Advancing mass spectrometry methods for glycosylation analysis and their application to disease-related glyco-alteration study

By
Zhengwei Chen

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Department of Chemistry)

at the
UNIVERSITY OF WISCONSIN-MADISON
2018

Date of final oral examination: July 9th, 2018

The dissertation is examined by the following members of the Final Oral Committee:
Lingjun Li, Professor, School of Pharmacy and Department of Chemistry
Josh Coon, Professor, Department of Chemistry
John C. Wright, Professor, Department of Chemistry
Wei Xu, Professor, Department of Oncology

Acknowledgments

First of all, I would like to give my sincere thanks to my advisor Prof. Lingjun Li. We met at a mass spectrometry conference in Beijing in 2012 where I got the first impression on the cutting-edge techniques developed in her lab and really enjoyed her talk. At that time, I was in the process of applying for graduate school, and I finally decided to apply for UW Madison, particularly aiming to join her lab, after further communicating with her. It is not only because of the innovative research being performed in her lab, but more importantly, Dr. Li is such a wonderful mentor who really cares about her students' growth and always tries to provide the best resources to her students. During the 5 years' study in her lab, I have received so many good suggestions for research directions. Also, she is always so supportive, without which I would not survive my graduate studies. Outside the lab, Dr. Li is also such an easygoing, kind and caring person, and she is truly my role model in terms of this aspect as well. Under the leadership of Dr. Li, the Li lab has grown bigger and bigger, with more than 40 people, including current students and alumni in the Li Lab reunion during 2018 ASMS. I believe the Li lab will continue to grow and prosper with such a wonderful advisor!

I would also like to thank my committee members, Dr. Josh Coon, Dr. John Wright and Dr. Wei Xu, for their guidance and suggestions on the thesis background oral, research proposal, 4th year committee meeting, and, finally, this dissertation. Your guidance is vital for me to establish myself as a qualified chemist.

Furthermore, I would also like to acknowledge Dr. Ling Tong for her mentoring at the pharmaceutical analysis institute in the Tasly group in Tianjin, China. Under her guidance, I started

to work on developing mass spectrometry (MS)-based methods for metabolite analysis, and later developed an interest to pursue a graduate career centered on MS. Dr. Tong was the very first mentor on the MS training, I truly appreciate her guidance to help me start the MS journey.

Throughout my 5-year graduate career, I have been fortunate to have highly experienced mentors to teach and train me through various projects. Dr. Chenxi Jia introduced me to ion mobility (IM)-MS and taught me a lot of fundamental knowledge of this technique. Later, Dr. Christopher Lietz helped me further dig into IM-MS by working on the NPY project. I also learned a lot from him in terms of experiment design, data analysis and how to present data. Besides sharing the same interest in research, we also shared a lot of interests in other things such as philosophy, movies, and food. I was also very fortunate that I was mentored by Dr. Xuefei Zhong, who introduced me into the glycoscience world. She was very detail-oriented and urged me to dig into the textbooks and fully understand the concept before I started, which really helped me to lay a solid foundation of the principles and theories involved.

I'd also like to thank my co-workers for their great help on different projects. Dr. Qing Yu really taught me a lot in moving from glycomics towards glycoproteomics. Dr. Ling Hao helped me design the PKM2 project and gave a lot of good suggestions in this project and other projects. I also want to thank Jillian Johnson for teaching me cell culture and continuing to supply me with cells for method development. I also have many good discussions with Dr. Junfeng Huang, who provided valuable input in the AD project. I wanted to thank Dr. Matt Glover on the collaboration of writing the IM-MS review in glycan analysis. My undergraduates Richard Shipman and Eric Miesbauer were a great help to analyze numerous data. My thanks also go to Dr. Amanda

Buchberger's efforts in lab management, and language editing of numerous manuscripts. Thanks Xiaofang Zhong and Qinying Yu for taking over the AD projects and continuing to move this project forward. Additionally, I am also fortunate enough to have worked with Dr. Dustin Frost, Dr. Fabao Liu, Dr. Bingming Chen, Dr. Erin Gemperline, Dr. Shan Jiang, Yatao Shi, Zichuan Zhang, Kellen DeLaney, Qinjingwen Cao, and Xueqing Pang. They are all great people to work with and I benefited a lot from interacting with and talking to them.

I would also like to extend my thanks to my collaborators Dr. Cindy Carlsson and Dr. Ozioma Okonkwo for their valuable input in the AD project. And thanks Dr. Keqi Tang for the collaboration in the SPIN source project. Thanks Dr. Xinxiang Zhang for providing us with the robust glycan tags for quantitative glycomics study. I would also like to thank Prof. David Lubman and Prof. Wei Xu for their collaboration on the haptoglobin project and PKM2 project, respectively.

Furthermore, I would like to thank Dr. Cameron Scarlett and Molly Pellitteri Hahn of the Analytical Instrumentation Center for maintaining instruments in the School of Pharmacy. They are always willing to help when something went wrong with the instrument and willing to spend their time to work with us to fix the problem.

My final thanks go to my family and friends. My parents have been always with me providing numerous support since I was born. My girlfriend Xiaofan was there when I was struggling the most and I have been so fortunate to meet her in Madison. Without her support, I honestly could not have made it.

Table of Contents

Advancing mass spectrometry methods for glycosylation analysis and their application to disease-related glyco-alteration study

Zhengwei Chen

Under the supervision of Professor Lingjun Li

At the University of Wisconsin-Madison

Abstract

Mass spectrometry (MS) has evolved as a powerful tool for glycosylation analysis benefiting from its high speed, high resolution and sensitivity. It enables detailed structure characterization of glycans, glycosylation site and the peptide sequences of complex biological samples in a high-throughput. This dissertation is dedicated to developing MS-based analytical methods for the structure characterization and quantitative analysis of both released glycans and intact glycopeptides utilizing a bottom-up approach. We started by developing a hydrophilic interaction liquid chromatography (HILIC) coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging (MALDI-MSI) platform for improved quantitative analysis of N-glycans compared to MALDI-MS. Faced with the challenges of N-glycan structure complexity, capillary electrophoresis (CE) emerges as a robust separation technique affording very small amount of sample consumption, fast separation speed, and high

resolving power. Therefore, we developed a CE-ESI-MS/MS based 6-plex aminoxy tandem mass tag (TMT) labeling method for quantitative analysis of protein N-glycosylation analysis from human serum, providing the ability of N-glycan quantitation across up to six samples. Improved resolution and quantification accuracy of the labeled human milk oligosaccharides (HMOs) isomers was also achieved by coupling CE with traveling wave ion mobility (TWIM)-CID-MS/MS. The advantage of the glycan-focused approach (glycomics) is that the wealthy structure information and glycan isomer separation could be achieved; however, this approach resulted in loss of the glycosylation site-specific information. Therefore, we developed a powerful glycoproteomics workflow for system-wide structure characterization and quantitation of intact glycopeptides. Improvements of this workflow include improved glycoprotein extraction, sequential enrichment strategies for improved N-/O- glycopeptide enrichment, hybrid novel fragmentation technique electron transfer and higher-energy collision dissociation (EThcD), multiplex quantitation enabled by our custom-made isobaric N, N-dimethyl leucine (DiLeu) tags, and automated FDR-based large-scale data analysis by Byonic. We have successfully applied this workflow to analyze complex biological samples for disease-related glyco-alteration study, including the glycosylation alteration in PKM2 knockout breast cancer cells vs. parental cells and glycosylation pattern alteration in AD patients.

