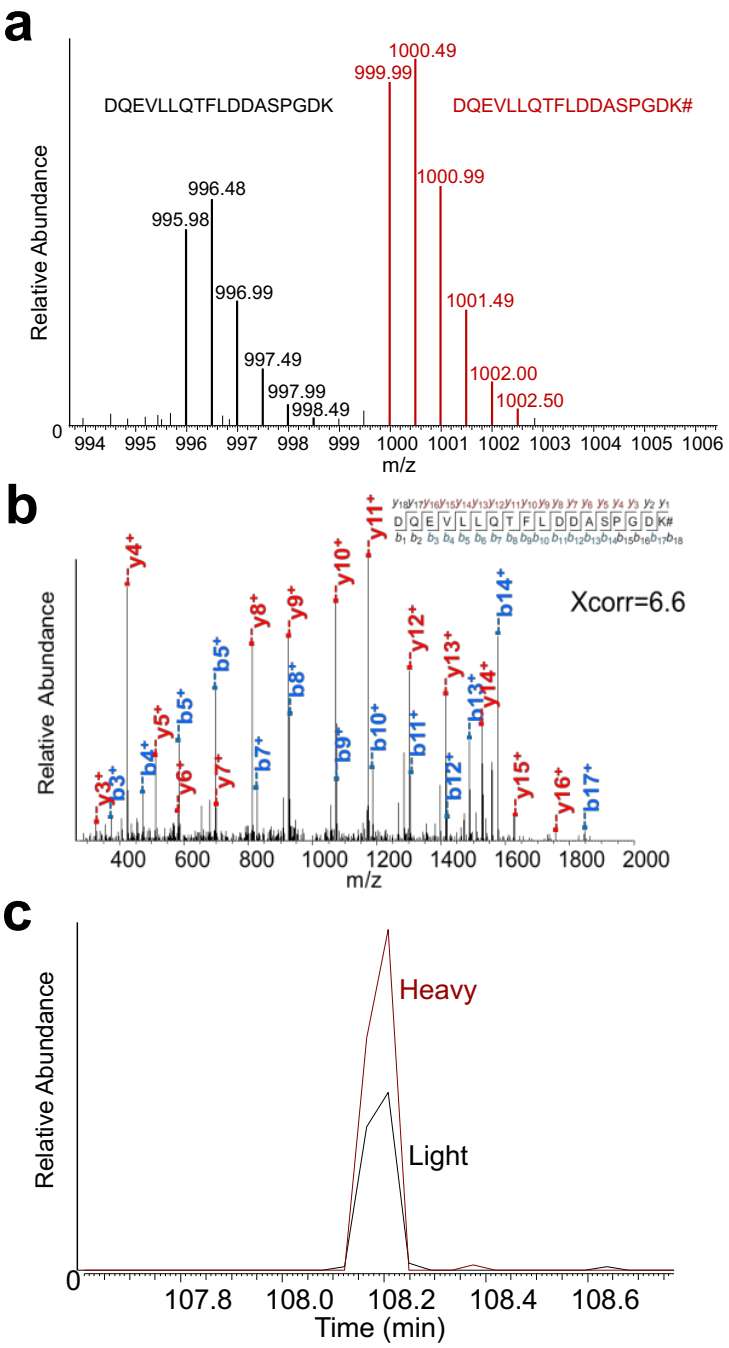


Figure 6.2 Examples of (a) a mass spectrum, (b) a tandem mass spectrum and (c) the elution profiles of heavy (atorvastatin-treated) and light (untreated) versions of the peptide DQEVLLQTFLDDASPGDK, which is from the protein APOB.

6.1.3.2 Up-regulated proteins related to lipid metabolic process

Upon inhibition of HMGCR by atorvastatin, a considerable number of proteins related to the lipid metabolic process were up-regulated. About 21% of up-regulated proteins (17 out of 81), listed in Table 6.1, are related to the lipid metabolic process, which is highly enriched with a *P* value of 8.7 x 10^{-8} (Figure 6.3b) based on the analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID).[69]

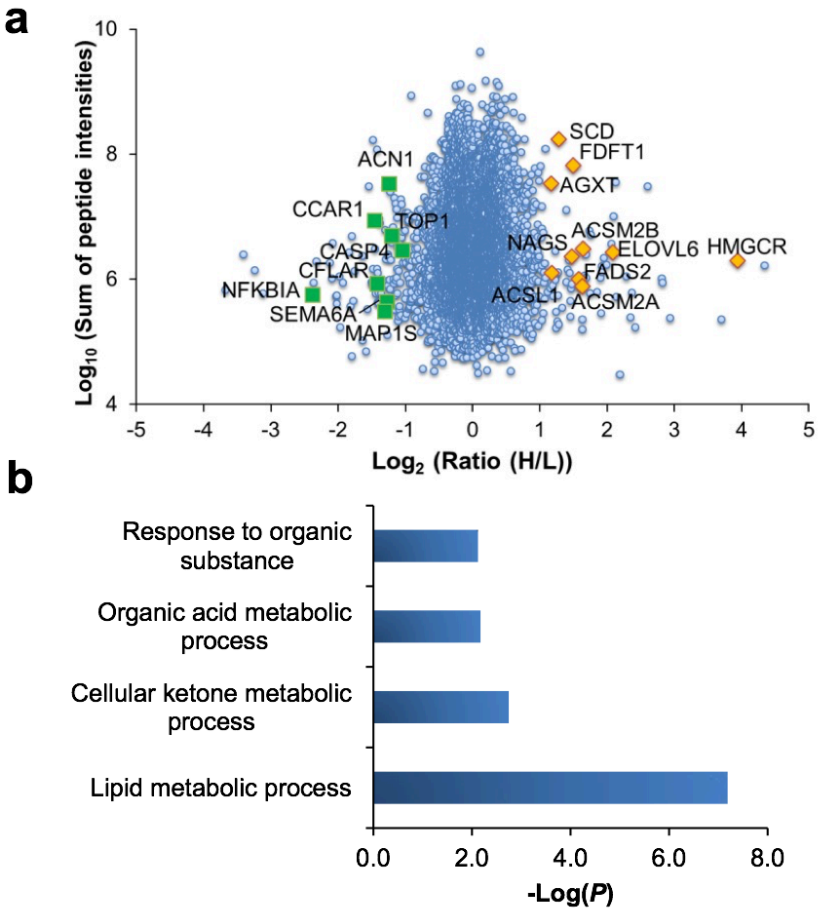


Figure 6.3 (a) Protein abundance changes for cells treated by atorvastatin *vs.* untreated, and (b) clustering of up-regulated proteins.

For example, ACSS2, acetyl-coenzyme A synthetase, which is located in the cytoplasm and activates acetate for lipid synthesis or energy generation, was up-regulated by 2.6-fold. CYP19A1, aromatase, which is crucial for cholesterol homeostasis, was up-regulated by 2.2-fold. It catalyzes a rate-limiting step in cholesterol catabolism and bile acid biosynthesis by introducing a hydrophilic moiety at position 7 of cholesterol (http://www.uniprot.org/uniprot/P22680). CNBP, a cellular nucleic acid-binding protein involved in sterol-mediated repression, was also quantified to be up-regulated by 2.7-fold in treated cells.

Table 6.1 Up-regulated proteins related to lipid metabolic processes.

Reference	Gene symbol	Peptide hits	Protein ratio	Annotation
P33121	ACSL1	39	2.3	Long chain fatty acid CoA ligase 1
Q08AH3	ACSM2A	3	3.1	Acyl-coenzyme A synthetase ACSM2A, mitochondrial
Q68CK6	ACSM2B	3	3.1	Acyl-coenzyme A synthetase ACSM2B, mitochondrial
Q9NR19	ACSS2	22	2.6	Acetyl-coenzyme A synthetase, cytoplasmic
P06727	APOA4	24	4.9	Apolipoprotein A-IV
P04114	APOB	239	2.0	Apolipoprotein B-100
P62633	CNBP	20	2.0	Cellular nucleic acid-binding protein
P11511	CYP19A1	8	2.2	Aromatase
Q9H5J4	ELOVL6	1	4.3	Elongation of very long chain fatty acids protein 6
O95864	FADS2	3	3.0	Fatty acid desaturase 2
P37268	FDFT1	69	2.8	Squalene synthase
P04035	HMGCR	21	15.3	3-hydroxy-3-methylglutaryl-coenzyme A reductase
Q14693	LPIN1	2	2.1	Phosphatidate phosphatase LPIN1
Q8NBP7	PCSK9	9	3.8	Proprotein convertase subtilisin/kexin type 9
Q5T2R2	PDSS1	3	2.4	Decaprenyl-diphosphate synthase subunit 1
O00767	SCD	13	2.4	Acyl-CoA desaturase
Q14534	SQLE	27	4.4	Squalene monooxygenase

6.1.3.3 Abundance changes of proteins in the mevalonate pathway

Atorvastatin inhibited HMGCR which prevented the synthesis of cholesterol and other lipids in the mevalonate pathway. HMGCR was highly up-regulated by 15.3-fold, which is consistent with the up-regulation of HMGCR mRNA in a previous study.[70] We identified 26 total peptides, and 18 unique peptides from this protein. After filtering based on the criterion of S/N>3, the final abundance change was calculated from 12 unique quantified peptides. In the mevalonate pathway, the abundance of the upstream protein HMGCS1, HMG-CoA synthase, was elevated by 1.8-fold. The abundances of both ACAT1, acetyl-CoA acetyltransferase in mitochondria, and ACAT2, acetyl-CoA acetyltransferase in cytoplasm, were also increased by 1.4- and 1.5-fold, respectively. Downstream proteins, including MVK, PMVK, MVD, FDPS and GGPS1, were not significantly regulated, as shown in Figure 6.4. The essential intermediate in this pathway, farnesyl-PP, can be converted into different types of lipids including squalene, ubiquinones, sterols, heme A, dolichols, and geranylgeranyl-PP. Interestingly, FDFT1, which catalyzes farnesyl-PP to squalene, was up-regulated by 2.8-fold, and another downstream enzyme, SQLE, squalene monooxygenase, was also up-regulated by 4.4-fold.

Low-density lipoprotein receptor (LDLR), which binds LDL, is the major cholesterol-carrying plasma lipoprotein, and is transported into cells by endocytosis. It has been reported that statin can up-regulate the hepatic LDLR, resulting in lower serum LDL levels.[71] It was also documented that an increase at the LDLR mRNA level occurred in human circulating mononuclear cells as a response to atorvastatin.[72] In the current work, LDLR was quantified to be up-regulated by 2.0-fold in cells treated by atorvastatin, which is correlated with increased cholesterol intake due to intracellular cholesterol synthesis inhibition.

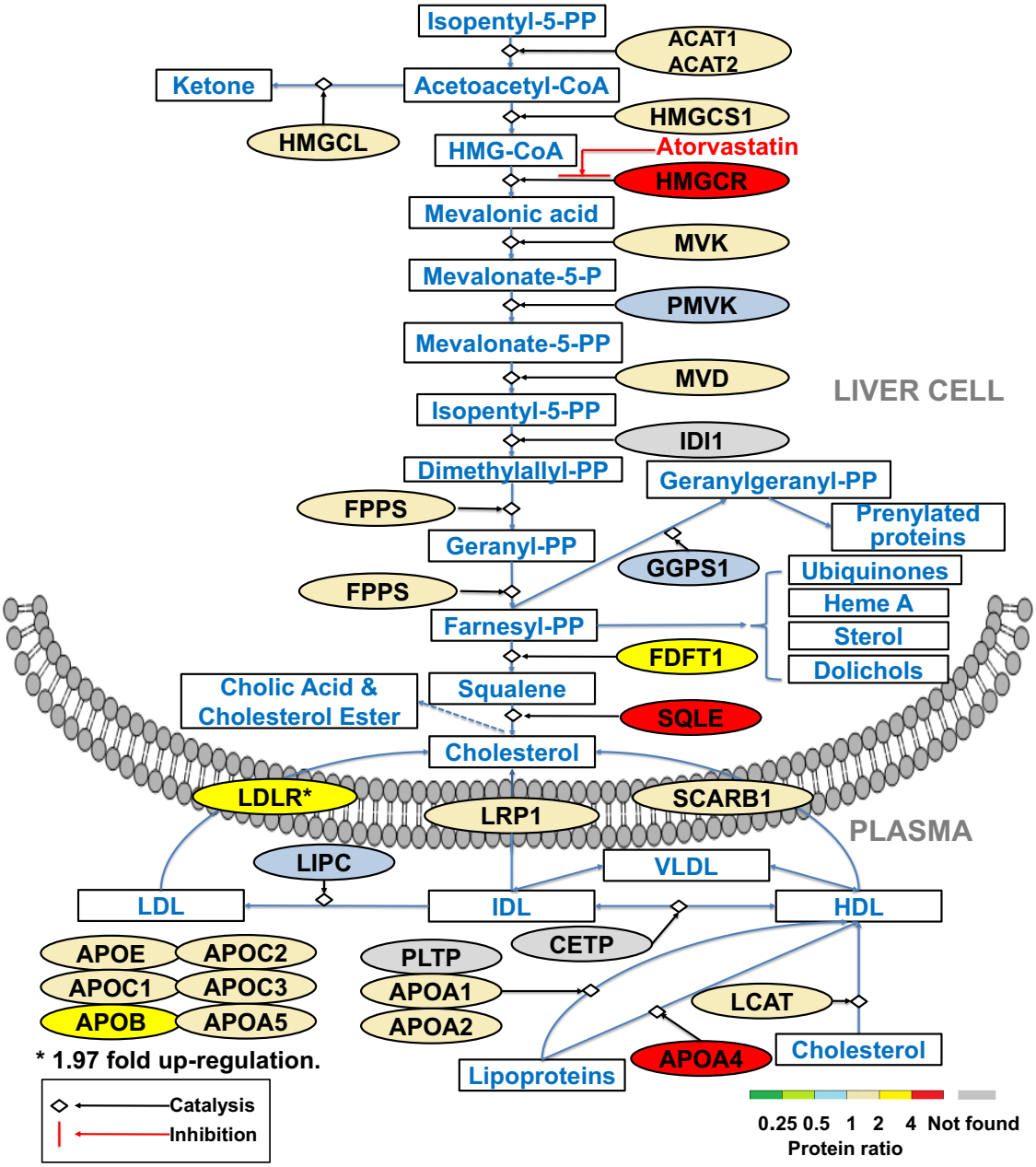


Figure 6.4 Abundance changes of proteins in the mevalonate pathway and some proteins related to cholesterol transportation (all abundance changes refer to intracellular proteins).

APOA4 is another critical protein in the secretion and catabolism of chylomicrons and VLDL. It is required for efficient activation of lipoprotein lipase by ApoC-II and is a potent activator of LCAT. Furthermore, the anti-oxidant and anti-antherogenic properties of APOA4[73, 74] may contribute to the pleiotropic effects of statins. This protein is highly up-regulated by 4.9-fold. Proteome analysis can provide a global view of protein abundance changes in atorvastatin-treated cells. Most proteins in this pathway are not significantly regulated, while several very important proteins, including HMGCR, FDFT1, SQLE, LDLR and APOA4, are up-regulated.