List of Abbreviations and Acronyms

ACN	Acetonitrile
AD	Alzheimer's disease
AGC	Automatic gain control
AminoxyTMT	Aminoxy tandem mass tag
APP	Amyloid precursor protein
Asn, N	Asparagine
BCA	Bicinchoninic acid assay
CID	Collision-induced dissociation
CCS	Collision cross section
CSF	Cerebrospinal fluid
CE	Capillary electrophoresis
CNS	Central nervous system
Da	Dalton, unit of mass measurement (1 Da = 1 g/mol)
DiLeu	N,N-dimethyl leucine
DDA	Data-dependent acquisition
DNA	Dexoyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetracetic acid

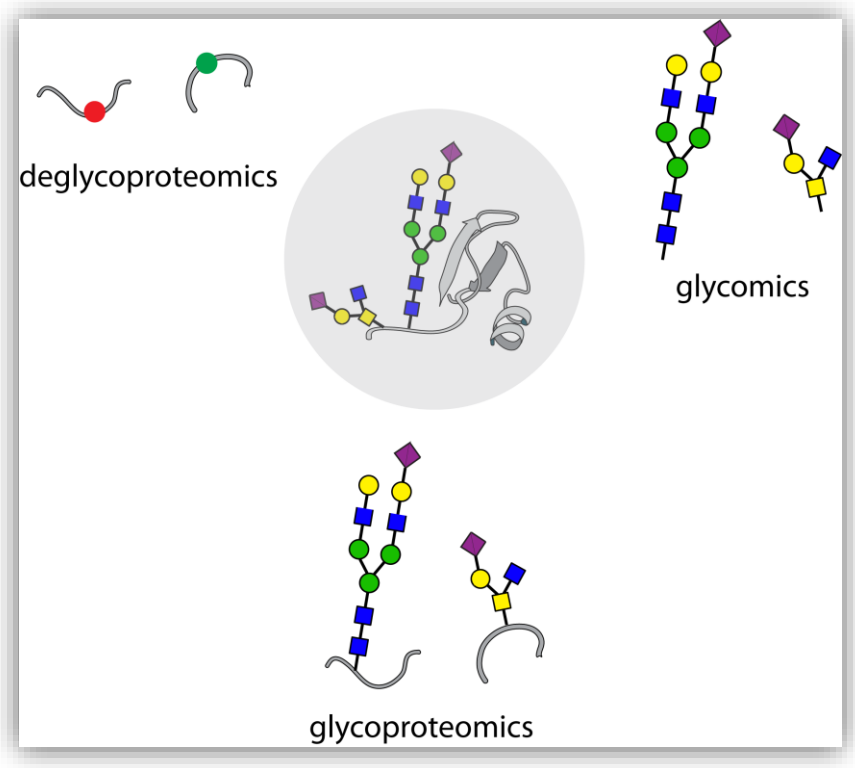
EIEs	Extracted ion electropherograms
ESI	Electrospray ionization
ETD	Electron transfer dissociation
EThcD	Electron-transfer/higher-energy collision dissociation
FA	Formic acid
FASP	Filter-aided sample preparation
FDR	False discovery rate
Fuc	Fucose
GO	Gene ontology
HCD	Higher-energy collisional dissociation
HILIC	Hydrophilic interaction chromatography
HMOs	Human milk oligosaccharides
HPLC	High-performance liquid chromatography
IM-MS	Ion mobility mass spectrometry
IT	Ion trap
LC	Liquid chromatography
m/z	mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
MCI	Mild cognitive impairment

MLA	Multi-lectin affinity
MS	Mass spectrometry
MS^1	Precursor mass analysis
MS^2	Tandem mass spectrometry
MSI	Mass spectrometric imaging
MWCO	Molecular weight cut-off
NCE	Normalized collision energy
ND	Neurodegenerative disease
PGC	Porous graphitic carbon
PNGase F	Peptide-N-glycosidase F
ppm	Parts per million
PSM	Peptide-spectrum match
PTM	Post-translational modification
RPLC	Reversed phase liquid chromatography
RT	Retention time
S/N	Signal-to-noise ratio
SAX	Strong anion exchange chromatography
SCX	Strong cation exchange chromatography
SDS	Sodium dodecyl sulfate

SPE	Solid phase extraction
S/T	Serine/threonine
TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
TOF	Time-of-flight
TMT	Tandem mass tag
TWIM	Traveling wave ion mobility
UA	Urea
XIC	Extracted ion chromatogram
z	charge
ZIC	zwitterionic ion chromatography

Chapter 1

Introduction and Research Summary



Mass spectrometry (MS) based bottom-up approaches for glycoprotein analysis

Introduction

In this dissertation, I have developed a multi-dimensional mass spectrometry (MS) platform, including HILIC-MALDI-MSI (hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging) and LC/CE/IM-ESI-MS (liquid chromatography/capillary electrophoresis/ion mobility-electrospray ionization-MS), to characterize both released glycans and intact glycopeptides at a system-wide level. Quantitation strategies at MS^1 and MS^2 level with isotopic and isobaric tags were also developed for quantitative analysis of glycans and glycopeptides across different disease/control samples. The developed analytical strategies helped to reveal glycosylation alteration in breast cancer and Alzheimer's disease (AD) and contributed to an improved understanding of the role of glycosylation in pathological processes.

Chapter 1 is a general introduction and summary of research reported in each chapter in the dissertation. **Chapter 2** reviews the recent advances of ion mobility-mass spectrometry for the improved structural characterization of glycans and glycoconjugates. **Chapter 3** reviews the recent advances in MS-based glycoproteomics in complex biological samples. **Chapter 4** describes my efforts to improve the current MALDI-MS platform by incorporating the HILIC separation and MALDI imaging for quantitative N-glycan analysis. **Chapter 5** reports on developing a CE-ESI-MS based strategy for quantitative analysis of N-glycans across up to six different samples enabled by aminoxyTMT tags. **Chapter 6** illustrates the development of an enhanced site-specific N-glycoproteomics workflow and its application to the study of PKM2 knockout breast cancer cells to better understand the role of glycosylation in breast cancer. **Chapter 7** describes the in-depth N-glycoproteome profiling in human cerebrospinal fluids (CSF) and the glycosylation landscape changes in Alzheimer's disease (AD). **Chapter 8** shows the results of CSF O-glycoproteome

profiling and the comprehensive post-translational modifications (PTMs) analysis of CSF endogenous peptides, as well as how the O-glycosylation and PTMs of endogenous peptides change in AD. **Chapter 9** concludes the dissertation and discusses future research directions. Additional information about the author's publications, presentations, background information and developed protocols can be found in **Appendices I, II and III.**

Chapter 2 Recent advances in ion mobility–mass spectrometry for improved structural characterization of glycans and glycoconjugates

Glycans and glycoconjugates are involved in regulating a vast array of cellular and molecular processes. Despite the importance of glycans in biology and disease, characterization of glycans remains difficult due to their structural complexity and diversity. Mass spectrometry (MS)-based techniques have emerged as the premier analytical tools for characterizing glycans. However, traditional MS-based strategies struggle to distinguish the large number of coexisting isomeric glycans that are indistinguishable by mass alone. Because of this, ion mobility spectrometry coupled to MS (IM–MS) has received considerable attention as an analytical tool for improving glycan characterization due to the capability of IM to resolve isomeric glycans before MS measurements. In this review, I present recent improvements in IM–MS instrumentation and methods for the structural characterization of isomeric glycans. In addition, we highlight recent applications of IM–MS that illustrate the enormous potential of this technology in a variety of research areas, including glycomics, glycoproteomics, and glycobiology.[1]

Chapter 3 Recent advances in MS-based glycoproteomics in complex biological samples

Protein glycosylation has been well-known to play a key role in various biological processes as well as in disease-related pathological progression. To better understand glycosylation's role in an underlying mechanism, comprehensive characterization and quantitation of glycoproteome in different biological systems are needed. MS-based glycoproteomics is a powerful approach that provides a system-wide profiling of the glycoproteome in a high-throughput manner. There have been significant technological advances in this field, including improved glycopeptide enrichment, hybrid fragmentation techniques, maturing specialized software, and effective quantitation strategies, as well as more dedicated workflows. With increasingly sophisticated glycoproteomics tools on hand, researchers have extensively employed this tool to explore different biology systems both in terms of in-depth glycoproteome profiling and comparative glycoproteome analysis. Quantitative glycoproteomics enables researchers to discover novel glycosylation-based biomarkers in various diseases with a potential to offer better sensitivity and specificity for disease diagnosis. In this review, we present the recent methodological developments in MS-based glycoproteomics and highlight its applications in answering different biological questions in complex biological systems.