6.1.3.4 Clustering of down-regulated proteins

Among 104 down-regulated proteins, 30 proteins (29%) with gene expression function are enriched with a *P* value of 1.1×10^{-3}. Some proteins related to nucleobase and nucleic acid metabolic process, cytoskeleton organization, macromolecular biosynthetic process, and cell cycle process are also down-regulated. In addition, proteins that have roles in cellular response to stress and DNA damage are also enriched among down-regulated proteins.

For down-regulated proteins, 44 are located in the nucleus, which is the most highly enriched location based on cellular compartment clustering, with a *P* value of 2.8×10^{-5}. In addition, 32 proteins are related to nucleic acid binding with a *P* value of 5.7×10^{-4}. Three proteins corresponding to chronic myeloid leukemia were down-regulated: E2F3, MDM2, and NFKBIA. A group of eight proteins associated with programmed cell death, CFLAR, TOP1, SEMA6A, CASP4, MAP1S, NFKBIA, ACIN1, and CCAR1, were also found to be down-regulated, and they are marked in green in Figure 6.3a. For instance, CCAR1, cell division cycle and apoptosis regulator protein 1, may play an important role in transcriptional regulation and apoptosis signaling. Here we quantified 15 unique peptides from this protein, with its abundance down-regulated by

2.8-fold in cells treated by atorvastatin. CASP4 (Caspase-4) is involved in the activation cascade of caspases responsible for apoptosis execution, and was also decreased by 2.1-fold.

6.1.3.5 Global analysis of protein phosphorylation

For protein phosphorylation, effective enrichment is vital to achieve large-scale identification and quantification by MS. A two-step enrichment method was used for protein phosphorylation analysis. SCX can be used to separate phosphopeptides from non-phophopeptides based on charge. At pH~3, the phosphate group is negatively charged, which makes phosphopeptides carry a lower overall positive charge. Therefore they elute from the SCX column before non-phosphopeptides. Following SCX fractionation, each of the twelve fractions was further enriched by TiO_2. In this experiment, we identified 27,369 total phosphopeptides with a false positive rate of less than 0.1% at the phosphopeptide level, and 15,513 unique phosphopeptides were identified from 2,302 proteins.

The overlap between identified proteins and phosphoproteins is shown in Figure 6.5a. Although the amount of starting material for protein analysis was over 50 times less than that for protein phosphorylation analysis, over 82% overlap between the datasets clearly demonstrated that the protein coverage was very high in the protein analysis. In total, 6,730 proteins were identified in HepG2 cells.

Overall, 9,791 unique phosphopeptides in 2,238 proteins were quantified (listed in a table online at doi.org/10.1021/pr501277g), among which 759 unique phosphopeptides from 354 proteins were down-regulated by at least twofold while 431 phosphopeptides in 266 proteins were up-regulated. Compared to protein abundance changes, phosphopeptide abundance changes were

more dynamic, which is consistent with the fact that proteins have much longer half-lives than protein phosphorylation, which can occur under a second.

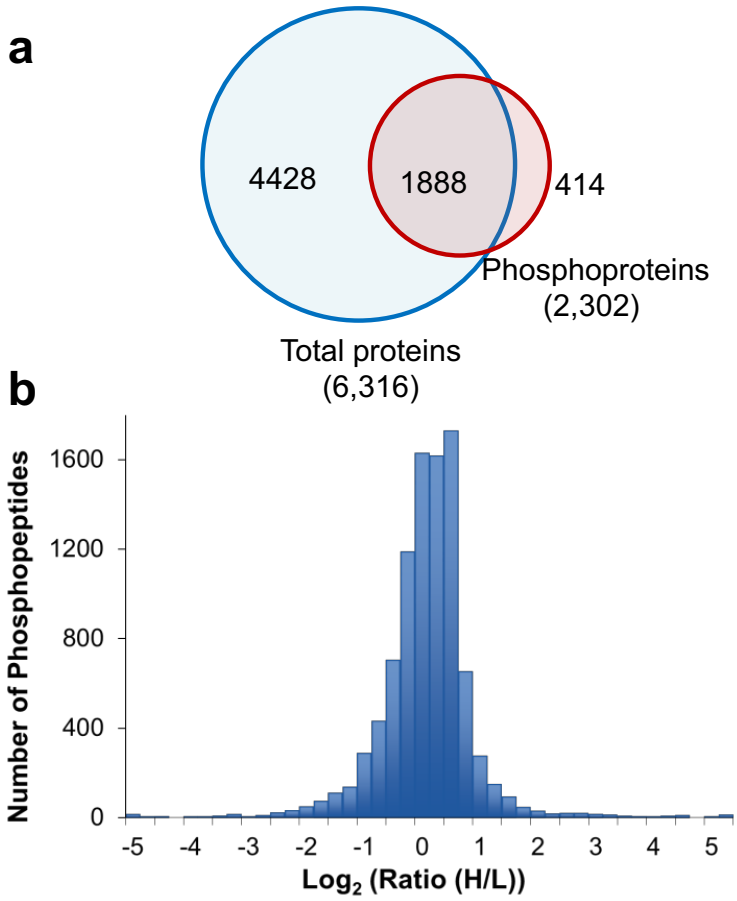


Figure 6.5 (a) Comparison of proteins (6,316) identified in the protein experiment and phosphoproteins (2,302) in the phosphorylation experiment, and (b) the abundance distribution of quantified phosphopeptides.

6.1.3.6 Motif analysis of regulated phosphorylation sites

Due to the specific structures of kinase catalytic domains, every type of kinase has a preferential motif surrounding the phosphorylation sites of substrates. Motif analysis of regulated phosphorylation sites may provide valuable information regarding the effect of atorvastatin on the activities of kinases in cells. In order to pinpoint the up- and down-regulated phosphorylation sites,

only singly phosphorylated peptides with well-localized phosphorylation sites (Ascore > 13) were considered. Based on the dataset of over 9,000 quantified unique phosphopeptides, 3,409 phosphopeptides met these requirements, among which 218 were down-regulated and 106 were up-regulated. The motif analysis results showed that the [S/T]P motif was found among both up- and down-regulated phosphorylation sites, which is the typical motif of cyclin-dependent kinases (CDKs) and mitogen-activated protein kinases (MAPKs). The up- and down-regulated sites with the [S/T]P motif could be from the direct and indirect effects of atorvastatin. Besides [S/T]P, no other motif was identified among up-regulated phosphorylation sites.

In contrast, several other types of motifs were found exclusively among down-regulated phosphorylation sites, including SPK, SD, E….S, K.T, K..S and K….S (“.” represents any amino acid residue), as shown in Figure 6.6. Basophilic enzymes such as PKA and PKC can recognize basic side chains preceding the target serine or threonine residues. Three identified motifs have a K before down-regulated phosphorylation sites. For example, K..S is the preferred motif of PKC.[75] PKC family members play significant roles in a variety of intracellular signal transduction processes, and are involved in receptor desensitization, modulating membrane structure events, regulating transcription, mediating immune responses, and regulating cell growth. These functions are achieved by PKC-mediated phosphorylation of their protein substrates. PKC, activated by tumor promoter phorbol ester, may phosphorylate potent activators of transcription, and thus lead to increased expression of oncogenes, promoting cancer progression.[76] Here, phosphorylation sites with the preferred basic motif K..S were enriched among down-regulated sites in cells treated by astorvastatin, which suggests that the activity of PKC was attenuated by the statin.

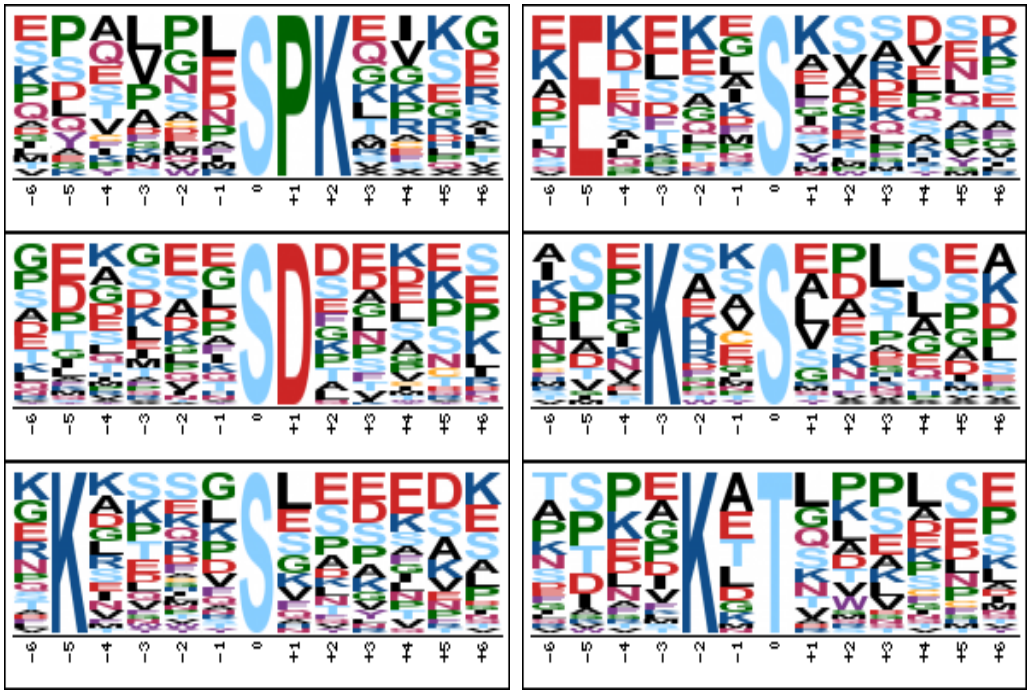


Figure 6.6 The results of motif analysis among down-regulated phosphorylation sites.

6.1.3.7 Pathway analysis based on regulated protein phosphorylation

Phosphoproteins containing up- or down-regulated phosphopeptides were clustered using DAVID.[69] Proteins related to gene expression, cell cycle and macromolecular metabolic process are highly enriched among up- and down-regulated phosphoproteins. For example, 99 out of 354 down-regulated phosphoproteins were related to gene expression with a *P* value of 1.9×10^{-8}, while 68 of 266 up-regulated phosphoproteins were found to belong to this category ($P=3.6 \times 10^{-4}$). Regulation of gene expression by protein phosphorylation is very complex in cells treated by statin.

Up-regulated phosphoproteins: Similar to the protein abundance changes, after statin treatment, phosphopeptides from several proteins with functions relating to metabolism of lipids and lipoproteins were up-regulated, including HMGCR, HMGCS1, HMGCL, FDFT1, FASN, HADHA and LDLR. Protein activities are often regulated by their phosphorylation and the up-

regulation of protein phosphorylation very likely corresponds to the increased activities of these proteins.

While statins inhibit HMGCR to lower cholesterol, the synthesis of other lipids through the mevalonate pathway is also inhibited, including farnesy-PP and geranylgeranyl-PP. Therefore, protein lipidation will be dramatically impacted. Lipidation of proteins plays critical roles in protein localization and signal transduction, for example the lipidated proteins Ras, Rho and Rap.[13] In this experiment, twelve G-protein modulators that contained up-regulated phosphopeptides were quantified, listed in Table 6.2. For example, ARHGAP11A, Rho GTPase-activating protein 11A, is involved in the regulation of small GTPase mediated signal transduction. Four unique phosphopeptides were quantified, but only the peptide AGCFS@PK (@ refers to the phosphorylation site) with a site at S422 was highly up-regulated by 6.6-fold. This site contains the SP motif. The doubly phosphorylated peptide with sites at S718 and S719 has a ratio of 1.5, and the other two phosphopeptides with sites at S484 or S868 have ratios of 1.0 and 1.3, respectively. Without protein lipidation, Ras, Rho and Rap cannot be localized to the plasma membrane and correspondingly cannot transduce signals effectively, which may be the explanation of why many G-protein modulators have up-regulated phosphorylation. In addition, some phosphopeptides from these G-protein modulators were down-regulated (listed in a table online at doi.org/10.1021/pr501277g) as a result of atorvastatin treatment. It is well-known that signal transduction by protein phosphorylation is extremely intricate in cells. The current experiments further demonstrated that the statin treatment impacts G-protein related signal transduction.

Table 6.2 Up-regulated phosphopeptides from G-protein modulators.

Reference	Gene symbol	Phosphopeptide	PPM	XCorr	Site	Ascore	Peptide ratio	Annotation
P53367	ARFIP1	K.LKT@PGVDAPSWLEEQ	0.8	2.98	361	78.2	9.5	**Arfaptin-1, interacts with GTP-bound ARF3**
Q6P4F7	ARHGAP11A	K.AGCFS@PK	-0.1	1.72	422	1000	6.6	**Rho GTPase-activating protein 11A**
		SS@QDGDTPAQASPPEEK	0.7	3.36	456	0.0	16.6	
Q6ZUM4	ARHGAP27	SSQDGDTPAQAS@PPEEK#	0.8	2.60	466	17.3	2.0	**Rho GTPase-activating protein 27**
Q9UBC2	EPS15L1	DSLRSTPS@HGSVSSLNSTGSLS@PK#	0.5	2.87	241, 255	6.0, 5.7	3.0	**Epidermal growth factor receptor substrate 15-like 1**
P85299	PRR5	FMSSPSLS@DLGK#	1.7	1.67	16	6.6	7.1	**Subunit of mTORC2, which regulates cell growth and survival.**
Q9NYI0	PSD3	SHS@SPSLNPDT@SPITAK#	-1.2	3.78	1011, 1019	9.3, 0.0	3.4	**PH and SEC7 domain-containing protein 3**
Q96QF0	RAB3IP	STSSAMSGS@HQDLSVIQPIVK	0.5	2.77	296	0.0	2.2	**Rab-3A-interacting protein**
Q9H6Z4	RANBP3	NESSNAS@EEEACEK	0.0	4.47	244	73.3	5.3	**Ran-binding protein 3**
Q684P5	RAP1GAP2	QEVFVYSPSPSSESPS@LGAAATPIIMSRSPTDAK#	-1.5	3.96	687	4.4	3.7	**Rap1 GTPase-activating protein 2**
Q92609	TBC1D5	SQAPVCSPLVFSDPLMGPASASSSNPSS@SPDDDSSK	-0.1	4.89	775	0.0	3.3	**TBC1 domain family member 5, a GTPase-activating protein for Rab family protein(s)**
		DSLSVSSNDAS@PPASVASLQPHMIGAQSS@PGPK	3.2	4.12	1745, 1763	7.4, 9.3	64.1	**Triple functional domain protein, promotes the exchange of GDP by GTP, positive regulation of GTPase activity**
O75962	TRIO	DSLSVSSNDASPPAS@VASLQPHMIGAQSSPGPK	0.4	4.91	1749	2.6	43.3	

The activities of kinases are often regulated by their phosphorylation, and many phosphopeptides from several kinases were up-regulated, including EPHA2, PTK2, CAMK1, CDK1, CAMK2D, MAP3K4, PAK2, PKN1, RPS6KC1, STK33 and TLK1, TGFBR2 and TPR. MAP3K4, mitogen-activated protein kinase kinase kinase 4, is a component of a protein kinase signal transduction cascade and activates the CSBP2, P38 and JNK MAPK pathways. Three out of four quantified phosphopeptides were up-regulated by 2.0-2.5-fold, respectively. The three up-regulated peptides have the same sequence, but different phosphorylation sites. This could be due to incorrect site assignments because of insufficient fragment information in the MS^2 spectra. TLK1, a serine/threonine-protein kinase tousled-like 1, is rapidly and transiently inhibited by

phosphorylation following the generation of DNA double-stranded breaks during S-phase.[77, 78] In our experiment, two triply phosphorylated peptides were up-regulated, which suggested that the activity of TLK1 was attenuated in cells treated by atorvastatin. However, several quantified phosphopeptides from TLK2 were not regulated. TGFBR2 is a transmembrane serine/threonine kinase forming with the TGF-beta type I serine/threonine kinase receptor, TGFBR1, the non-promiscuous receptor for the TGF-beta cytokines TGFB1 (transforming growth factor beta-1), TGFB2 and TGFB3. It transduces the TGFB1, TGFB2 and TGFB3 signal from the cell surface to the cytoplasm and is thus regulating a plethora of physiological and pathological processes including cell cycle arrest in epithelial and hematopoietic cells, control of mesenchymal cell proliferation and differentiation, wound healing, extracellular matrix production, immunosuppression and carcinogenesis (http://www.uniprot.org/uniprot/P37173). The C-terminal peptides (located in the cytoplasm) with the phosphorylation sites at S562 and T566 were up-regulated by 2.8-fold. More detailed information about quantified phosphopeptides in kinases is included in listed in a table online at doi.org/10.1021/pr501277g.

Down-regulated phosphoproteins: In this experiment, more phosphopeptides were down-regulated than up-regulated. In addition to many phosphoproteins related to gene expression and the cell cycle, other interesting pathways, including the splicesome, tight junction, apoptosis, and CARM1 (coactivator-associated arginine methyltransferase 1) and regulation of estrogen receptor pathways, contain down-regulated phosphoproteins.

Tight junctions, also known as occluding junctions, are the closely associated areas of two cells whose membranes join together forming a virtually impermeable barrier to fluid. In the tight junction pathway, phosphopeptides from seven proteins were found to be down-regulated, including EPB41L2, INADL, CTTN, PARD3, OCLN, CGN and MLLT4. Down-regulation of

these phosphoproteins in the tight junctions may interfere with cell-cell interactions, which may be related to the pleiotropic effects of statins.

Four proteins (SRA1, BRCA1, MED1 and POLR2A) that have roles in the CARM1 and regulation of estrogen receptor pathway were found to contain down-regulated phosphopeptides in cells treated by the statin. SRA1, steroid receptor RNA activator 1, enhances cellular proliferation and differentiation, and promotes apoptosis *in vivo* and may play a role in tumorigenesis.[79] It is highly expressed in the liver and skeletal muscle and is up-regulated in human tumors of the breast, ovary, and uterus. At the protein level, SRA1 is slightly down-regulated with a ratio of 0.77, which is similar to literature reports.[57] In the phosphorylation experiment, one peptide RVAAPQDGS@PRVPAS@ETSPGPPPMGPPPPSSK was down-regulated by 2.5-fold, but the other singly phosphorylated peptide at the C-terminus of the protein was not regulated. MED1 is a component of the Mediator complex, and a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. In cells treated by atorvastatin, the abundance of MED1 was decreased by a ratio of 0.62, which is in a very good agreement with the ~ 30% decrease previously reported in HL-60 cells treated by lovastatin.[57] Based on the quantification of protein phosphorylation, 21 out of 24 unique quantified phosphopeptides were down-regulated in MED1. Our phosphorylation results clearly demonstrate that atorvastatin has an impact on the CARM1 and regulation of estrogen receptor pathway, which may contribute to its anti-cancer activity.

6.1.4 Conclusions

Statins are the most common and effective drugs for lowering cholesterol in patients. The have pleiotropic effects possibly due to off-target effects and/or secondary effects from lipid

synthesis inhibition in the mevalonate pathway. Here we systemically investigated the protein and protein phosphorylation abundance changes in HepG2 cells treated by atorvastatin. Over 6,000 proteins were quantified, but only a very small portion of them were regulated, *i.e.* 104 down-regulated and 81 up-regulated. As expected, many lipid-related proteins were up-regulated, including HMGCR, FDFT, SQLE and LDLR, while proteins related to gene expression, cellular response to stress and apoptosis were down-regulated. We quantified almost 10,000 unique phosphopeptides, which were more dynamic than proteins. The protein phosphorylation results demonstrate that many proteins with gene expression and cell cycle function have regulated phosphorylation, including both up- and down-regulated phosphorylation. Several basic motifs found among down-regulated sites indicated that kinases with preferences for these motifs have attenuated activities, including PKA and PKC. In addition to phosphoproteins related to lipid metabolism, phosphopeptides on a group of G-protein modulators were up-regulated, which may be due to cell signal transduction changes resulted from the effect of protein lipidation by the statin. Phosphopeptides from several proteins related to the tight junction, apoptosis, and CARM1 and regulation of estrogen receptor pathways were down-regulated. MS-based proteomics techniques provide an ideal way to gain insight into the protein and modified protein changes in cells treated by statin. A more comprehensive understanding of the cellular response to statins and the underlying molecular mechanisms of their pleiotropic effects will be beneficial in applying them to treat noncardiac vascular diseases, and minimizing their potential side effects.

6.2 Mass Spectrometric Analysis of the Human N-glycoproteome in Statin-Treated Liver Cells with Two Lectin-Independent Chemical Enrichment Methods

6.2.1 Introduction

Protein glycosylation is critical in determining protein folding, trafficking, stability and activity [80, 81]. Among multiple types of protein glycosylation, N- and O-linked glycosylation are the two major types [82, 83]. N-linked glycosylation occurs on the side chain of the asparagine residue and often has an N-X-S/T/C (X stands for any amino acid residues other than proline) [84, 85], while O-linked glycosylation is on the side chains of serine and threonine residues [86-88]. N-glycosylation typically begins with the synthesis of the dolichol-linked precursor oligosaccharide ($GlcNAc_2Man_9Glc_3$), followed by *en bloc* transfer of the precursor oligosaccharide to newly synthesized peptides in the endoplasmic reticulum (ER) [89, 90]. Due to its importance in biological systems [91-93], N-glycosylation has also brought extensive attention for its role in human disease, such as Alzheimer's disease (AD), cancer, and infectious diseases [92, 94, 95].

With the development of mass spectrometry (MS) instrumentation and computation techniques, current MS-based proteomics is very powerful in analyzing protein modifications, including glycosylation, in complex biological samples [96-107]. Due to the low abundance of many glycoproteins, sub-stoichiometry of protein glycosylation, and the complexity of biological samples, it is imperative to enrich glycoproteins prior to MS analysis [106, 108, 109]. Conventional lectin-based enrichment methods have been extensively used [110, 111]. However, due to the binding specificity of lectin, no single or several types of lectin can cover all glycoproteins that have highly diverse glycans in human cells.

In recent years, several very elegant methods have been developed and tremendously advanced the glycoproteomics field [85, 106, 107, 112-114]. In this work, we systematically compared two lectin-independent chemical methods to enrich and analyze glycoproteins in human cells: one based on boronic acid and *cis*-diol interactions [108, 115] and the other benefited from metabolic labeling and click reaction [116-118]. For the first method, we utilized the universal and reversible interactions between boronic acid and sugar molecules. Boronic acid and *cis*-diols can form reversible covalent bonds in basic solutions, and conversely, the bonds are prone to cleavage under acidic conditions. The reversible nature of this bond ensures that glycopeptides can be effectively released after capturing. The second method takes advantage of the endogenous glycoprotein synthesis machinery to incorporate a chemical handle into glycans for further click chemistry and biotin avidin-based glycopeptide enrichment. An unnatural sugar analog containing an azide group was employed to feed cells in order to generate the chemical handle mentioned above. Comparing to BA, we reasoned that MC has the advantage of better reflecting the dynamic changes in cells since only the newly-synthesized glycoproteins are labeled by the sugar analog, while BA may be more universal for glycoprotein enrichment.

In this work, we designed an experiment to comprehensively compare the identification and quantification of glycoproteins with these two methods. We analyzed the glycoproteome changes in statin-treated liver cells using these two methods. Statins are a group of cholesterol-lowering drugs that target 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCR), which is the rate-limiting enzyme of the mevalonate pathway. Upon inhibition of HMGCR, the synthesis of many intermediate and end products in this pathway was affected, which induced many well-known pleiotropic effects of statins. Dolichol is one of the end products and is involved in protein N-glycosylation, functioning as the membrane anchor for precursor oligosaccharides formation.

Therefore, we expected that protein N-glycosylation was attenuated in statin-treated cells. Using liver cells (HepG2) as a biological model, we systematically evaluated the performance of these two methods and explained the underlying mechanisms for the differences observed. The current work may provide useful information for future selection of enrichment methods to study the cell glycoproteome under different circumstances.

6.2.2 Experimental section

6.2.2.1 Cell Culture and metabolic labeling

HepG2 (C3A) cells (from American type culture collection (ATCC)) were grown in "heavy" and "light" SILAC (stable isotope labeling with amino acids in cell culture) Dulbecco's modified eagle's medium (DMEM) (Sigma-Aldrich) for five generations before treatment with a statin. The medium also contained 1000 mg/L glucose and 10% dialyzed fetal bovine serum (diFBS) (Corning). "Heavy" and "light" SILAC media were freshly prepared by adding 0.146 g/L $^{13}C_6{}^{15}N_2$ L-lysine (Lys-8) and 0.84 g/L $^{13}C_6$ L-arginine (Arg-6) (Cambridge Isotope Lab) or the corresponding non-labeled L-lysine (Lys-0) and L-arginine (Arg-0). When cells reached about 70% confluency, we switched to SILAC media without diFBS and added 15 μM atorvastatin to the heavy group. Meanwhile, dimethyl sulfoxide (DMSO) was used to treat the light group as a vehicle control. For the MC experiments, 100 μM tetra-acetylated N-azidoacetylgalactosamine (Ac_4GalNAz) (Click Chemistry Tools) was added into both heavy and light cells at the statin or mock treatment time. Cells were then maintained in a humidified incubator at 37 °C and 5.0% CO_2 for 24 h.

6.2.2.2 Cell lysis, click reaction, and protein digestion

Cells were washed twice with phosphate buffered saline (PBS), harvested by scraping in PBS, and pelleted by centrifugation at 500 g for 3 min. Two trial runs using about 2% of total cells were conducted to calibrate the heavy and light cell ratios in the BA and MC experiments. For the real experiments, heavy and light cells were mixed based on the protein ratio of 1:1 according to the results from the trial runs. The cell pellets were lysed through end-over-end rotation at 4 ºC for 45 minutes in a lysis buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH=7.6, 150 mM NaCl, 0.5% sodium deoxycholate (SDC), 25 units/mL benzonase and 1 tablet/ 10 mL protease inhibitor (EDTA-free, Roche)). Lysates were centrifuged, and the resulting supernatant was transferred to new tubes. For the MC experiment, the supernatant was reacted with 100 μM dibenzocyclooctyne (DBCO)-sulfo-biotin to have the metabolically labeled glycoproteins tagged with biotin through the specific click reaction between the azido group and DBCO [119-121]. Proteins were subjected to disulfide reduction with 5 mM 1,4-dithiothreitol (DTT) (56 ºC, 25 min) and alkylation with 14 mM iodoacetamide (RT, 20 min in the dark). Detergent was removed by the methanol-chloroform protein precipitation method. The purified proteins were digested with 10 ng/μL Lys-C (Wako) in 50 mM HEPES pH 8.2, 1.6 M urea, 5% ACN at 31 ºC for 16 h, and 10 ng/uL trypsin (Promega) at 37 ºC for 4 h.

6.2.2.3 Glycopeptide separation, enrichment and deglycosylation

Digestion mixtures were acidified by addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%, clarified by centrifugation, desalted using tC18 SepPak cartridge (Waters), and lyophilized. For the BA experiment, purified peptides were dried and enriched with boronic acid-conjugated magnetic beads (Figure 6.7a). Briefly, beads were washed three times with 100

mM ammonium acetate. Peptides were dissolved in the same buffer and mixed with the beads. The mixture was incubated in a shaking incubator at room temperature for an hour, and the beads were washed with the buffer mentioned above to remove non-glycopeptides. Finally, the beads were eluted with ACN:H_2O:TFA = 49:50:1 with shaking. The elution was lyophilized and purified with tC18 SepPak cartridge, dried overnight, and treated with four units of peptide-*N*-glycosidase F (PNGase F, Sigma-Aldrich) in 80 μL buffer containing 50 mM NH_4HCO_3 in heavy oxygen water ($H_2{}^{18}O$) at 37 °C for 3 h. The reaction was quenched by addition of 1% TFA to pH~2, desalted, and dried. The glycopeptides were fractionated by high pH reversed-phase high-performance liquid chromatography (HPLC) into 10 fractions with a 40-min gradient of 5-55% ACN in 10 mM ammonium acetate (pH=10). The fractions were dried and further purified with the stage-tip method.

For the MC experiment, purified and dried peptides were enriched with NeutrAvidin beads (Thermo) at 37 °C for 30 min (Fig. 5.7b). The samples were transferred to spin columns and washed according to manufacturer's protocol. Peptides were eluted from the beads by 3-min incubations with 300 μL of 8 M guanidine-HCL (pH=1.5) at 56 °C three times. Eluates were combined, desalted using tC18 SepPak cartridge, and lyophilized overnight. Dried peptides were deglycosylated as described in the BA experiment and quenched using the same method. Subsequently, we also attempted to fractionate the glycopeptide sample using HPLC, but the results were not ideal because the enriched sample amount in the MC experiment was much lower than the sample amount in the BA experiment since only the newly-synthesized and metabolically labeled glycopeptides were enriched. Finally, we fractionated the deglycosylated peptides during the stage-tip step, and the sample was separated into 3 fractions using the elution buffer with 20%, 50% and 80% ACN, respectively, containing 1% HOAc.