Chapter 4 Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope

This chapter describes the development of a novel hydrophilic interaction liquid chromatography (HILIC) coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging (MALDI-MSI) platform for quantitative analysis of N-glycans. The relative quantitation was achieved by heavy and light stable-isotope labeled hydrazide reagents, and feasibility of the HILIC-MALDI-MSI platform for reliable quantitative analysis of N-glycans was demonstrated using the labeled maltodextrin ladder. The N-glycans released from human

serum was used to demonstrate the utility of this novel platform in quantitative analysis of N-glycans from a complex sample. Benefiting from the minimized ion suppression provided by HILIC separation, the comparison between MALDI-MS and the newly developed platform HILIC-MALDI-MSI revealed that HILIC-MALDI-MSI provided higher N-glycan coverage as well as better quantitation accuracy in the quantitative analysis of N-glycans released from human serum.[2]

Chapter 5 Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags

Recently developed carbonyl-reactive aminoxy tandem mass tag (aminoxyTMT) reagents enable multiplexed characterization and quantitative comparison of structurally complex glycans between different biological samples. Compared to some previously reported isotopic labeling strategies for glycans, the use of the aminoxyTMT method features a simple labeling procedure, excellent labeling efficiency, and reduced spectral complexity at the MS1 level. We developed an online capillary electrophoresis (CE)-ESI-MS/MS analyses of multiplexed aminoxyTMT-labeled human milk oligosaccharides (HMOs) and different types of N-glycans released from glycoprotein standards. Improved resolution and quantification accuracy of the labeled HMO isomers was achieved by coupling CE with traveling wave ion mobility (TWIM)-CID-MS/MS. N-Glycans released from human serum protein digests were labeled with six-plex aminoxyTMT and subjected to CE-ESI-MS/pseudo-MS3 analysis, which demonstrated the potential utility of this glycan relative quantification platform for more complex biological samples.[3]

Chapter 6 Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD)

This chapter describes the development of a dedicated workflow with the capability to simultaneously characterize and quantify intact glycopeptides in a site-specific and high-throughput manner. This enhanced workflow includes improved specific extraction of membrane-bound glycoproteins using the filter-aided sample preparation (FASP) method, enhanced enrichment of N-glycopeptides using sequential HILIC and multi-lectin affinity (MLA) enrichment, site-specific N-glycopeptide characterization enabled by EThcD, relative quantitation utilizing isobaric N,N-dimethyl leucine (DiLeu) tags and automated FDR-based large-scale data analysis by Byonic. When this approach was applied to study the glycosylation alterations in PKM2 knockout cells vs. parental breast cancer cells, it revealed altered N-glycoprotein/N-glycopeptide patterns and very different glycosylation microheterogeneity for different types of glycans. With glycosylation being one of the most important signaling modulators, our results provide additional evidence that signaling pathways are closely regulated by metabolism.[4]

Chapter 7 In-depth site-specific analysis of N-glycoproteome in human cerebrospinal fluid (CSF) and glycosylation landscape changes in Alzheimer's disease (AD)

In chapter 7, we developed a large-scale site-specific glycoproteomics approach for in-depth CSF N-glycoproteome analysis, including sequential HILIC and boronic acid enrichment for improved N-glycopeptide coverage, intact N-glycopeptide characterization enabled by EThcD and automated FDR-based large-scale data analysis by Byonic. This approach allows us to analyze thousands of intact N-glycopeptides from CSF proteins in a high-throughput manner, generating information about glycopeptide sequences, glycosylation site and glycan composition. In total, 3596 intact N-glycopeptides from 676 N-glycosylation sites and 358 N-glycoproteins were identified in CSF, representing the largest reported site-specific CSF N-glycoproteome dataset so far. This developed strategy was also applied to N-glycoproteome analysis of CSF samples from

AD patients, allowing us to conduct a glycosylation pattern comparison between healthy control and AD. A comparable N-glycoproteome coverage was detected in AD, but very different glycoform patterns were detected for glycoproteins such as alpha-1-antichymotrypsin, Ephrin-A3, Carnosinase CN1, voltage-dependent T-type calcium channel subunit alpha-1H etc., which serve as promising glycosylation-based biomarker candidates for AD.

Chapter 8 In-depth site-specific O-glycosylation analysis of glycoproteins and endogenous peptides in cerebrospinal fluid (CSF) from healthy individual, mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients

In this chapter, we first optimized the boronic acid-based enrichment strategy after PNGase F pretreatment to efficiently enrich both sialylated and non-sialylated O-glycopeptides in CSF. The optimized approach was applied to CSF O-glycoproteome study, where 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins were confidently identified. Up until now, the dataset represents the largest site-specific O-glycoproteome reported for CSF, including 154 novel O-glycosites. For endogenous peptide analysis, we developed a peptidomics workflow that combined CSF endogenous peptide extraction by 10k molecular weight cut-off (MWCO), EThcD fragmentation, and a three-step database searching strategy for comprehensive PTMs analysis. In this workflow, the use of EThcD enabled preserving of labile PTMs including glycosylation and phosphorylation, facilitating the accurate site localization. The three-step database searching strategy using Byonic as the search engine allowed a comprehensive PTM analysis of endogenous peptides, including both N-/O-glycosylation, phosphorylation, amidation, acetylation, Gln to pyro-Glu conversion, which allowed the discovery of 95 O-glycosylated CSF endogenous peptides for the first time. To provide a more complete picture of the O-glycosylation state of CSF glycoproteins in MCI and AD patients, in-depth O-glycosylation profiling

experiments were also conducted in parallel, showing a decreased fucosylation in MCI and AD. The PTM analysis of endogenous CSF peptides in MCI and AD patients showed an increased percentage of PTMs modified peptides, including O-glycosylation, Gln to pyro-Glu and acetylation.

Chapter 9 Conclusions and future directions

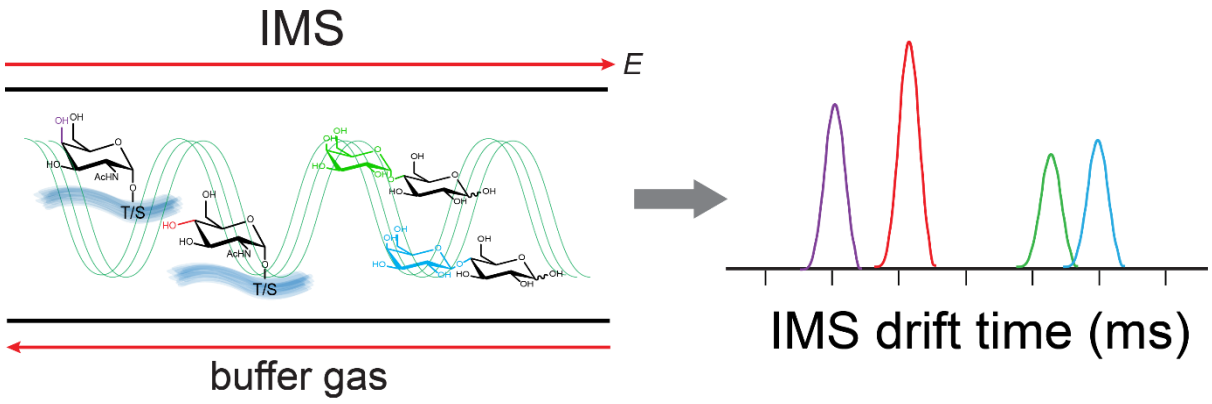
The final chapter presents the conclusions both in terms of the analytical method development and its application to disease studies. This chapter also includes the experiences gained during the method development process, and also the perspectives on the limitations of the current methodology and future directions.

References:

1. Chen, Z.; Glover, M. S.; Li, L., Recent advances in ion mobility–mass spectrometry for improved structural characterization of glycans and glycoconjugates. *Curr. Opin. Chem. Biol.* **2018,** 42, 1-8.
2. Chen, Z.; Zhong, X.; Tie, C.; Chen, B.; Zhang, X.; Li, L., Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents. *Anal. Bioanal. Chem.* **2017,** 409, (18), 4437-4447.
3. Zhong, X.; Chen, Z.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L., Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags. *Anal. Chem.* **2015,** 87, (13), 6527-6534.
4. Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L., Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD). *Analyst* **2018,** 143, (11), 2508-2519.

Chapter 2

Recent advances in ion mobility-mass spectrometry for improved structural characterization of glycans and glycoconjugates



- Adapted with permission from **Chen, Z.**,[#] Glover, M. S.,[#] & Li, L. (2018). Recent advances in ion mobility–mass spectrometry for improved structural characterization of glycans and glycoconjugates. Current opinion in chemical biology, 42, 1-8. (# These authors contributed equally). Copyright (2018) Elsevier.
- **Author contribution**: manuscript was written by Chen, Z., Glover, M. and edited by Chen, Z., Glover, M. and L. Li.

Abstract

Glycans and glycoconjugates are involved in regulating a vast array of cellular and molecular processes. Despite the importance of glycans in biology and disease, characterization of glycans remains difficult due to their structural complexity and diversity. Mass spectrometry (MS)-based techniques have emerged as the premier analytical tools for characterizing glycans. However, traditional MS-based strategies struggle to distinguish the large number of coexisting isomeric glycans that are indistinguishable by mass alone. Because of this, ion mobility spectrometry coupled to MS (IM-MS) has received considerable attention as an analytical tool for improving glycan characterization due to the capability of IM to resolve isomeric glycans prior to MS measurements. In this review, we present recent improvements in IM-MS instrumentation and methods for the structural characterization of isomeric glycans. In addition, we highlight recent applications of IM-MS that illustrate the enormous potential of this technology in a variety of research areas, including glycomics, glycoproteomics, and glycobiology.