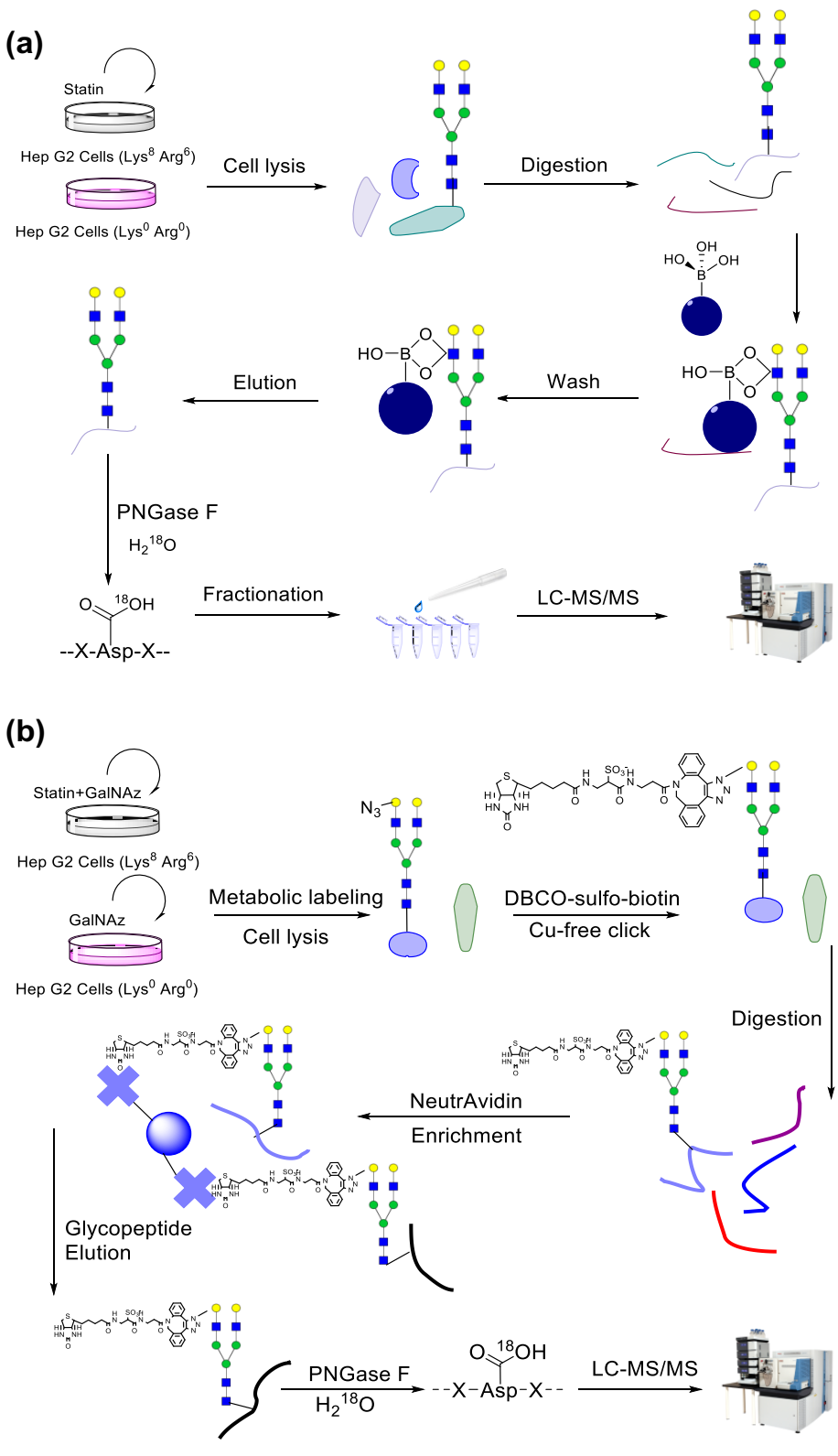


Figure 6.7 Experimental schemes of the (a) BA and (b) MC experiments.

6.2.2.4 LC-MS/MS analyses

Purified and dried peptide samples were dissolved in 10 μL solution of 5% ACN and 4% formic acid (FA) each, and 4 μL of the resulting solutions were loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 3 μm, 200 Å, 100 μm x 16 cm, Michrom Bioresources) by Dionex WPS-3000TPLRS autosampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Wellplate Sampler). Peptides were separated by reversed-phase chromatography using UltiMate 3000 binary pump with a 110 min gradient with increasing concentration of ACN (in 0.125% FA). Peptides were detected with a data-dependent Top20 method [42, 60] in a hybrid dual-cell quadrupole linear ion trap - Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher, with Xcalibur 3.0.63 software). For each cycle, one full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 20 MS/MS in the LTQ for the most intense ions. The selected ions were excluded from further analysis for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

6.2.2.5 Database searches and data filtering

Raw data files from the mass spectrometer were converted into mzXML format, and precursor ion mass measurements were refined by checking the monoisotopic peak assignments [50]. All spectra were searched using the SEQUEST algorithm (version 28) [61] and matched against a database encompassing sequences of all proteins in the UniProt Human (*Homo sapiens*) database containing common contaminants. Each protein sequence was listed in both forward and reverse orders to control the false discovery rate (FDR) of glycopeptide identifications. We performed the database search using the following parameters: 10 ppm precursor mass tolerance; 1.0 Da product

ion mass tolerance; fully digested with trypsin; up to two missed cleavages; variable modifications: oxidation of methionine (+15.9949), O^{18} tag of asparagine (+2.9883), heavy lysine (+8.0142), and heavy arginine (+6.0201); fixed modifications: carbamidomethylation of cysteine (+57.0214).

The target-decoy method was employed to estimate and control FDRs at the glycopeptide levels [62, 63]. Through linear discriminant analysis (LDA), which is similar to other methods reported in the literature [64-66], several parameters (such as XCorr, ΔCn, precursor mass error, and charge state) were used to distinguish correct and incorrect peptide identifications [50]. After scoring, peptides shorter than six amino acid residues were removed, and the dataset was restricted to glycopeptides when determining FDRs. Glycopeptide FDRs were filtered to <1 % based on the number of decoy sequences in the final data set.

6.2.2.6 Glycopeptide quantification and glycosylation site localization

For peptide quantification, we required an S/N value larger than 3 for both heavy and light species. If the S/N value of one member of a heavy and light pair was less than 3, the partner was required to be greater than 5. A probabilistic algorithm was used to localize N-glycosylation sites and to estimate the assignment confidence [67, 122]. A ModScore was calculated for each glycosylation site, and sites with a ModScore >13 (P<0.05) were considered to be confidently localized.

6.2.3 Results and discussion

6.2.3.1 Examples of glycopeptide identification

The enriched glycopeptides were treated with PNGase F in heavy oxygen water to remove N-glycans and to generate a common tag. When this enzymatic reaction occurs in heavy oxygen

water, the converted Asp from the glycosylation site contains heavy oxygen, which creates a mass shift of +2.9883 Da for MS analysis. In this case, heavy oxygen on Asp enabled us to distinguish authentic N-glycosylation sites from those caused by spontaneous asparagine deamidation, which may happen *in vivo* and during sample preparation. It could also occur during PNGase F treatment, which may result in false positive identifications of protein N-glycosylation sites. To minimize false positive identifications, we ran the reaction for 3 h, during which the effect of deamidation was nearly negligible [118].

Two examples of N-glycopeptide identifications are shown in Figure 6.8. Formerly glycosylated peptide YHYN*GTLLDGTLFDSSYSR@ (*-N-glycosylation site, @-heavy arginine) was confidently identified with XCorr of 5.9 from the BA experiment (Figure 6.8a). This peptide is from protein FKBP9, one of the peptidyl-prolyl cis-trans isomerases (PPIases) that accelerates the folding of proteins during protein synthesis. The other deglycosylated peptide SSCGKEN*TSDPSLVIAFGR shown in Figure 6.8b is from protein LAMP1- lysosome-associated membrane glycoprotein 1, which has the major function of presenting carbohydrate ligands to selectins. This peptide was identified in the MC experiment with an even higher XCorr of 6.2 and mass accuracy of -0.23 PPM. The site N84 is confidently identified to be glycosylated with ModScore=1,000, and the score of 1,000 means that only one possible glycosylation site exists on the identified glycopeptide.

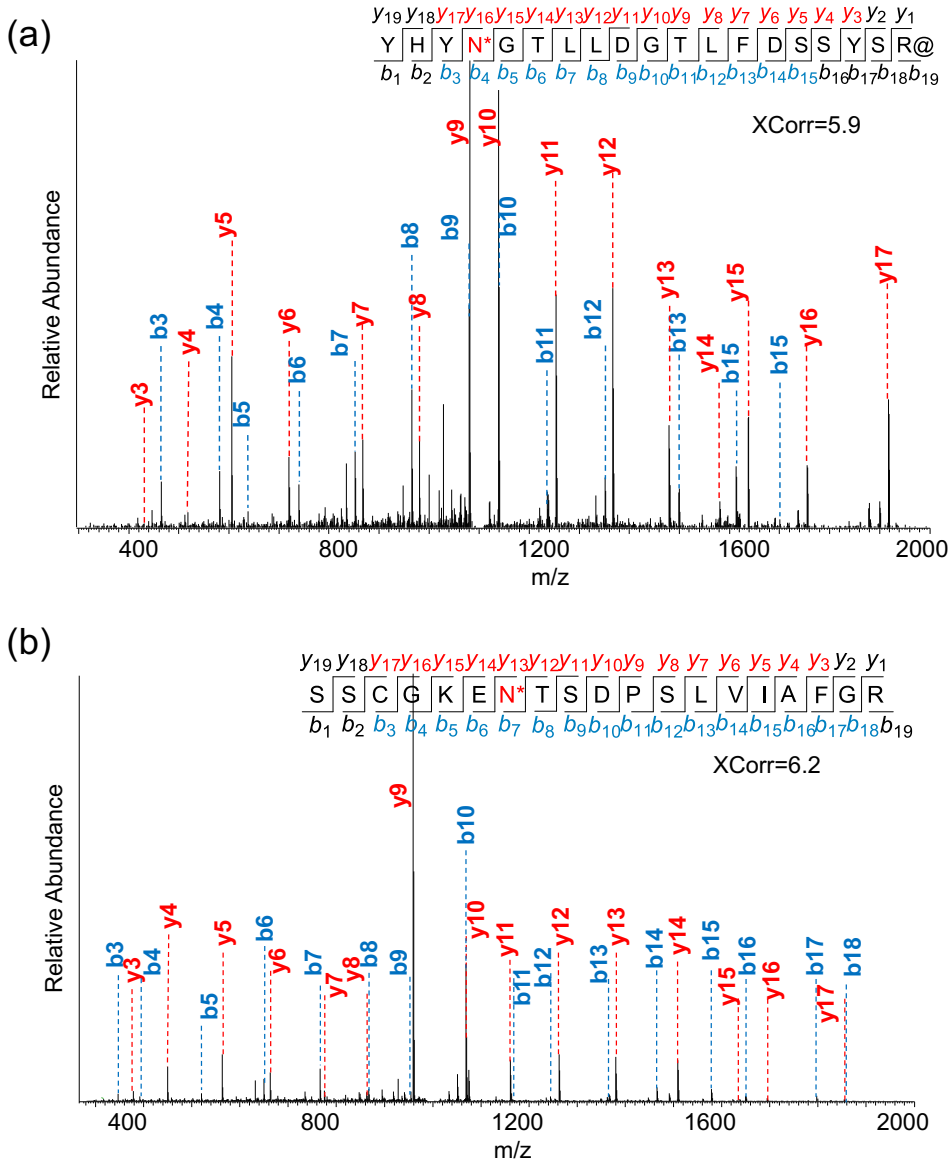


Figure 6.8 Tandem mass spectra of (a) the glycopeptide YHYN*GTLLDGTLFDSSYSR@ (*-N-glycosylation site, @-heavy arginine) from protein FKBP9 identified in the BA experiment, and (b) the glycopeptide SSCGKEN*TSDPSLVIAFGR from protein LAMP1 identified in the MC experiment.

In this work, a total of 2,641 unique formerly glycosylated peptides were identified with the boronic acid-based enrichment method and 1,493 with the method combining metabolic labeling with click chemistry. These results indicated that the boronic acid enrichment method is more universal.

6.2.3.2 N-glycosylation sites identified with the two lectin-independent enrichment methods

Boronic acids and sugars can form reversible covalent bonds. Based on this universal reaction, our lab has employed this chemical enrichment method to analyze the yeast glycoproteome [108]. The results demonstrated that this method can be used to effectively enrich glycopeptides from digested whole cell lysates. The potential pitfall of this reaction is that the interactions are relatively weak, which could affect the enrichment of glycopeptides from low-abundance glycoproteins.

Metabolic labeling can be employed to label proteins and/or modified proteins, and the labeled proteins and modified proteins may bind to fluorophore for visualization or be selectively enriched for further analysis [116, 123]. In this study, we incorporated an azide-containing sugar analog (Ac_4GalNAz) into glycans in glycoproteins, and this azide group was used as a chemical handle to tag a biotin molecule onto the metabolically labeled glycans through click chemistry [124, 125]. Tagging glycoproteins with biotin allowed further glycopeptide enrichment through the strong interaction between biotin on labeled peptides and NeutrAvidin beads after cell lysis and digestion. The detailed experimental procedure is shown in Figure 6.7b. Stringent wash was employed to remove non-glycopeptides. Compared to BA, MC has more steps, and the enrichment is largely dependent on metabolic labeling and click reaction efficiency. Since only sugar analog-labeled

peptides were enriched with the MC method, it can largely minimize sample complexity, which is an advantage when investigating cellular responses to drug treatment.

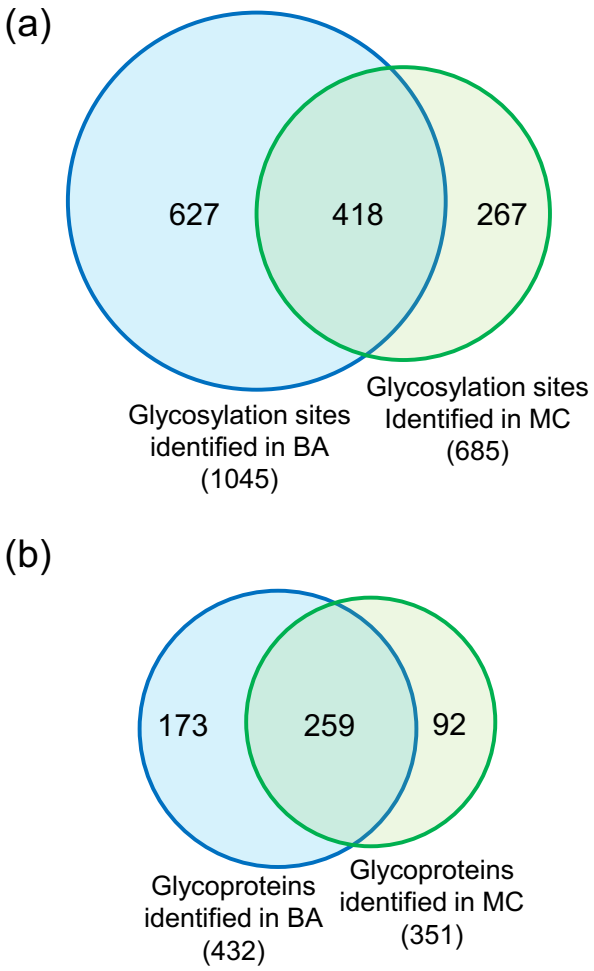


Figure 6.9 Comparison of glycosylation sites (a) and glycoproteins (b) identified using the two enrichment methods.

For glycosylation site identification, in addition to running the PNGase F treatment for three hours and filtering glycopeptides with <1% FDR, we also applied another criterion: all glycosylation sites must have the consensus motif of NXS/T/C [85] (X stands for a random amino

acid other than proline). We confidently identified 1,045 N-glycosylation sites on 432 proteins with the boronic acid-based enrichment method, and the sites are listed in a table online at doi.org/10.1016/j.ijms.2017.05.010. Using the enrichment method combined with metabolic labeling and click chemistry, 685 N-glycosylation sites were identified on 351 proteins (listed in a table online at doi.org/10.1016/j.ijms.2017.05.010). 418 common sites were identified in both experiments (Figure 6.9a). Many proteins contain one glycosylation site while some proteins are highly glycosylated. For instance, LRP1, prolow-density lipoprotein receptor-related protein 1 is a large protein with molecular weight 504,606 Da and 4,454 amino acid residues. We identified very similar number of glycosylation sites on this protein through the two experiments: 21 sites in BA and 20 in MC. As expected, the overlap at the protein level was higher, and 259 common glycoproteins were identified from the two experiments (Figure 6.9b).