Keywords: Ion mobility, Mass spectrometry, Glycans, Glycoconjugates, Isomers, Glycomics

Highlights:

- IM-MS is a powerful tool for glycan and glycoconjugate isomer analysis.
- Coupling IM with orthogonal separation techniques along with gas-phase manipulations greatly enhance isomer separation.
- Coupling IM with MS-based fragmentation and spectroscopic techniques improves structure characterization.
- CCS database provides another dimension of information for confident structural

identification.

- IM-MS contributes to improved characterization of complex biological samples.

Introduction

As one of the most abundant and complex protein post-translational modifications (PTMs), glycosylation is associated with many key biological processes including cell adhesion, molecular trafficking, receptor activation, and signal transduction [1]. The analysis of glycans and glycoconjugates is challenging due to the large diversity of structures resulting from the non-template driven biosynthesis [2, 3]. In addition, many of the monosaccharides that compose larger glycans are structural isomers, and they can be connected via either α- or β-stereochemistry at multiple linkage positions, resulting in many glycan isomers (**Figure 1a**). Isomeric glycans can also have a variety of connectivities to numerous sites on other classes of molecules such as proteins, contributing to their strucutral complexity (**Figure 1b**).

Separation and detailed structural characterization of glycan or glycoconjugate isomers is crucial for understanding their roles in various biological processes. Benefiting from speed and sensitivity of analysis, liquid chromatography (LC)-MS and capillary electrophoresis (CE)-MS have emerged as powerful techniques for glycan characterization [4-6]. Despite recent improvements to nearly all aspects of MS-based analyical workflows for glycan and glycoconjugate chatacterization, it remains challenging to achieve complete structural elucidation due to the complexity of glycans and lack of standard reference databases. Therefore, new techniques and methods that enhance the differentiation of glycan and glycoconjugate isomers would be highly desirable.

Although it has been two decades since IM-MS was orginally used to separate glycan isomers,

recent technological advancements have sparked increased interest in IM-MS for glycan and glycoconjugate analysis [7]. Unlike other commonly used separation techniques such as LC and CE, IM-MS is a post-ionization gas-phase technique that separates ions based on differences in shape and charge as they travel through a buffer gas under the influence of an electric field [8, 9]. The time it takes for an analyte ion to travel through the IM cell can be used to calculate rotationally averaged collision cross section (CCS) which provides an additional parameter that can be used to identify compounds as well as information about molecular conformation [8-10].

Initial applications of IM to carbohydrate analysis focused on distinguishing small isomeric carbohydrate standards [11, 12]. Due to the advancement and commercialization of IM-MS instrumentation, a growing number of labs continue to demonstrate that IM-MS is a fast, sensitive, and effective method for resolving carbohydrate isomers. For example, IM-MS has been used to separate a variety of isomeric species, including connectivity and configurational isomers [13]. Furthermore, studies have been extended to more complex systems such as N- and O-glycans and intact glycopeptides [14-20]. Here, we discuss the latest developments in IM-MS methods and technology that have allowed for enhanced sepration and strucural characterization of glycans and glycoconjugates and discuss advances necessary for IM to become more widely used in glycomics and glycoproteomics workflows.

Improving IM-Based Isomer Separations.

Although many proof-of-principle experiments have shown the potential of IM to separate glycan isomers, baseline separation of isomeric glycans is difficult as they often have minor differences in CCS. This is especially problematic as studies are expanded to larger glycans and glycoconjugates because minor changes in the glycan composition often result in subtle differences

in the overall structure. Furthermore, improved glycan separation will be crucial for extending the applications of IM-MS technology to large-scale studies of glycans and glycoproteins in complex mixtures from biological systems (i.e., glycomics and glycoproteomics). Thus, various analytical workflows that include IM separation have been developed to enhance the separation of isomeric glycans.

One of major factors that has limited the utility of IM-MS for many applications is that many instrument platforms lack the mobility resolution necessary to resolve isomeric species that have minor differences in CCS. The development of high resolution IM instrumentation using structures for lossless ion manipulations (SLIM) technology has shown great potential to enable separations of a variety of isomeric species [21]. Recently, a novel instrument was developed capable of ultralong pathlength travelling wave ion mobility (TWIM) separations on a serpentine-shaped SLIM device that has a 30-fold increase in IM resolution compared to traditional drift tube IM and TWIM instruments [22]. In addition to providing baseline separation of isomers lacto-N-hexaose and lacto-N-neohexaose, high resolution SLIM IM-MS revealed a new conformation of lacto-N-neohexaose. This suggests that SLIM-based IM separations will provide a level of conformational information about glycans that was previously inaccessible.

An alternative approach to increase isomer separation by improving IM-MS instrumentation is to optimize the charge state or polarity of the glycan ions to yield optimal separation of isomeric species. Numerous studies have demonstrated that mobility separations of glycan isomers can be optimized by manipulating the ion charge state or charge polarity [23, 24]. Because of this it is important to consider a variety of charge carriers, such as metal cations and anions, and ionization methods for improving mobility separations (**Figure 2a**) [25-29]. Furthermore, it was demonstrated that electron transfer reactions with group II metal-coordinated carbohydrates improve separation

of isomeric species, suggesting the potential for ion-ion reactions in the gas-phase for differentiation of isomeric oligosaccharides [30].

As the detection and the characterization of glycans is often hindered by the lack of a chromophore and their poor ionization properties in either spectroscopic or mass-spectrometric detection, a wide variety of glycan labeling reagents have been developed, which provide the opportunity to manipulate the conformation of the glycans in the gas-phase. Although 1-phenyl-3-methyl-5-pyrazolone (PMP) was originally developed to enhance UV detection, a recent study showed increased separation of structural isomers after PMP derivatization [31]. Reacting with *cis*-diols on carbohydrates, boronic acids (BAs) dereivatization has also been shown great potential in increasing isobaric carbohydrates differentiation as an ion mobility shift strategy [32]. Besides covalent binding, non-covalent binding such as crown ethers for peptides and metal cations for carbohydrates is another promising approach to enhance structural selectivity [24, 33, 34]. Recently, non-covalent complexes between monosaccharides and combinations of metal cations, peptides, and amino acids were demonstrated to improve differentiation of 8 pairs of enantiomeric glucose isomers [35]. Although there is no universal strategy to use sample preparation or gas-phase chemistry to improve glycan isomer separation, the strategies discussed above are important considerations for achieving optimal separation.

Coupling IM with orthogonal separation techniques.

In addition to improving IM-MS technology, it is important to consider the enhanced analytical capability by coupling IM-MS with a variety of orthogonal separation techniques. Because mobility separations typically occur on the order of milliseconds it is possible to couple IM between LC [36, 37] or CE [38] and mass spectrometry for glycan analysis (**Figure 2a**). The

combination of orthogonal separation methods has been demonstrated to offer improved characterization of isomeric glycans. For example, the combination of hydrophilic interaction chromatography (HILIC) and TWIM was used for separation of isomeric pectic oligosaccharides [37]. In addition, a straightforward reversed-phase LC-IM-MS platform was developed for integrated proteomic and glycomic studies [39]. Furthermore, the combination of CE with (TWIM)-CID-MS/MS provided improved separation and quantitation of aminoxy tandem mass tag (aminoxyTMT)-labeled human milk oligosaccharides (HMOs) (**Figure 2b**) [38].

Coupling IM with MS-based fragmentation and spectroscopic techniques.

The development of instrumentation that couples IM with a variety of fragmentation techniques has been crucial for improving the identification of glycan and glycopeptide isomers. It is important to note that coupling IM with MS/MS is mutually beneficial for both techniques. That is, ion mobility separations can be used to deconvolute complex fragmentation spectra that arise from coeluting isomeric species (**Figure 3a**). In addition, fragmentation spectra can be used to deconvolute CCS distributions of partially resolved isomeric species (**Figure 3a**).

An increasing number of studies have demonstrated that it is beneficial to perform ion mobility experiments on both precursor and fragment ions of isomeric species (**Figure 3b**). For example, IM analysis of precursor and fragmentation ions was used to differentiate Lewis and blood group epitopes [40]. Furthermore, it was recently demonstrated that IM analysis of MS/MS derived fragments of glycans can be used to differentiate anomeric glycosidic linkages [41]. Another promising approach capable of performing mobility separation of precursor and fragment ions is tandem IM (IM-CID-IM) in which two mobility regions are separated by a region where ions can be mobility selected and collisionally activated [42]. IM-IM-MS was recently used to distinguish

underivatized carbohydrate isomers based on differences in mobilities of fragments ions. By probing ion mobility profiles of product ions, IM was also used to distinguish α2,3 or α2,6 sialic-acid linkage [15, 43].