6.2.3.3 Protein clustering based on molecular function

We clustered the glycoproteins identified exclusively in either BA or MC experiment according to the molecular function analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [69] (Figure 6.10). Interestingly, we found that the most enriched categories for glycoproteins identified in the BA experiment are intracellular enzyme activity-related, such as hydrolase and transferase activities. However, among glycoproteins identified in the MC experiment, the top enriched categories are binding, receptor, and molecular transducer activities. These activities are known to occur prominently on the cell surface. The differences may be attributed to the following reasons. The boronic acid-based enrichment method is universal, which may unbiasedly enrich cell surface and intracellular glycoproteins. However, the enrichment method based on metabolic labeling and click chemistry is very dependent on the metabolic

labeling efficiency, and the latter relies on the endogenous glycan synthesis machinery. In this work, we used Ac_4GalNAz to feed the cells and labeled the glycans containing GalNAc or GlcNAc. Normally, cell surface glycoproteins have mature glycan structures in order to be transported to the plasma membrane and/or be secreted, thus these glycans are more likely to have GalNAc moieties that can be substituted by GalNAz.

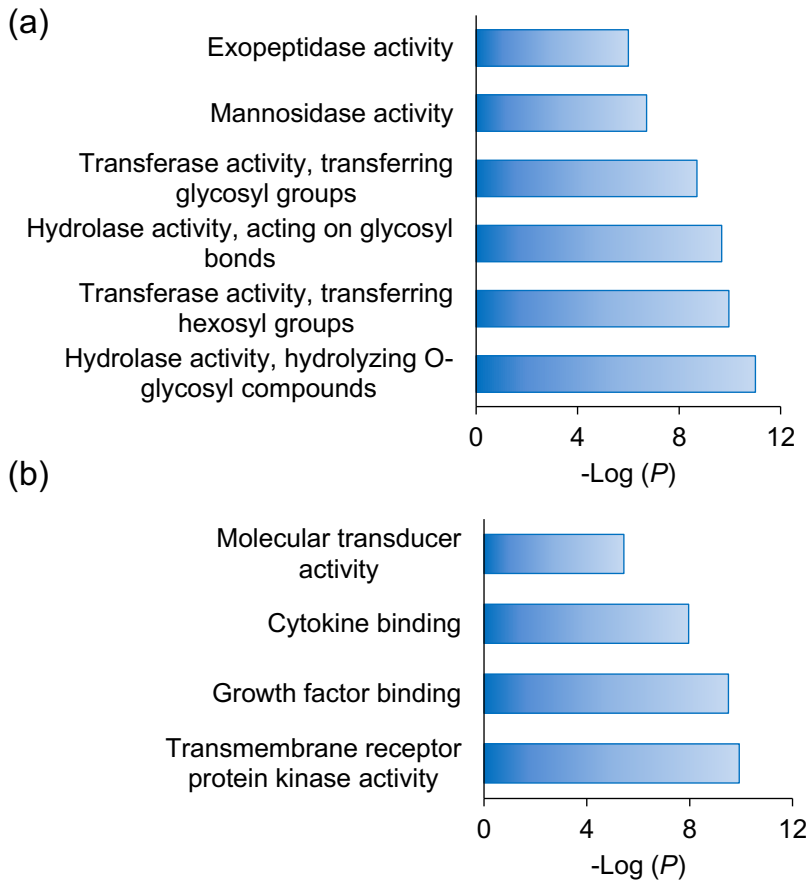


Figure 6.10 Clustering of the glycoproteins identified only in the (a) BA or (b) MC experiment based on molecular function.

Although all N-Glycans have GlcNAc, GalNAz must convert into GlcNAz before labeling. Therefore, the labeling of GlcNAz may not be as efficient as GalNAz over a relatively short labeling period. For intracellular glycoproteins, especially those still in the ER and Golgi, because their glycan structures are likely immature, the chance of being labeled would be lower than those on the cell surface. Overall, BA is a more global and universal method, while MC has better performance on the identification of glycoproteins located on the cell surface.

6.2.3.4 Quantification of cell glycoproteome changes in statin-treated cells

Statins are a family of popular drugs for lowering cholesterol, but they may affect protein N-glycosylation because the inhibition of HMGCR by statins also prevents the synthesis of other products in the mevalonate pathway, including ubiquinone, dolichol, and farnesyl-pyrophosphate (farnesyl-PP) [7]. Dolichol is essential to protein N-glycosylation in the form of dolichyl phosphate (Dol-P), which serves as the carrier in pyrophosphate-linked oligosaccharide assembly as well as acting as the acceptor in the synthesis of the sugar donors Dol-P-Man and Dol-P-Glc from GDP-Man and UDP-Glc, respectively. Thus, protein N-glycosylation is expected to be impacted while the dolichol synthesis is hindered by the statin treatment. Perturbation of protein N-glycosylation by statins may contribute to the well-known "pleiotropic effects" of statins [7, 126]. Systematic and quantitative investigation of protein N-glycosylation changes by statins will provide insight into the molecular mechanisms of the pleiotropic effects and allow patients to benefit further from the drug.

Statin is a relatively mild drug, and patients typically take it for months or years [3]. Here, we used it to treat cells only for one day, and the drug indirectly affected N-glycosylation; in

addition, dolichol in cells was not depleted. Therefore, we did not expect that N-glycosylation would be dramatically influenced over a short period of the treatment.

An example of the full MS and elution profiles of the heavy and light versions of glycopeptide WSFSN*GTSWR are shown in Figure 6.11a and 5.11b. Based on the areas under the curves from both elution profiles, we were able to accurately quantify the ratio of the glycopeptide as 2.17. This peptide is from protein NEU1 (sialidase-1), which catalyzes the removal of sialic acid moieties from glycoproteins and glycolipids. The protein abundance was up-regulated by 2.07 fold under the drug treatment.

We quantified a total of 1,247 unique glycopeptides from BA and 682 glycopeptides from MC with an overlap of 376 peptides (Figure 6.11c). As anticipated, majority of the quantified peptides were not regulated when two-fold change was used as a threshold. With the BA method, 59 glycopeptides were down-regulated while 93 were up-regulated before normalization. With the MC method, 24 glycopeptides were down-regulated while 62 were up-regulated.

6.2.3.5 Glycosylation site quantification and normalization by their corresponding parent protein abundance changes

For the N-glycosylation site quantification, besides the identification criteria discussed above, all glycopeptides must be singly glycosylated with the site ModScore > 13. With the site localization confidence, the quantitation can be site-specific. Although 1,045 glycosylation sites were identified from BA, only 719 sites (listed in a table online at doi.org/10.1016/j.ijms.2017.05.010) were quantified; while in MC, the combination of MC and SILAC led to the confident quantitation of 584 glycosylated sites (listed in a table online at doi.org/10.1016/j.ijms.2017.05.010) out of 685 unique glycosylation sites identified. 330 sites

were quantified in both experiments (Figure 6.11d). Relatively low overlap was expected because the principles of two enrichment methods are different. The powerful MS-based proteomics can allow us to site-specifically quantify protein glycosylation changes.

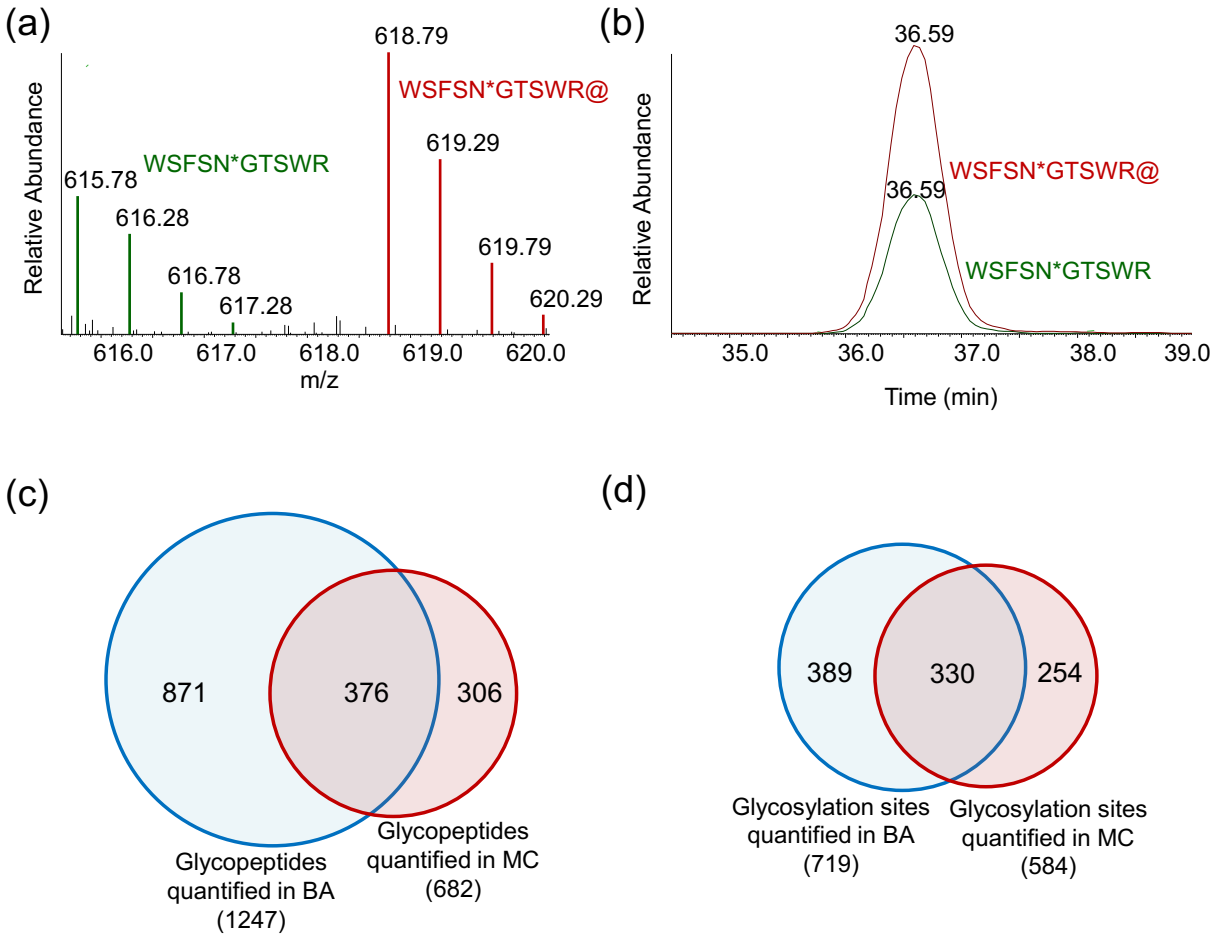


Figure 6.11 (a) An example of the full MS of the heavy (WSFSN*GTSWR@) and light (WSFSN*GTSWR) glycopeptides with the same sequence; (b) the elution profiles of the two glycopeptides; (c) comparison of unique glycopeptides quantified in the two experiments; (d) comparison of the glycosylation sites quantified from the two experiments.

When performing quantitative study of protein modifications, we need to pay attention to the abundance changes of their parent proteins. For instance, if a protein is dramatically up-regulated in treated cells while the stoichiometry of the modification sites from this protein are largely unaffected or even down-regulated, we could still profile these sites to be up-regulated because site down-regulation cannot cancel out the effect of protein up-regulation. As shown in Figure 6.12a, for example, if we assume that two out of three copies of a certain protein are N-glycosylated, then it will result in 66.7% glycosylation rate. After the drug treatment, although four copies of this protein are glycosylated, the glycosylation percentage is significantly lowered. This is due to protein expression up-regulation in the treated cells. Therefore, we normalized the raw site ratios by the corresponding parent protein ratios we obtained previously [117] to provide more quantitative information. This normalization strategy was previously applied for phosphorylation analysis in the literature [52].

The site ratio distributions in the BA and MC experiments before and after normalization are shown in Figure 6.12b and c. The whole series were shifted towards the down-regulation side after normalization. We listed a few quantified sites as examples in Table 6.3; two-fold was set as the threshold for defining a site to be regulated. All the listed sites have raw ratios larger than 2. However, their protein ratios are also larger than 2, which demonstrated that these proteins were up-regulated in the statin-treated cells. For instance, the first site in the list is from protein HMGCR, the rate-liming enzyme in the mevalonate pathway and the direct target of statins. Since the function of this protein was inhibited by the statin, this protein expression was up-regulated dramatically in the statin-treated cells, and the protein ratio for HMGCR increased by 15.4 fold. Without normalization by the protein ratio, N-glycosylation site 281 on HMGCR was up-regulated

by 25.9 fold, while the normalized ratio was 1.7. After normalization, six out of seven N-glycosylation sites were determined to be not regulated.

Among 640 normalized sites from the BA experiment (listed in a table online at doi.org/10.1016/j.ijms.2017.05.010), 22 were up-regulated, and 35 were down-regulated. In contrast, 30 sites were up-regulated, and 50 were down-regulated among 518 normalized sites from the MC experiment. Although we quantified fewer sites from the MC experiment, more sites were down-regulated. This phenomenon is in very good agreement with the expectation that the results from the MC experiment may be more dynamic because it only enriches the newly-synthesized glycoproteins during the statin treatment.

Table 6.3 Some example N-glycosylation sites quantified in the BA experiment.