In addition to coupling IM with collision-based fragmentation methods, IM has recently been combined with a variety of fragmentation and spectroscopic techniques, including electron activated dissociation (ExD) [44], UV photodissociation [45, 46], and cryogenic IR spectroscopy [47, 48]. The combination of selected accumulation-trapped IM (SA-TIMS) with Fourier transform ion cyclotron resonance (FTICR) mass spectrometers makes it possible to perform ExD on mobility-selected ions [44]. It was recently demonstrated that coupling IM with cryogenic IR spectroscopy can be used to identify isomeric glycosaminoglycans that are partially resolved by IM [49].

Collision Cross Section Databases.

Another benefit IM-MS provides is the ability to measure CCS values which can be implemented into databases and used as additional criteria for structural identification (**Figure 3c**) [43, 50]. Because CCS values are an additional parameter to improve the identification of glycan and glycopeptides, several groups have compiled CCS databases of glycans and glycopeptides that have potential to aide in glycan identifications [25, 28, 51-53]. For example, GlycoMob is an online database of >900 CCS values of glycans and their fragments. In addition, a database containing glycopeptide CCS values was recently presented that has the potential to aide in identification of unique glycoforms [54].

Accurate glycan identification based on CCS values is still in its early stage with several challenges remaining before this approach could be used routinely for glycan identification. One of the challenges lies in the shortage of available CCS values resulting from the difficulty in

synthesizing glycan standards especially for these complex N- or O-glycans [55]. Another limitation of CCS databases is that CCS values are not intrinsic properties of ions in the same way as *m*/*z* values. CCS values depend on a variety of factors such as ionization conditions, buffer gas, instrument parameters, and the calibration method if not measured in a linear drift tube. Because of this, the development of robust standard operating procedures, quality controls, and calibration methods is necessary for CCS databases to be used effectively. Despite the challenges listed above, implementation of CCS databases into analytical workflows to improve glycan and glycoconjugate identification will likely be a crucial step for expanding the role and utility of IM-MS in the glycosciences. With increasing amounts of IM-MS data acquired each day, especially after coupling with LC, powerful bioinformatics tools that enable easy data acquisition, analysis and processing in a high-throughput and rapid way is key in advancing the IM-MS enhanced glycomics workflow development. Continuous efforts into platform development to support the storage of IM-MS data is also highly needed to facilitate the database query for the glycoscience community. Such platform should have the capability to support the complexity of IM-MS data, including the precursor ion and fragments m/z information, CCS information and LC information such as PGC-LC retention time reference [56], and also their connection to each other. Besides, the advancement of instrument with higher mobility resolution and CCS measurement accuracy will certainly largely improve the effectiveness and accuracy of IM-MS assisted glycomics workflow.

Improved characterization of intact glycoconjugates.

Although most examples mentioned above illustrate the utility of IM-MS for separating free or released glycans, an emerging application of IM-MS is the structural characterization of glycans bound to other classes of molecules such as proteins, peptides, and lipids (i.e., glycoconjugates).

The analysis of IM-MS can reveal information about the macro- and micro- heterogeneity of glycosylation of glycoconjugates. For example, IM can separate intact glycopeptides that differ only in the glycosylation sites [16]. Furthermore, IM-MS analysis of glycopeptide fragments was demonstrated to be an effective strategy to distinguish α2,3 versus α2,6 sialic acid linkages on intact glycopeptides [15, 16]. In addition to glycopeptides, IM-MS was recently used to separate glycolipid isomers [57]. It is also important to note that as IM-MS technologies evolve, they have the potential to be extended to characterize glycosylation of larger systems such as intact glycoproteins, antibodies, and virus capsids. For example, IM-MS and collision induced unfolding was used to provide qualitative information about glycosylation levels for intact antibodies [58].

Improved characterization of biological samples.

One of the most promising applications of IM-MS is the characterization of glycans and glycoconjugates in complex mixtures from biological systems. Several studies have shown the benefit of adding IM into MS-based workflows for the analysis of biological samples [20, 59]. For example, incorporation of field asymmetric ion mobility spectrometry (FAIMS) into a bottom-up proteomic workflow increased the number of glycopeptides identified from flagellin from *Campylobacter jejuni* 11168 [20]. It is important to note that IM separation of isomeric and isobaric species prior to MS analysis has been shown to improve quantification accuracy of both peptides and glycans [38, 60]. In addition to separating isomers, IM can also be used to extract glycan and glycopeptide regions from interference ions or other molecular species [39, 61-63]. Being able to separate various molecular species, IM has the potential to aide multi-omics studies, which could greatly simplify sample preparation procedures and provide more detailed information about molecular and cellular processes [64, 65]. Due to the recent improvements in technology and methods

described here, the application of IM-MS will likely continue to be extended beyond proof-of-principle experiments performed on glycan standards to complex mixtures from a variety of biological samples.

Conclusions.

IM-MS continues to emerge as a powerful method for characterizing the enormous structural diversity and complexity of glycans and glycoconjugates. Recent advances in IM-MS instrumentation and methods have positioned this technique to enable discoveries in the glycosciences. In the future, the development of hybrid methods that couples high-resolution IM with orthogonal separation techniques and MS-based fragmentation and spectroscopic techniques will likely provide the analytical capability to extend the application of IM-MS to more complex biological systems in order to unravel the role of glycans and glycoconjugates in biology and disease.

Acknowledgements

This research was supported in part by the National Institutes of Health (NIH) grants R21AG055377, R01 DK071801, and National Science Foundation (NSF) grant CHE-1710140. This reserach was supported by the National Institutes of Health, under Ruth L. Kirschstein National Research Service Award T32 HL 007936 from the National Heart Lung and Blood Institute to the University of Wisconsin-Madison Cardiovascular Research Center (MSG Postdoctoral Fellowship). LL acknowledges a Vilas Distinguished Achievement Professorship and Janis Apinis Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

1. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.
2. Rakus, J. F.; Mahal, L. K., New technologies for glycomic analysis: toward a systematic understanding of the glycome. *Annual review of analytical chemistry* **2011,** 4, 367-392.
3. Zhu, M.; Bendiak, B.; Clowers, B.; Hill, H. H., Ion mobility-mass spectrometry analysis of isomeric carbohydrate precursor ions. *Anal. Bioanal. Chem.* **2009,** 394, (7), 1853-1867.
4. Pabst, M.; Altmann, F., Glycan analysis by modern instrumental methods. *Proteomics* **2011,** 11, (4), 631-643.
5. Mariño, K.; Bones, J.; Kattla, J. J.; Rudd, P. M., A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **2010,** 6, (10), 713-723.
6. Zaia, J., Mass spectrometry and the emerging field of glycomics. *Chem. Biol.* **2008,** 15, (9), 881-892.
7. Gray, C.; Thomas, B.; Upton, R.; Migas, L.; Eyers, C.; Barran, P.; Flitsch, S., Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2016,** 1860, (8), 1688-1709.
8. Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E., Biomolecule analysis by ion mobility spectrometry. *Annu. Rev. Anal. Chem.* **2008,** 1, 293-327.
9. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Ion mobility–mass spectrometry. *J. Mass Spectrom.* **2008,** 43, (1), 1-22.
10. Jurneczko, E.; Barran, P. E., How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **2011,** 136, (1), 20-28.
11. Lee, S.; Wyttenbach, T.; Bowers, M. T., Gas phase structures of sodiated oligosaccharides by ion mobility/ion chromatography methods. *IJMSI* **1997,** 167, 605-614.
12. Liu, Y.; Clemmer, D. E., Characterizing oligosaccharides using injected-ion mobility/mass spectrometry. *Anal. Chem.* **1997,** 69, (13), 2504-2509.
13. Hofmann, J.; Hahm, H.; Seeberger, P.; Pagel, K., Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* **2015,** 526, (7572), 241-244.
14. Harvey, D. J.; Scarff, C. A.; Edgeworth, M.; Struwe, W. B.; Pagel, K.; Thalassinos, K.; Crispin, M.; Scrivens, J., Travelling‐wave ion mobility and negative ion fragmentation of high‐mannose N‐glycans. *J. Mass Spectrom.* **2016,** 51, (3), 219-235.
15. Guttman, M.; Lee, K. K., Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Anal. Chem.* **2016,** 88, (10), 5212-5217.
16. Hinneburg, H.; Hofmann, J.; Struwe, W. B.; Thader, A.; Altmann, F.; Silva, D. V.; Seeberger, P. H.; Pagel, K.; Kolarich, D., Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *ChCom* **2016,** 52, (23), 4381-4384.
17. Creese, A. J.; Cooper, H. J., Separation and identification of isomeric glycopeptides by high field asymmetric waveform ion mobility spectrometry. *Anal. Chem.* **2012,** 84, (5), 2597-2601.
18. Campbell, J. L.; Baba, T.; Liu, C.; Lane, C. S.; Le Blanc, J. Y.; Hager, J. W., Analyzing Glycopeptide Isomers by Combining Differential Mobility Spectrometry with Electron-and Collision-Based Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, 1-8.
19. Zhu, F.; Trinidad, J. C.; Clemmer, D. E., Glycopeptide site heterogeneity and structural diversity determined by combined lectin affinity chromatography/IMS/CID/MS techniques. *J. Am. Soc. Mass Spectrom.* **2015,** 26, (7), 1092-1102.

20. Ulasi, G. N.; Creese, A. J.; Hui, S. X.; Penn, C. W.; Cooper, H. J., Comprehensive mapping of O‐glycosylation in flagellin from Campylobacter jejuni 11168: A multienzyme differential ion mobility mass spectrometry approach. *Proteomics* **2015,** 15, (16), 2733-2745.
21. Ibrahim, Y. M.; Hamid, A. M.; Deng, L.; Garimella, S. V.; Webb, I. K.; Baker, E. S.; Smith, R. D., New frontiers for mass spectrometry based upon structures for lossless ion manipulations. *Analyst* **2017,** 142, (7), 1010-1021.
22. Deng, L.; Webb, I. K.; Garimella, S. V.; Hamid, A. M.; Zheng, X.; Norheim, R. V.; Prost, S. A.; Anderson, G. A.; Sandoval, J. A.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D., Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS using Structures for Lossless Ion Manipulations. *Anal. Chem.* **2017,** 89, (8), 4628-4634.
23. Struwe, W.; Baldauf, C.; Hofmann, J.; Rudd, P.; Pagel, K., Ion mobility separation of deprotonated oligosaccharide isomers–evidence for gas-phase charge migration. *ChCom* **2016,** 52, (83), 12353-12356.
24. Zheng, X.; Zhang, X.; Schocker, N. S.; Renslow, R. S.; Orton, D. J.; Khamsi, J.; Ashmus, R. A.; Almeida, I. C.; Tang, K.; Costello, C. E.; Smith, R. D.; Michael, K.; Baker, E. S., Enhancing glycan isomer separations with metal ions and positive and negative polarity ion mobility spectrometry-mass spectrometry analyses. *Anal. Bioanal. Chem.* **2017,** 409, (2), 467.
25. Struwe, W.; Benesch, J.; Harvey, D.; Pagel, K., Collision cross sections of high-mannose N-glycans in commonly observed adduct states–identification of gas-phase conformers unique to [M− H]− ions. *Analyst* **2015,** 140, (20), 6799-6803.
26. Huang, Y.; Dodds, E. D., Ion-neutral collisional cross sections of carbohydrate isomers as divalent cation adducts and their electron transfer products. *Analyst* **2015,** 140, (20), 6912-6921.
27. Hoffmann, W.; Hofmann, J.; Pagel, K., Energy-resolved ion mobility-mass spectrometry—a concept to improve the separation of isomeric carbohydrates. *J. Am. Soc. Mass Spectrom.* **2014,** 25, (3), 471-479.
28. Huang, Y.; Dodds, E. D., Ion mobility studies of carbohydrates as group I adducts: isomer specific collisional cross section dependence on metal ion radius. *Anal. Chem.* **2013,** 85, (20), 9728-9735.
29. Zhu, F.; Glover, M. S.; Shi, H.; Trinidad, J. C.; Clemmer, D. E., Populations of metal-glycan structures influence MS fragmentation patterns. *J. Am. Soc. Mass Spectrom.* **2015,** 26, (1), 25-35.
30. Huang, Y.; Dodds, E. D., Discrimination of isomeric carbohydrates as the electron transfer products of group II cation adducts by ion mobility spectrometry and tandem mass spectrometry. *Anal. Chem.* **2015,** 87, (11), 5664-5668.
31. Yang, H.; Shi, L.; Zhuang, X.; Su, R.; Wan, D.; Song, F.; Li, J.; Liu, S., Identification of structurally closely related monosaccharide and disaccharide isomers by PMP labeling in conjunction with IM-MS/MS. *Sci. Rep.* **2016,** 6, srep28079.
32. Fenn, L. S.; McLean, J. A., Enhanced carbohydrate structural selectivity in ion mobility-mass spectrometry analyses by boronic acid derivatization. *ChCom* **2008**, (43), 5505-5507.
33. Hilderbrand, A. E.; Myung, S.; Clemmer, D. E., Exploring crown ethers as shift reagents for ion mobility spectrometry. *Anal. Chem.* **2006,** 78, (19), 6792-6800.
34. Bohrer, B. C.; Clemmer, D. E., Shift reagents for multidimensional ion mobility spectrometry-mass spectrometry analysis of complex peptide mixtures: evaluation of 18-Crown-6 ether complexes. *Anal. Chem.* **2011,** 83, (13), 5377-5385.
35. Gaye, M.; Nagy, G.; Clemmer, D.; Pohl, N., Multidimensional analysis of 16 glucose isomers by ion mobility spectrometry. *Anal. Chem.* **2016,** 88, (4), 2335-2344.

36. Yamaguchi, Y.; Nishima, W.; Re, S.; Sugita, Y., Confident identification of isomeric N-glycan structures by combined ion mobility mass spectrometry and hydrophilic interaction liquid chromatography. *Rapid Commun. Mass Spectrom.* **2012,** 26, (24), 2877-2884.
37. Leijdekkers, A. G.; Huang, J.-H.; Bakx, E. J.; Gruppen, H.; Schols, H. A., Identification of novel isomeric pectic oligosaccharides using hydrophilic interaction chromatography coupled to traveling-wave ion mobility mass spectrometry. *Carbohydr. Res.* **2015,** 404, 1-8.
38. Zhong, X.; Chen, Z.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L., Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags. *Anal. Chem.* **2015,** 87, (13), 6527-6534.
39. Lareau, N. M.; May, J. C.; McLean, J. A., Non-derivatized glycan analysis by reverse phase liquid chromatography and ion mobility-mass spectrometry. *Analyst* **2015,** 140, (10), 3335-3338.
40. Hofmann, J.; Stuckmann, A.; Crispin, M.; Harvey, D. J.; Pagel, K.; Struwe, W. B., Identification of Lewis and Blood Group Carbohydrate Epitopes by Ion Mobility-Tandem-Mass Spectrometry Fingerprinting. *Anal. Chem.* **2017,** 89, (4), 2318-2325.
41. Gray, C. J.; Schindler, B.; Migas, L. G.; Picmanova, M.; Allouche, A. R.; Green, A. P.; Mandal, S.; Motawia, M. S.; Sánchez-Pérez, R.; Bjarnholt, N.; Møller, B. L.; Rijs, A. M.; Barran, P. E.; Compagnon, I.; Eyers, C. E.; Flitsch, S. L., Bottom-up elucidation of glycosidic bond stereochemistry. *Anal. Chem.* **2017,** 89, (8), 4540-4549.
42. Gaye, M.; Kurulugama, R.; Clemmer, D., Investigating carbohydrate isomers by IMS-CID-IMS-MS: precursor and fragment ion cross-sections. *Analyst* **2015,** 140, (20), 6922-6932.
43. Both, P.; Green, A.; Gray, C.; Šardzík, R.; Voglmeir, J.; Fontana, C.; Austeri, M.; Rejzek, M.; Richardson, D.; Field, R.; Widmalm, G.; Flitsch, S.; Eyers, C., Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat. Chem.* **2014,** 6, (1), 65-74.
44. Pu, Y.; Ridgeway, M. E.; Glaskin, R. S.; Park, M. A.; Costello, C. E.; Lin, C., Separation and identification of isomeric glycans by selected accumulation-trapped ion mobility spectrometry-electron activated dissociation tandem mass spectrometry. *Anal. Chem.* **2016,** 88, (7), 3440-3443.
45. Morrison, K. A.; Clowers, B. H., Differential Fragmentation of Mobility-Selected Glycans via Ultraviolet Photodissociation and Ion Mobility-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017,** 6, (28), 1236-1241.
46. Lee, S.; Valentine, S. J.; Reilly, J. P.; Clemmer, D. E., Analyzing a mixture of disaccharides by IMS-VUVPD-MS. *Int. J. Mass spectrom.* **2012,** 309, 161-167.
47. Masellis, C.; Khanal, N.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R., Cryogenic Vibrational Spectroscopy Provides Unique Fingerprints for Glycan Identification. *J. Am. Soc. Mass Spectrom.* **2017**, 1-6.
48. Hernandez, O.; Isenberg, S.; Steinmetz, V.; Glish, G. L.; Maitre, P., Probing mobility-selected saccharide isomers: selective ion–molecule reactions and wavelength-specific IR activation. *The Journal of Physical Chemistry A* **2015,** 119, (23), 6057-6064.
49. Khanal, N.; Masellis, C.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R., Glycosaminoglycan analysis by cryogenic messenger-tagging IR spectroscopy combined with IMS-MS. *Anal. Chem.* **2017**.
50. May, J. C.; Morris, C. B.; McLean, J. A., Ion Mobility Collision Cross Section Compendium. *Anal. Chem.* **2016,** 89, (2), 1032-1044.
51. Pagel, K.; Harvey, D. J., Ion mobility–mass spectrometry of complex carbohydrates: collision cross sections of sodiated N-linked glycans. *Anal. Chem.* **2013,** 85, (10), 5138-5145.