Gene Symbol	PPM	XCorr	Peptide	Site	Mod Score	Site ratio	Protein ratio	Site ratio normalized
HMGCR	-0.43	4.24	WIADPSPQN*STA DTSK#	281	1000	25.89	15.35	1.69
KLB	-0.28	2.17	FALDWASVLPTGN *LSAVNR@	611	31.94	14.24	2.02	7.05
SLCO4C1	-1.06	1.93	VYYN*CSCIER	544	1000	5.39	4.18	1.29
KLB	-1.18	2.32	MGQN*VSLNLR	391	59.30	2.88	2.02	1.43
A1BG	0.50	3.50	EGDHEFLEVPEAQ EDVEATFPVHQPG N*YSCSYR@	179	1000	2.75	3.05	0.90
NEU1	0.06	1.43	VN*LTLR@	343	1000	2.47	2.07	1.19
	0.30	1.77	WSFSN*GTSWRK	352	1000	2.17	2.07	1.05

*-glycosylation site, #-heavy lysine, @-heavy arginine

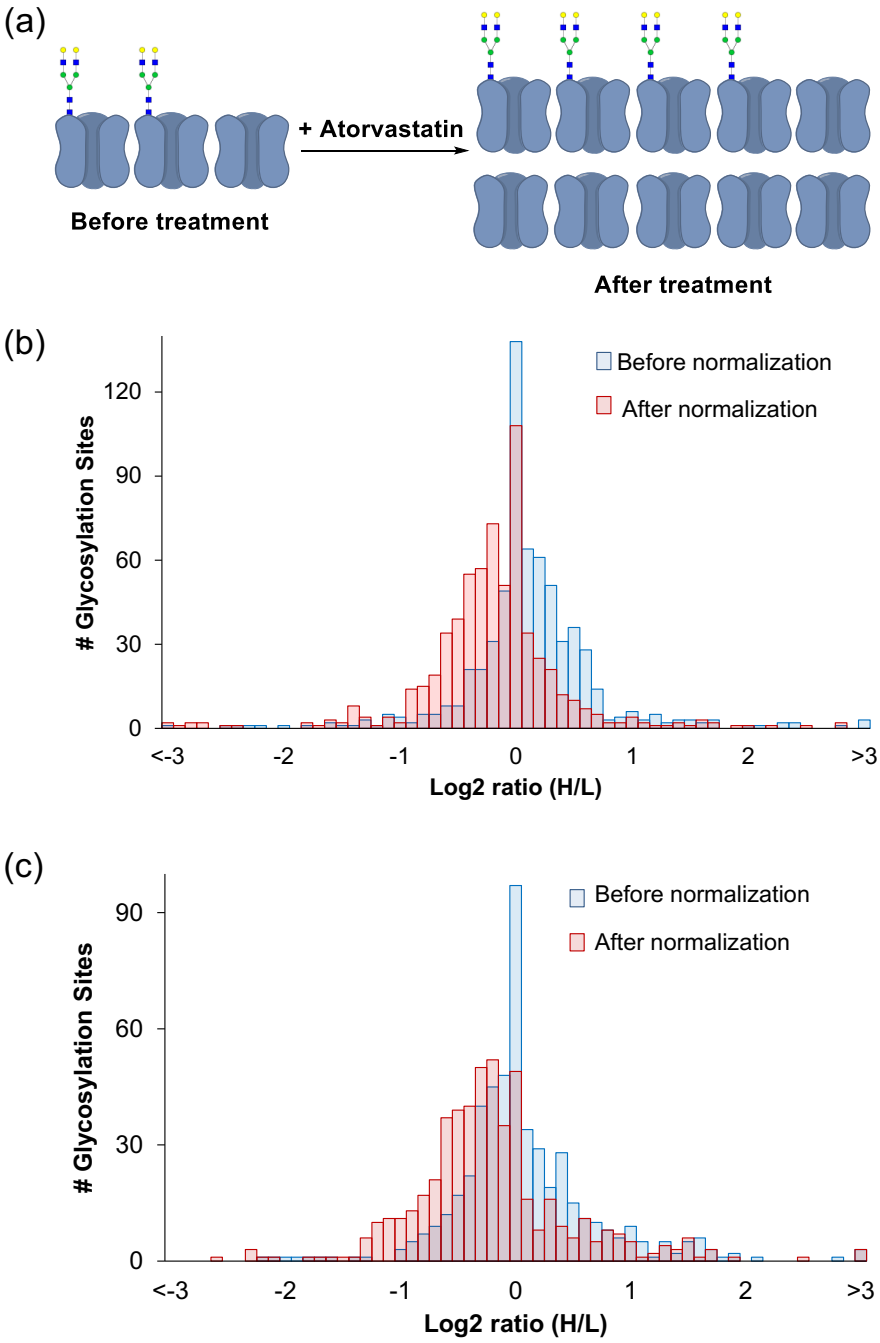


Figure 6.12 (a) An illustration of glycosylation site and glycoprotein abundance changes; glycosylation site regulation distributions before and after normalization using corresponding protein ratios in the (b) BA and (c) MC experiments.

6.2.4. Conclusions

Protein glycosylation alteration is often a hallmark of human disease. In-depth analysis of glycoprotein changes may aid in a better understanding of glycoprotein functions and lead to the discovery of disease biomarkers and drug targets. Modern MS-based proteomics is very powerful in globally analyzing protein modifications, but it is pivotal to enrich modified proteins in complex biological samples prior to MS analysis. Lectin-based enrichment methods have been used extensively to enrich glycopeptides. However, the binding specificity of lectin prevents high coverage of glycopeptides. Here, we evaluated two lectin-independent chemical enrichment methods (namely, BA and MC) for global analysis of protein N-glycosylation. BA is based on the reversible interactions between boronic acids and hydroxyl groups on glycans; MC utilizes the endogenous glycan synthesis pathways in human cells to incorporate a sugar analog with a chemically functional, but biologically inert group, into the glycan structure, followed by biorthogonal reactions and affinity enrichment. BA is more universal and helped identify a greater number of glycosylation sites, whereas MC has better performance on cell surface glycoprotein identification. Furthermore, the quantitative results from the MC experiment were more dynamic because it enriched the newly synthesized glycoproteins under the drug treatment. For the quantification of protein modification, normalization using the parent protein ratios can provide more quantitative information regarding the protein expression and modification changes. Because of the high abundance of proteins and sugars in human cells, the interactions between proteins and sugars are ubiquitous. Global analysis of protein glycosylation will dramatically facilitate glycoscience research in the biological and biomedical fields.

6.3 References

1. Murray, C.J.L. et al. The state of US health, 1990-2010 burden of diseases, injuries, and risk factors. *JAMA-J. Am. Med. Assoc.* **310**, 591-608 (2013).
2. Ton, V.K., Martin, S.S., Blumenthal, R.S. & Blaha, M.J. Comparing the new european cardiovascular disease prevention guideline with prior american heart association guidelines: an editorial review. *Clin. Cardiol.* **36**, E1-E6 (2013).
3. Baigent, C. et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* **366**, 1267-1278 (2005).
4. Stone, N.J. et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults a report of the american college of cardiology/American heart association task force on practice guidelines. *J. Am. Coll. Cardiol.* **63**, 2889-2934 (2014).
5. http://www.mercurynews.com/nation-world/ci_24507796/u-s-calls-one-third-all-adults-take.
6. "Doing things differently", Pfizer 2008 Annual Review,. (2009).
7. Liao, J.K. & Laufs, U. in Annual Review of Pharmacology and Toxicology, Vol. 45 89-118 (Annual Reviews, Palo Alto; 2005).
8. Hamelin, B.A. & Turgeon, J. Hydrophilicity/lipophilicity: relevance for the pharmacology and clinical effects of HMG-CoA reductase inhibitors. *Trends Pharmacol. Sci.* **19**, 26-37 (1998).
9. Ernster, L. & Dallner, G. Biochemical, physiological and medical aspects of ubiquinone function. *Biochim. Biophys. Acta-Mol. Basis Dis.* **1271**, 195-204 (1995).
10. Dutton, P.L. et al. 4 Coenzyme Q oxidation reduction reactions in mitochondrial electron transport. (CRC Press, Boca Raton; 2000).
11. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta-Gen. Subj.* **1426**, 239-257 (1999).
12. Farh, L., Mitchell, D.A. & Deschenes, R.J. Farnesylation and proteolysis are sequential, but distinct steps in the CAAX modification pathway. *Arch. Biochem. Biophys.* **318**, 113-121 (1995).
13. Resh, M.D. Trafficking and signaling by fatty-acylated and prenylated proteins. *Nat. Chem. Biol.* **2**, 584-590 (2006).
14. Resh, M.D. Covalent lipid modifications of proteins. *Curr. Biol.* **23**, R431-R435 (2013).
15. Tsunekawa, T. et al. Cerivastatin, a hydroxymethylglutaryl coenzyme A reductase inhibitor, improves endothelial function in elderly diabetic patients within 3 days. *Circulation* **104**, 376-379 (2001).
16. Jain, M.K. & Ridker, P.M. Anti-inflammatory effects of statins: Clinical evidence and basic mechanisms. *Nat. Rev. Drug Discov.* **4**, 977-987 (2005).
17. Lahera, V. et al. Endothelial dysfunction, oxidative stress and inflammation in atherosclerosis: Beneficial effects of statins. *Curr. Med. Chem.* **14**, 243-248 (2007).

18. Zhou, Q. & Liao, J.K. Pleiotropic effects of statins - basic research and clinical perspectives. *Circ. J.* **74**, 818-826 (2010).

19. Jasinska, M., Owczarek, J. & Orszulak-Michalak, D. Statins: a new insight into their mechanisms of action and consequent pleiotropic effects. *Pharmacol. Rep.* **59**, 483-499 (2007).

20. Weis, M., Heeschen, C., Glassford, A.J. & Cooke, J.P. Statins have biphasic effects on angiogenesis. *Circulation* **105**, 739-745 (2002).

21. Sacco, R.L. & Liao, J.K. Drug insight: statins and stroke. *Nat. Clin. Pract. Cardiovasc. Med.* **2**, 576-584 (2005).

22. Liao, J.K. Clinical implications for statin pleiotropy. *Curr. Opin. Lipidology* **16**, 624-629 (2005).

23. Kobashigawa, J.A. et al. Effect of pravastatin on outcomes after cardiac transplantation. *N. Engl. J. Med.* **333**, 621-627 (1995).

24. Cafforio, P., Dammacco, F., Gernone, A. & Silvestris, F. Statins activate the mitochondrial pathway of apoptosis in human lymphoblasts and myeloma cells. *Carcinogenesis* **26**, 883-891 (2005).

25. Graaf, M.R., Richel, D.J., van Noorden, C.J.F. & Guchelaar, H.J. Effects of statins and farnesyltransferase inhibitors on the development and progression of cancer. *Cancer Treat. Rev.* **30**, 609-641 (2004).

26. Campbell, M.J. et al. Breast cancer growth prevention by statins. *Cancer Res.* **66**, 8707-8714 (2006).

27. Klawitter, J., Shokati, T., Moll, V., Christians, U. & Klawitter, J. Effects of lovastatin on breast cancer cells: a proteo-metabonomic study. *Breast Cancer Res.* **12** (2010).

28. Singh, S., Singh, A.G., Singh, P.P., Murad, M.H. & Iyer, P.G. Statins are associated with reduced risk of esophageal cancer, particularly in patients with barrett's esophagus: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **11**, 620-629 (2013).

29. Demierre, M.F., Higgins, P.D.R., Gruber, S.B., Hawk, E. & Lippman, S.M. Statins and cancer prevention. *Nat. Rev. Cancer* **5**, 930-942 (2005).

30. Wang, P.S., Solomon, D.H., Mogun, H. & Avorn, J. HMG-CoA reductase inhibitors and the risk of hip fractures in elderly patients. *JAMA-J. Am. Med. Assoc.* **283**, 3211-3216 (2000).

31. Pandey, R.D., Gupta, P.P., Jha, D. & Kumar, S. Role of statins in Alzheimer's disease: a retrospective meta-analysis for commonly investigated clinical parameters in RCTs. *Int. J. Neurosci.* **123**, 521-525 (2013).

32. Silva, T., Teixeira, J., Remiao, F. & Borges, F. Alzheimer's disease, cholesterol, and statins: the junctions of important metabolic pathways. *Angew. Chem.-Int. Edit.* **52**, 1110-1121 (2013).

33. Kurata, T. et al. Statins have therapeutic potential for the treatment of Alzheimer's disease, likely via protection of the neurovascular unit in the AD brain. *J. Neurol. Sci.* **322**, 59-63 (2012).

34. Ceriello, A. et al. Effect of atorvastatin and irbesartan, alone and in combination, on postprandial endothelial dysfunction, oxidative stress, and inflammation in type 2 diabetic patients. *Circulation* **111**, 2518-2524 (2005).

35. Matafome, P. et al. Metformin and atorvastatin combination further protect the liver in type 2 diabetes with hyperlipidaemia. *Diabetes-Metab. Res. Rev.* **27**, 54-62 (2011).

36. Kearney, P.M. et al. Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis. *Lancet* **371**, 117-125 (2008).

37. Katz, D.H., Intwala, S.S. & Stone, N.J. Addressing statin adverse effects in the clinic: The 5 Ms. *J. Cardiovasc. Pharmacol. Ther.* **19**, 533-542 (2014).

38. Cannon, J. et al. High-throughput middle-down analysis using an orbitrap. *J. Proteome Res.* **9**, 3886-3890 (2010).

39. Yates, J.R. A century of mass spectrometry: from atoms to proteomes. *Nat. Methods* **8**, 633-637 (2011).

40. Hebert, A.S. et al. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat. Methods* **10**, 332-+ (2013).

41. Dai, L.Z. et al. Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10**, 365-U373 (2014).

42. Wu, R.H. et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* **8**, 677-683 (2011).

43. Munoz, J. & Heck, A.J.R. From the human genome to the human proteome. *Angew. Chem.-Int. Edit.* **53**, 10864-10866 (2014).

44. Choudhary, C., Weinert, B.T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 536-550 (2014).

45. Mann, M. & Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255-261 (2003).

46. Kim, J.E., Tannenbaum, S.R. & White, F.M. Global phosphoproteome of HT-29 human colon adenocarcinoma cells. *J. Proteome Res.* **4**, 1339-1346 (2005).

47. Albuquerque, C.P. et al. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* **7**, 1389-1396 (2008).

48. Zhou, W.D. et al. An initial characterization of the serum phosphoproteome. *J. Proteome Res.* **8**, 5523-5531 (2009).

49. 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).

50. Huttlin, E.L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189 (2010).

51. Pan, L., Iliuk, A., Yu, S., Geahlen, R.L. & Tao, W.A. Multiplexed quantitation of protein expression and phosphorylation based on functionalized soluble nanopolymers. *J. Am. Chem. Soc.* **134**, 18201-18204 (2012).

52. Wu, R.H. et al. Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. *Mol. Cell. Proteomics* **10**, 10.1074/mcp.M1111.009654 (2011).

53. Ham, B.M. et al. Novel Ser/Thr protein phosphatase 5 (PP5) regulated targets during DNA damage identified by proteomics analysis. *J. Proteome Res.* **9**, 945-953 (2010).

54. Melo-Braga, M.N. et al. Comprehensive quantitative comparison of the membrane proteome, phosphoproteome, and sialiome of human embryonic and neural stem cells. *Mol. Cell. Proteomics* **13**, 311-328 (2014).

55. Brioschi, M., Lento, S., Tremoli, E. & Banfi, C. Proteomic analysis of endothelial cell secretome: A means of studying the pleiotropic effects of HMG-CoA reductase inhibitors. *J. Proteomics* **78**, 346-361 (2013).

56. Gu, M.X. et al. Proteomic Analysis of endothelial lipid rafts reveals a novel role of statins in antioxidation. *J. Proteome Res.* **11**, 2365-2373 (2012).

57. Dong, X.L., Xiao, Y.S., Jiang, X.N. & Wang, Y.S. Quantitative proteomic analysis revealed lovastatin-induced perturbation of cellular pathways in HL-60 cells. *J. Proteome Res.* **10**, 5463-5471 (2011).

58. Kuntz, E. & Kuntz, H.-D. Hepatology: Textbook and Atlas, Edn. 3rd. (Springer, Heidelberg, Germany; 2008).

59. Villen, J. & Gygi, S.P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630-1638 (2008).

60. 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).

61. 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).

62. 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).

63. 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).

64. Du, X. et al. A computational strategy to analyze label-free temporal bottom-up proteomics data. *J. Proteome Res.* **7**, 2595-2604 (2008).