52. Hofmann, J.; Struwe, W. B.; Scarff, C. A.; Scrivens, J. H.; Harvey, D. J.; Pagel, K., Estimating collision cross sections of negatively charged N-glycans using traveling wave ion mobility-mass spectrometry. *Anal. Chem.* **2014,** 86, (21), 10789-10795.
53. Glaskin, R. S.; Khatri, K.; Wang, Q.; Zaia, J.; Costello, C. E., Construction of a Database of Collision Cross Section Values for Glycopeptides, Glycans, and Peptides Determined by IM-MS. *Anal. Chem.* **2017,** 89, (8), 4452-4460.
54. Struwe, W. B.; Pagel, K.; Justin, L.; Benesch, P.; Harvey, D. J.; Campbell, M. P., GlycoMob: an ion mobility-mass spectrometry collision cross section database for glycomics. *Glycoconjugate J.* **2016,** 33, (3), 399.
55. Wiederschain, G. Y., Essentials of glycobiology. *Biochemistry (Moscow)* **2009,** 74, (9), 1056-1056.
56. Pabst, M.; Bondili, J. S.; Stadlmann, J.; Mach, L.; Altmann, F., Mass+ retention time= structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal. Chem.* **2007,** 79, (13), 5051-5057.
57. Wojcik, R.; Webb, I. K.; Deng, L.; Garimella, S. V.; Prost, S. A.; Ibrahim, Y. M.; Baker, E. S.; Smith, R. D., Lipid and glycolipid isomer analyses using ultra-high resolution ion mobility spectrometry separations. *International journal of molecular sciences* **2017,** 18, (1), 183.
58. Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T., Collision induced unfolding of intact antibodies: rapid characterization of disulfide bonding patterns, glycosylation, and structures. *Anal. Chem.* **2015,** 87, (22), 11509-11515.
59. Sarbu, M.; Robu, A. C.; Ghiulai, R. M.; Vukelić, Z. e.; Clemmer, D. E.; Zamfir, A. D., Electrospray ionization ion mobility mass spectrometry of human brain gangliosides. *Anal. Chem.* **2016,** 88, (10), 5166-5178.
60. Sturm, R. M.; Lietz, C. B.; Li, L., Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry. *Rapid Commun. Mass Spectrom.* **2014,** 28, (9), 1051-1060.
61. Harvey, D. J.; Scarff, C. A.; Edgeworth, M.; Crispin, M.; Scanlan, C. N.; Sobott, F.; Allman, S.; Baruah, K.; Pritchard, L.; Scrivens, J. H., Travelling wave ion mobility and negative ion fragmentation for the structural determination of N‐linked glycans. *Electrophoresis* **2013,** 34, (16), 2368-2378.
62. Harvey, D. J.; Sobott, F.; Crispin, M.; Wrobel, A.; Bonomelli, C.; Vasiljevic, S.; Scanlan, C. N.; Scarff, C. A.; Thalassinos, K.; Scrivens, J. H., Ion mobility mass spectrometry for extracting spectra of N-glycans directly from incubation mixtures following glycan release: application to glycans from engineered glycoforms of intact, folded HIV gp120. *J. Am. Soc. Mass Spectrom.* **2011,** 22, (3), 568-581.
63. Harvey, D. J.; Crispin, M.; Bonomelli, C.; Scrivens, J. H., Ion Mobility Mass Spectrometry for Ion Recovery and Clean-Up of MS and MS/MS Spectra Obtained from Low Abundance Viral Samples. *J. Am. Soc. Mass Spectrom.* **2015,** 26, (10), 1754-1767.
64. Fenn, L. S.; McLean, J. A., Structural separations by ion mobility-MS for glycomics and glycoproteomics. *Mass Spectrometry of Glycoproteins: Methods and Protocols* **2013**, 171-194.
65. Fenn, L. S.; McLean, J. A., Simultaneous glycoproteomics on the basis of structure using ion mobility-mass spectrometry. *Molecular BioSystems* **2009,** 5, (11), 1298-1302.

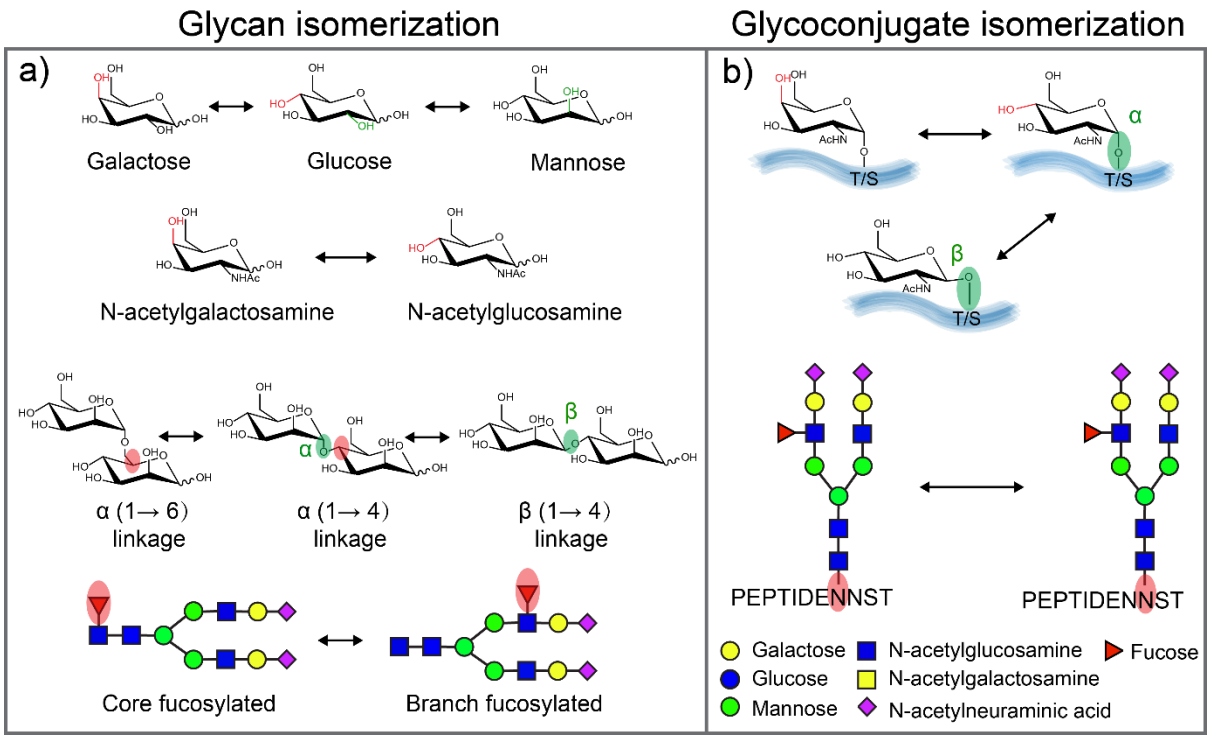


Figure 1. The isomerization of glycan and glycoconjugates. **a)** The building blocks (monosaccharides) that compose larger glycans are structural isomers (Hexose: galactose, glucose, mannose, N-acetylhexosamine: N-acetylgalactosamine, N-acetylglucosamine); monosaccharides can be connected either α- or β-stereochemistry at multiple potential linkage position; fucose could be either be attached to N-glycan core or branches. **b)** Epimeric glycoconjugates results from alternative configurations (α- or β-) at the anomeric linkages or the presence of epimeric glycan monomers (galactose or glucose), scheme modified from reference [43]; two isomeric N-glycopeptides differ in the site of N-glycan attachment.