65. 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).

66. Zhang, J.Y. et al. Bayesian nonparametric model for the validation of peptide identification in shotgun proteomics. *Mol. Cell. Proteomics* **8**, 547-557 (2009).

67. 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).

68. Schwartz, D. & Gygi, S.P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391-1398 (2005).

69. 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).

70. Rudling, M. et al. Regulation of hepatic low-density lipoprotein receptor, 3-hydroxy-3-methylglutaryl coenzyme A reductase, and cholesterol 7 alpha-hydroxylase mRNAs in human liver. *J. Clin. Endocrinol. Metab.* **87**, 4307-4313 (2002).

71. Dong, B., Wu, M.H., Cao, A.Q., Li, H. & Liu, J.W. Suppression of Idol expression is an additional mechanism underlying statin-induced up-regulation of hepatic LDL receptor expression. *Int. J. Mol. Med.* **27**, 103-110 (2011).

72. Pocathikorn, A., Taylor, R.R. & Mamotte, C.D.S. Atorvastatin increases expression of low-density lipoprotein receptor mRNA in human circulating mononuclear cells. *Clin. Exp. Pharmacol. Physiol.* **37**, 471-476 (2010).

73. Wong, W.M.R. et al. The APOA4 T347S variant is associated with reduced plasma TAOS in subjects with diabetes mellitus and cardiovascular disease. *J Lipid Res* **45**, 1565-1571 (2004).

74. Delgado-Lista, J. et al. Effects of variations in the APOA1/C3/A4/A5 gene cluster on different parameters of postprandial lipid metabolism in healthy young men. *J Lipid Res* **51**, 63-73 (2010).

75. Nishikawa, K., Toker, A., Johannes, F.J., Zhou, S.Y. & Cantley, L.C. Determination of the specific substrate sequence motifs of protein kinase C isozymes. *J. Biol. Chem.* **272**, 952-960 (1997).

76. Yamasaki, T., Takahashi, A., Pan, J.Z., Yamaguchi, N. & Yokoyama, K.K. Phosphorylation of activation transcription factor-2 at serine 121 by protein kinase C controls c-Jun-mediated activation of transcription. *J. Biol. Chem.* **284**, 8567-8581 (2009).

77. Sillje, H.H.W. & Nigg, E.A. Identification of human Asf1 chromatin assembly factors as substrates of Tousled-like kinases. *Curr. Biol.* **11**, 1068-1073 (2001).

78. Groth, A. et al. Human Tousled like kinases are targeted by an ATM- and Chk1-dependent DNA damage checkpoint. *Embo J.* **22**, 1676-1687 (2003).

79. Lanz, R.B. et al. Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo. *Mol. Cell. Biol.* **23**, 7163-7176 (2003).

80. Ryan, M.C., Notterpek, L., Tobler, A.R., Liu, N. & Shooter, E.M. Role of the peripheral myelin protein 22 N-linked glycan in oligomer stability. *J. Neurochem.* **75**, 1465-1474 (2000).

81. Varki, A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126**, 841-845 (2006).

82. Konrad, R.J. & Kudlow, J.E. The role of O-linked protein glycosylation in beta-cell dysfunction (review). *Int. J. Mol. Med.* **10**, 535-539 (2002).

83. Spiro, R.G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43r-56r (2002).

84. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struc. Biol.* **21**, 576-582 (2011).

85. Zielinska, D.F., Gnad, F., Wisniewski, J.R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907 (2010).

86. Goto, M. Protein O-glycosylation in fungi: Diverse structures and multiple functions. *Biosci. Biotech. Bioch.* **71**, 1415-1427 (2007).

87. Kamemura, K. & Hart, G.W. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: A new paradigm for metabolic control of signal transcluction and transcription. *Prog. Nucleic. Acid Re.* **73**, 107-136 (2003).

88. Steentoft, C. et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977-982 (2011).

89. Breitling, J. & Aebi, M. N-linked protein glycosylation in the endoplasmic reticulum. *Csh. Perspect. Biol.* **5** (2013).

90. Ruiz-Canada, C., Kelleher, D.J. & Gilmore, R. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* **136**, 272-283 (2009).

91. Ibraghimovbeskrovnaya, O. et al. Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular-matrix. *Nature* **355**, 696-702 (1992).

92. Ohtsubo, K. & Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).

93. Roth, J. Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **102**, 285-303 (2002).

94. Ju, T.Z., Otto, V.I. & Cummings, R.D. The Tn antigen-structural simplicity and biological complexity. *Angew. Chem.-Int. Edit.* **50**, 1770-1791 (2011).

95. 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).

96. Harvey, D.J. Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert. Rev. Proteomic* **2**, 87-101 (2005).

97. Morelle, W., Canis, K., Chirat, F., Faid, V. & Michalski, J.C. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* **6**, 3993-4015 (2006).

98. Xiao, H.P. & Wu, R.H. Quantitative investigation of human cell surface N-glycoprotein dynamics. *Chemical Science* **8**, 268-277 (2017).

99. Xu, S.L., Medzihradszky, K.F., Wang, Z.Y., Burlingame, A.L. & Chalkley, R.J. N-glycopeptide profiling in *Arabidopsis Inflorescence*. *Mol. Cell. Proteomics* **15**, 2048-2054 (2016).

100. Yang, Y. et al. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 10 (2016).

101. Wang, X.S. et al. A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol. Cell. Proteomics* **15**, 2462-2475 (2016).

102. Yang, N. et al. Quantitation of site-specific glycosylation in manufactured recombinant monoclonal antibody drugs. *Anal. Chem.* **88**, 7091-7100 (2016).

103. Zacharias, L.G. et al. HILIC and ERLIC enrichment of glycopeptides derived from breast and brain cancer cells. *J. Proteome Res.* **15**, 3624-3634 (2016).

104. Khatri, K., Klein, J.A. & Zaia, J. Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Anal. Bioanal. Chem.* **409**, 607-618 (2017).

105. Zhu, Z.K. & Desaire, H. in Annual Review of Analytical Chemistry, Vol 8, Vol. 8. (eds. R.G. Cooks & J.E. Pemberton) 463-483 (Annual Reviews, Palo Alto; 2015).

106. Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* **12**, 561-567 (2015).

107. Sun, S.S. et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **34**, 84-88 (2016).

108. Chen, W.X., Smeekens, J.M. & Wu, R.H. A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol. Cell. Proteomics* **13**, 1563-1572 (2014).

109. Madera, M., Mechref, Y. & Novotny, M.V. Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal. Chem.* **77**, 4081-4090 (2005).

110. Dong, L.P., Feng, S., Li, S.S., Song, P.P. & Wang, J.D. Preparation of concanavalin A-chelating magnetic nanoparticles for selective enrichment of glycoproteins. *Anal. Chem.* **87**, 6849-6853 (2015).

111. Calvano, C.D., Zambonin, C.G. & Jensen, O.N. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J. Proteomics* **71**, 304-317 (2008).

112. 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).

113. Wollscheid, B. et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol.* **27**, 378-386 (2009).

114. Zeng, Y., Ramya, T.N.C., Dirksen, A., Dawson, P.E. & Paulson, J.C. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat. Methods* **6**, 207-209 (2009).

115. Xiao, H.P., Smeekens, J.M. & Wu, R.H. Quantification of tunicamycin-induced protein expression and N-glycosylation changes in yeast. *Analyst* **141**, 3737-3745 (2016).

116. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chemical Science* **6**, 4681-4689 (2015).

117. Xiao, H.P., Chen, W.X., Tang, G.X., Smeekens, J.M. & Wu, R.H. Systematic investigation of cellular response and pleiotropic effects in atorvastatin-treated liver cells by MS-based proteomics. *J. Proteome Res.* **14**, 1600-1611 (2015).

118. Xiao, H.P., Tang, G.X. & Wu, R.H. Site-specific quantification of surface N-glycoproteins in statin-treated liver cells. *Anal. Chem.* **88**, 3324-3332 (2016).

119. Hong, V., Steinmetz, N.F., Manchester, M. & Finn, M.G. Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconjugate Chem.* **21**, 1912-1916 (2010).

120. Shelbourne, M., Chen, X., Brown, T. & El-Sagheer, A.H. Fast copper-free click DNA ligation by the ring-strain promoted alkyne-azide cycloaddition reaction. *Chem. Commun.* **47**, 6257-6259 (2011).

121. Debets, M.F. et al. Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun.* **46**, 97-99 (2010).

122. Chen, W.X., Smeekens, J.M. & Wu, R.H. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chemical Science* **7**, 1393-1400 (2016).

123. Hang, H.C., Yu, C., Kato, D.L. & Bertozzi, C.R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14846-14851 (2003).

124. McKay, C.S. & Finn, M.G. Click chemistry in complex mixtures: Bioorthogonal bioconjugation. *Chem. Biol.* **21**, 1075-1101 (2014).

125. Baskin, J.M. et al. Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16793-16797 (2007).

126. Forbes, K. et al. Statins inhibit insulin-like growth factor action in first trimester placenta by altering insulin-like growth factor 1 receptor glycosylation. *Mol. Hum. Reprod.* **21**, 105-114 (2015).

APPENDIX

This appendix summarizes the completed collaboration projects that are not presented in the prior chapters. A brief description is provided for each project. In addition, the abstract of a book chapter resulted from the original research is also included.

A1. Simultaneous Time-Dependent Surface Enhanced Raman Spectroscopy, Metabolomics and Proteomics Reveal Cancer Cell Death Mechanisms Associated with Au-Nanorod Photo-Thermal Therapy

Ali, M. R. K., Wu, Y., Han, T. G., Zang, X. L., Xiao, H. P., Tang, Y., Wu, R. H., Fernandez, F. M., El-Sayed, M. A. Simultaneous time-dependent surface enhanced raman spectroscopy, metabolomics and proteomics reveal cancer cell death mechanisms associated with Au-nanorod photo-thermal therapy, *Journal of the American Chemical Society*, 2016, 138, 15434-15442.

In cancer plasmonic photothermal therapy (PPTT), plasmonic nanoparticles are used to convert light into localized heat, leading to cancer cell death. Among plasmonic nanoparticles, gold nanorods (AuNRs) with specific dimensions enabling them to absorb near-infrared laser light have been widely used. The detailed mechanism of PPTT therapy, however, still remains poorly understood. Typically, surface-enhanced Raman spectroscopy (SERS) has been used to detect time-dependent changes in the intensity of the vibration frequencies of molecules that appear or disappear during different cellular processes. A complete proven assignment of the molecular identity of these vibrations and their biological importance has not yet been accomplished. Mass

spectrometry (MS) is a powerful technique that is able to accurately identify molecules in chemical mixtures by observing their m/z values and fragmentation patterns. Here, we complemented the study of changes in SERS spectra with MS-based metabolomics and proteomics to identify the chemical species responsible for the observed changes in SERS band intensities during PPTT. We observed an increase in intensity of the bands at around 1000, 1207, and 1580 cm–1, which were assigned in the literature to phenylalanine, albeit with dispute. Our metabolomics results showed increased levels of phenylalanine, its derivatives, and phenylalanine-containing peptides, providing evidence for more confidence in the SERS peak assignments. To better understand the mechanism of phenylalanine increase upon PPTT, we combined metabolomics and proteomics results through network analysis, which proved that phenylalanine metabolism was perturbed. Furthermore, several apoptosis pathways were activated via key proteins (e.g., HADHA and ACAT1), consistent with the proposed role of altered phenylalanine metabolism in inducing apoptosis. Our study shows that the integration of the SERS with MS-based metabolomics and proteomics can assist the assignment of signals in SERS spectra and further characterize the related molecular mechanisms of the cellular processes involved in PPTT.

A2. Evaluation and Optimization of Reduction and Alkylation methods to Maximize Peptide Identification with MS-based Proteomics

Suttapitugsakul, S., Xiao, H. P., Smeekens, J. M., Wu, R. H. Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics, *Molecular BioSystems*, 2017, 13, 2574-2582.

Mass spectrometry (MS) has become an increasingly important technique to analyze

proteins. In popular bottom-up MS-based proteomics, reduction and alkylation are routine steps to facilitate peptide identification. However, incomplete reactions and side reactions may occur, which compromise the experimental results. In this work, we systematically evaluated the reduction step with commonly used reagents, i.e., dithiothreitol, 2-mercaptoethanol, tris(2-carboxyethyl)phosphine, or tris(3-hydroxypropyl)phosphine, and alkylation with iodoacetamide, acrylamide, N-ethylmaleimide, or 4-vinylpyridine. By using digested peptides from a yeast whole-cell lysate, the number of proteins and peptides identified were very similar using four different reducing reagents. The results from four alkylating reagents, however, were dramatically different with iodoacetamide giving the highest number of peptides with alkylated cysteine and the lowest number of peptides with incomplete cysteine alkylation and side reactions. Alkylation conditions with iodoacetamide were further optimized. To identify more peptides with cysteine, thiopropyl-sepharose 6B resins were used to enrich them, and the optimal conditions were employed for the reduction and alkylation. The enrichment resulted in over three times more cysteine-containing peptides than without enrichment. Systematic evaluation of the reduction and alkylation with different reagents can aid in a better design of bottom-up proteomic experiments.

A3. Global Analysis of Secreted Proteins and Glycoproteins in Saccharomyces Cerevisiae

Protein secretion is essential for numerous cellular activities, and secreted proteins in bodily fluids are a promising and noninvasive source of biomarkers for disease detection.

Systematic analysis of secreted proteins and glycoproteins will provide insight into protein function and cellular activities. Yeast (Saccharomyces cerevisiae) is an excellent model system for eukaryotic cells, but global analysis of secreted proteins and glycoproteins in yeast is challenging due to the low abundances of secreted proteins and contamination from high-abundance intracellular proteins. Here, by using mild separation of secreted proteins from cells, we comprehensively identified and quantified secreted proteins and glycoproteins through inhibition of glycosylation and mass spectrometry-based proteomics. In biological triplicate experiments, 245 secreted proteins were identified, and comparison with previous experimental and computational results demonstrated that many identified proteins were located in the extracellular space. Most quantified secreted proteins were down-regulated from cells treated with an N-glycosylation inhibitor (tunicamycin). The quantitative results strongly suggest that the secretion of these down-regulated proteins was regulated by glycosylation, while the secretion of proteins with minimal abundance changes was contrarily irrelevant to protein glycosylation, likely being secreted through nonclassical pathways. Glycoproteins in the yeast secretome were globally analyzed for the first time. A total of 27 proteins were quantified in at least two protein and glycosylation triplicate experiments, and all except one were down-regulated under N-glycosylation inhibition, which is solid experimental evidence to further demonstrate that the secretion of these proteins is regulated by their glycosylation. These results provide valuable insight into protein secretion, which will further advance protein secretion and disease studies.

A4. Evidence for the Importance of Post-Transcriptional Regulatory Changes in Ovarian Cancer Metastasis and the Contribution of miRNAs

Zhang, M. N., Matyunina, L. V., Walker, L. D., Chen, W. X., Xiao, H. P., Benigno, B. B., Wu, R. H., McDonald, J. F. Evidence for the importance of post-transcriptional regulatory changes in ovarian cancer metastasis and the contribution of miRNAs, *Scientific Reports*, 2017, 7:8171. Copyright 2017 Macmillan Publishers Limited, part of Springer Nature.

High-throughput technologies have identified significant changes in patterns of mRNA expression over cancer development but the functional significance of these changes often rests upon the assumption that observed changes in levels of mRNA accurately reflect changes in levels of their encoded proteins. We systematically compared the expression of 4436 genes on the RNA and protein levels between discrete tumor samples collected from the ovary and from the omentum of the same OC patient. The overall correlation between global changes in levels of mRNA and their encoding proteins is low ($r = 0.38$). The majority of differences are on the protein level with no corresponding change on the mRNA level. Indirect and direct evidence indicates that a significant fraction of the differences may be mediated by microRNAs.

A5. Specific Identification of Glycoproteins Bearing the Tn antigen in human cells

Reproduced with permission from Wiley-VCH Verlag GmbH &Co.

Zheng, J. N., Xiao, H. P., Wu, R. H. Specific identification of glycoproteins bearing the Tn antigen in human cells, *Angewandte Chemie International Edition*, 2017, 56, 7107-7111. Copyright 2017 Wiley-VCH Verlag GmbH &Co.

Glycoproteins contain a wealth of valuable information regarding the development and disease status of cells. In cancer cells, some glycans (such as the Tn antigen) are highly up-regulated, but this remains largely unknown for glycoproteins with a particular glycan. Herein, an innovative method combining enzymatic and chemical reactions was first designed to enrich

glycoproteins with the Tn antigen. Using synthetic glycopeptides with O-GalNAc (the Tn antigen) or O-GlcNAc, we demonstrated that the method is selective for glycopeptides with O-GalNAc and can distinguish between these two modifications. The diagnostic ions from the tagged O-GalNAc further confirmed the effectiveness of the method and confidence in the identification of glycopeptides with the Tn antigen by mass spectrometry. Using this method, we identified 96 glycoproteins with the Tn antigen in Jurkat cells. The method can be extensively applied in biological and biomedical research.

A6. Gold Nanorod-Assisted Plasmonic Photothermal Therapy of Cancer: Efficacy, Toxicity and Mechanistic Studies *in vivo*

Ali, M. R. K., Rahman, M. A., Wu, Y., Han, T. G., Peng, X. H., Mackay, M. A., Wang, D. S., Shin, H. J., Chen, Z., Xiao, H. P., Wu, R. H., Tang, Y., Shin, D. M., El-Sayed, M. A. Gold nanorod-assisted plasmonic photothermal therapy of cancer: efficacy, toxicity and mechanistic studies in vivo, *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114, E3110-E3118.

Gold nanorods (AuNRs)-assisted plasmonic photothermal therapy (AuNRs-PPTT) is a promising strategy for combating cancer in which AuNRs absorb near-infrared light and convert it into heat, causing cell death mainly by apoptosis and/or necrosis. Developing a valid PPTT that induces cancer cell apoptosis and avoids necrosis in vivo and exploring its molecular mechanism of action is of great importance. Furthermore, assessment of the long-term fate of the AuNRs after treatment is critical for clinical use. We first optimized the size, surface modification [rifampicin (RF) conjugation], and concentration (2.5 nM) of AuNRs and the PPTT laser power (2 W/cm2) to

achieve maximal induction of apoptosis. Second, we studied the potential mechanism of action of AuNRs-PPTT using quantitative proteomic analysis in mouse tumor tissues. Several death pathways were identified, mainly involving apoptosis and cell death by releasing neutrophil extracellular traps (NETs) (NETosis), which were more obvious upon PPTT using RF-conjugated AuNRs (AuNRs@RF) than with polyethylene glycol thiol-conjugated AuNRs. Cytochrome c and p53-related apoptosis mechanisms were identified as contributing to the enhanced effect of PPTT with AuNRs@RF. Furthermore, Pin1 and IL18-related signaling contributed to the observed perturbation of the NETosis pathway by PPTT with AuNRs@RF. Third, we report a 15-month toxicity study that showed no long-term toxicity of AuNRs in vivo. Together, these data demonstrate that our AuNRs-PPTT platform is effective and safe for cancer therapy in mouse models. These findings provide a strong framework for the translation of PPTT to the clinic.

A7. Targeting Cancer Cell Integrins Using Gold Nanorods in Photothermal Therapy Inhibits Migration through Affecting Cytoskeletal Proteins

Ali, M. R. K., Wu, Y., Tang, Y., Xiao, H. P., Chen, K. C., Han, T. G., Fang, N., Wu, R. H., El-Sayed, M. A. Targeting cancer cell integrins using gold nanorods in photothermal therapy inhibits migration through affecting cytoskeletal proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114, E5655-E5663. Copyright National Academy of Sciences (U.S.).

Metastasis is responsible for most cancer-related deaths, but the current clinical treatments are not effective. Recently, gold nanoparticles (AuNPs) were discovered to inhibit cancer cell migration and prevent metastasis. Rationally designed AuNPs could greatly benefit their

antimigration property, but the molecular mechanisms need to be explored. Cytoskeletons are cell structural proteins that closely relate to migration, and surface receptor integrins play critical roles in controlling the organization of cytoskeletons. Herein, we developed a strategy to inhibit cancer cell migration by targeting integrins, using Arg–Gly–Asp (RGD) peptide-functionalized gold nanorods. To enhance the effect, AuNRs were further activated with 808-nm near-infrared (NIR) light to generate heat for photothermal therapy (PPTT), where the temperature was adjusted not to affect the cell viability/proliferation. Our results demonstrate changes in cell morphology, observed as cytoskeleton protrusions-i.e., lamellipodia and filopodia-were reduced after treatment. The Western blot analysis indicates the downstream effectors of integrin were attracted toward the antimigration direction. Proteomics results indicated broad perturbations in four signaling pathways, Rho GTPases, actin, microtubule, and kinases-related pathways, which are the downstream regulators of integrins. Due to the dominant role of integrins in controlling cytoskeleton, focal adhesion, actomyosin contraction, and actin and microtubule assembly have been disrupted by targeting integrins. PPTT further enhanced the remodeling of cytoskeletal proteins and decreased migration. In summary, the ability of targeting AuNRs to cancer cell integrins and the introduction of PPTT stimulated broad regulation on the cytoskeleton, which provides the evidence for a potential medical application for controlling cancer metastasis.

A8. Exosomes Isolated from Bone Marrow-Derived MSCs Support the *ex vivo* Survival of Human Peripheral Blood-Derived Plasma Cells

Lewis, C. H., Nguyen, D., Garimalla, S., Xiao, H. P., Gibson, G., Wu, R. H., Galipeau, J., Lee, F. E. Exosomes isolated from bone marrow-derived MSCs support the ex vivo survival of human

peripheral blood-derived plasma cells. *Journal of Extracellular Vesicles*, accepted. Copyright retained by the authors.

Conditioned medium (CM) was from marrow-derived mesenchymal stromal cells (MSCs) was previously demonstrated to be able to maintains blood ASCs cell function for up to 30 days in vitro. We here purified MSC-derived exosomes from CM, and tested whether these vesicles could recapitulate the homeostatic interactions of the marrow niche. We treated MSC CM with a liposome-disrupting agent, abolishing plasma cell antibody-secretion by 75%. We further interrogated exosome production from replicating and irradiated, growth-arrested MSCs to better mirror the physiology of endogenous mobilized or quiescent marrow MSCs. We isolated Exosomes and the accompanying Exosome-Depleted CM (Exo-Depl CM), and assessed their ability to sustain antibody-secreting cells (ASCs) from healthy adult humans in vitro. Purified exosomes from both irradiated (Irrad) and Non-Irrad MSCs were comparable in their ability to support ASC functionality. However, Exo-Depl CM derived from Non-Irrad-MSCs was 50% less effective than corresponding fractions from Irrad-MSCs. Taken together, these findings indicate that MSC exosomes are an effective support system for the ex vivo culture of ASCs, and that growth arrested MSCs also produce additional products which act additively on ASCs. To identify which factors account for such differential support, we used proteomics and pathway analysis, identifying proteins involved in the survival of B-lineage cells and stroma cells, that were enriched in exosomes, including the ectoenzyme bone marrow stromal cell antigen-1 (CD157). Our data support the hypothesis that the in vitro support of ASC by MSC can be recapitulated with purified exosomes, suggesting a niche support mechanism which can operate independently of cell contact and MSC cell cycle status. These findings have great import for the exosome field and elucidation of factors that modulate B-cell biology.

A9. A Boronic Acid-Based Enrichment for Site-Specific Identification of the N-glycoproteome Using MS-Based Proteomics

Xiao H. P., Tang G.X., Chen W. X., Wu R. H. A Boronic Acid-Based Enrichment for Site-Specific Identification of the N-glycoproteome Using MS-Based Proteomics. In: Grant J., Li H. (eds) *Analysis of Post-Translational Modifications and Proteolysis in Neuroscience. Neuromethods*, 2015, vol 114. Humana Press, New York, NY. Copyright 2017 Springer International Publishing AG.

Modification of proteins by N-linked glycans plays a critically important role in biological systems, including determining protein folding and trafficking as well as regulating many biological processes. Aberrant glycosylation is well known to be related to disease, including cancer and neurodegenerative diseases. Current mass spectrometry (MS)-based proteomics provides the possibility for site-specific identification of the N-glycoproteome; however, this is extraordinarily challenging because of the low abundance of many N-glycoproteins and the heterogeneity of glycans. Effective enrichment is essential to comprehensively analyze N-glycoproteins in complex biological samples. The covalent interaction between boronic acid and cis-diols allows us to selectively capture glycopeptides and glycoproteins, whereas the reversible nature of the bond enables them to be released after non-glycopeptides are removed. By virtue of the universal boronic acid-diol recognition, large-scale mapping of N-glycoproteins can be achieved by combining boronic acid-based enrichment, PNGase F treatment in the presence of heavy oxygen (^{18}O) water, and MS analysis. This method can be extensively applied for the comprehensive analysis of N-glycoproteins in a wide variety of complex biological samples.

List of Publications

Journal Papers

1. **Xiao, H. P.#**; Chen, W. X.#; Smeekens, J. M.; Wu, R. H., A chemical method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins, ***Nature Communications*** (MS#: NCOMMS-17-11780), accepted.
2. Lewis, C. H.; Nguyen, D.; Garimalla, S.; **Xiao, H. P.**; Gibson, G.; Wu, R. H.; Galipeau, J.; Lee, F. E. Exosomes isolated from bone marrow-derived MSCs support the *ex vivo* survival of human peripheral blood-derived plasma cells. ***Journal of Extracellular Vesicles*****,** accepted.
3. **Xiao, H. P.**; Hwang, J. E.; & Wu, R. H, Mass spectrometric analysis of the N-glycoproteome in statin-treated liver cells with two lectin-independent chemical enrichment methods. ***International Journal of Mass Spectrometry***, DOI: 10.1016/j.ijms.2017.05.010, **2018**.
4. **Xiao, H. P.**; Wu, R. H., Quantitative investigation of human cell surface N-glycoprotein dynamics, ***Chemical Science***, 8(1):268-77, **2017**.
5. **Xiao, H. P.**; Wu, R. H., Global and site-specific analysis revealing unexpected and extensive protein S-GlcNAcylation in human cells, ***Analytical Chemistry***, 89 (6), 3656–3663, **2017**.
6. **Xiao, H. P.**; Wu, R. H., Simultaneous quantitation of glycoprotein degradation and synthesis rates by integrating isotope labelling, chemical enrichment and multiplexed proteomics, ***Analytical Chemistry***, 89 (19), 10361-10367, **2017**.
7. Ali, M. R. K.; Wu, Y.; Tang, Y.; **Xiao, H. P.**; Chen, K. C.; Han, T. G.; Fang, N.; Wu, R. H.; El-Sayed, M. A., Targeting cancer cell integrins using gold nanorods in photothermal therapy inhibits migration through affecting cytoskeletal proteins. ***Proceedings of the National Academy of Sciences***, 114(28), E5655-E5663, **2017**.
8. Ali, M. R. K.; Rahman, M. A.; Wu, Y.; Han, T. G.; Peng, X. H.; Mackay, M. A.; Wang, D. S.; Shin, H. J.; Chen, Z.; **Xiao, H. P.**; Wu, R. H.; Tang, Y.; Shin, D. M.; El-Sayed, M. A., Gold nanorod-assisted plasmonic photothermal therapy of cancer: efficacy, toxicity and mechanistic studies *in vivo*, ***Proceedings of the National Academy of Sciences***, 114(15), E3110-E3118, **2017**.
9. Zheng, J. N.; **Xiao, H. P.**; Wu, R. H., Specific identification of glycoproteins with the Tn antigen in human cells, ***Angewandte Chemie International Edition***, *56*(25), 7107-7111, **2017**.
10. Zhang, M. N.; Matyunina, L. V.; Walker, L. D.; Chen, W. X.; **Xiao, H. P.**; Benigno, B. B.; Wu, R. H.; McDonald, J. F., Evidence for the importance of post-transcriptional regulatory changes in ovarian cancer metastasis and the contribution of miRNAs, ***Scientific Reports***, 7:8171, **2017**.

11. Smeekens, J. M.; **Xiao, H. P.**; Wu, R. H., Global analysis of secreted proteins and glycoproteins in *Saccharomyces cerevisiae*, ***Journal of Proteome Research***, 16(2), 1039-1049, **2017**.
12. Suttapitugsakul, S.; **Xiao, H. P.**; Smeekens, J. M.; Wu, R. H., Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics, ***Molecular BioSystems***, 13, 2574-2582, **2017**.
13. **Xiao, H. P.**; Tang, G.X.; Wu, R. H., Site-Specific quantification of the N-glycoproteome on the cell surface of statin-treated liver cells, ***Analytical Chemistry***, 88(6): 3324-3332, **2016**.
14. **Xiao, H. P.**; Smeekens, J. M.; Wu, R. H., Quantification of tunicamycin-induced protein expression and N-glycosylation changes in yeast, ***Analyst***, 141, 3737-3745, **2016**.
15. Ali, M. R. K.; Wu, Y.; Han, T. G.; Zang, X. L.; **Xiao, H. P.**; Tang, Y.; Wu, R. H.; Fernandez, F. M.; El-Sayed, M. A., Simultaneous time-dependent surface enhanced raman spectroscopy, metabolomics and proteomics reveal cancer cell death mechanisms associated with Au-nanorod photo-thermal therapy, ***Journal of the American Chemical Society***, 138(47), 15434-15442, **2016**.
16. **Xiao, H. P.**; Chen, W. X.; Tang, G. X.; Smeekens, J. M.; Wu, R. H., Systematic investigation of cellular response and pleiotropic effects in atorvastatin-treated liver cells by MS-based proteomics. ***Journal of Proteome Research***, 14 (3), 1600-1611, **2015**.

Book Chapter

1. **Xiao, H. P.**; Tang, G. X.; Chen, W. X.; Wu, R. H., A boronic acid-based enrichment for site-specific identification of the N-glycoproteome using MS-based proteomics. ***Analysis of Post-Translational Modifications and Proteolysis in Neuroscience***, Springer, New York, Volume 114, Page 31-41, **2016**.