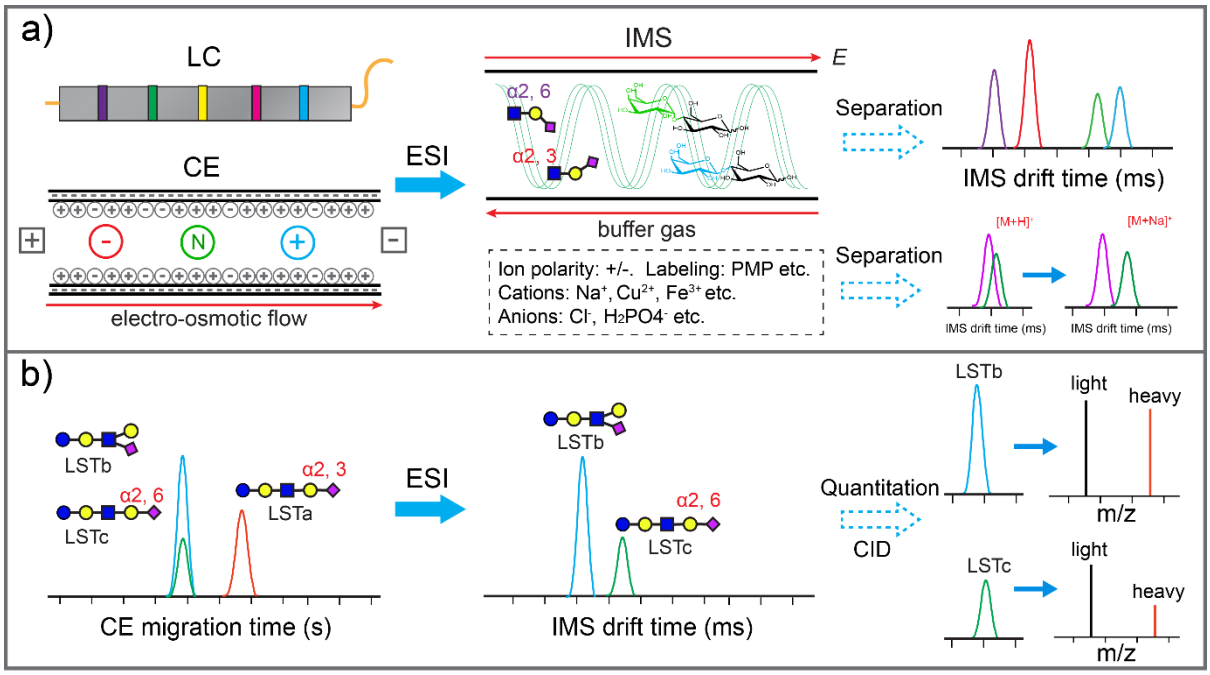


Figure 2. Coupling IM with LC or CE. **a)** After LC or CE separation, analytes were ionized by ESI and subject to another dimension of separation afforded by IM based on their shape and charge through a buffer gas under a weak electric field (*E*). Sample preparation or gas-phase chemistry could be manipulated to improve glycan isomer separation. **b)** CE-ESI-TWIM-MS/MS analysis of a mixture of aminxoyTMT^6-128 (light) and aminoxyTMT^6-131 (heavy) differentially labeled sialyllacto-N-tetraose a, b, c (LSTa, LSTb, LSTc). CE was able to separate LSTa with LSTb/c, but was unable to resolve LSTb and LSTc. Benefiting from another dimension of separation afforded by TWIM, baseline separation between LSTb and LSTc was achieved, which enables quantitative analysis of each isomer following MS/MS [38].

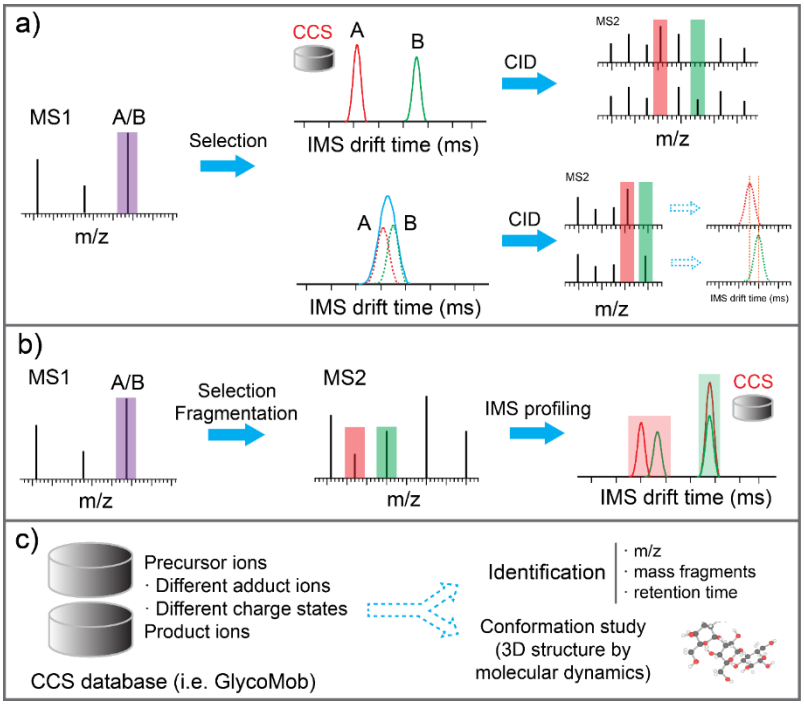


Figure 3. IM-MS analysis of precursor ions and their fragments. **a)** Two scenarios exist for co-eluted isomers A/B for IM-MS analysis after being selected by quadrupole. Scenario one: A and B could be baseline-separated by IM. The mobility-selected ions could be subject to CID separately and signature product ions could be obtained for each †isomer. Scenario two: A and B could not be completely resolved by IM. Signature products ions were obtained for unresolved species. Drift time profiles of these signature products ions were extracted from total drift time profiles to differentiate A and B. **b)** Co-eluted isomers A/B were selected by quadrupole for MS/MS and the drift time profiles could be obtained for all product ions. Those product ions that are indicative of the isomeric structures of the analyte could be distinguished by IM and be used to differentiate the isomer. **c)** The CCS of both precursor ions and product ions could be measured and implemented into a CCS database. The CCS values could be used as an additional parameter for glycan identification besides the commonly used m/z, mass fragments, and retention time. Furthermore, conformational study could be conducted by molecular dynamics to acquire the 3D structures of glycans.

Chapter 3

Recent advances in mass spectrometry (MS)-based glycoproteomics in complex biological samples

Adapted from **Chen, Z.**,[#] Huang, J.,[#] Li, L. S*ubmitted to Trends Analyt. Chem. (2018) (# These authors contributed equally)*

Author contribution: manuscript was written by Chen, Z., Huang, J. and edited by Chen, Z., Huang, J., L. Li.

Abstract

Protein glycosylation has been well-know to play a key role in various biological processes as well as in disease-related pathological progesssion. To better understand glycosylation's role in an undelying mechanism, comprehensive characterization and quantitation of glycoproteome in different biological systems is needed. Mass spectrometgry (MS)-based glycoproteomics is a powerful approach that provdies a system-wide profiling of the glycoproteome in a high-throughput manner. There has been a significant technology advances in this field, including improved glycopeptide enrichment, hybrid fragmentation techniques, maturing specialized softwares, and effective quantitation strategies, as well as more dedicated workflows. With incresingly sophiscated glycoproteomis tools on hand, researshers have extensively adpated this approach to explore different biology systems both in terms of in-depth glycoprteome profiling and comparative glycoproteome analysis. Quantitative glycoproteomics enables reserachers to discover novel glycosylation-based biomarker in various diseases with a potential to offer better sensitivity and specificity for disease diagnsois. In this review, we present the recent methodology developments in MS-based glycoproteomics and highlight its application in answering differernt biology questions in complex biological system.

Keywords: Mass Spectrometry, Glycoproteomics, N-glycosylation, O-glycosylation, Biological samples, Biomarker, Disease

1. Introduction

As one of the most common protein post-translational modifications (PTMs), protein glycosylation plays an important role in protein stability, intra- and intercellular signaling, fertilization, embryogenesis, organ development, hormone activity, and immunological regulation [1]. It is estimated that half of the proteins expressed in cells are glycoproteins. There are many types of protein glycosylation, and the most widely studied types are N-linked (amide nitrogen of asparagine residue) and O-linked (hydroxyl oxygen of serine or threonine residue). There is a consensus amino acid sequence (Asn-X-Thr/Ser (X is any amino acid except proline)) that contains the glycosylation site of N-glycoproteins, while no sequences has been found in O-linked glycoproteins yet.

Numerous studies have shown altered glycosylation played a key role in the pathological process along the disease progression. Mass spectrometry (MS)-based glycoproteomics is a powerful, high-throughput approach that enables system-wide screening of glycosylation-based biomarkers. In fact, many current disease biomarker are glycoproteins, such as CEA for colorectal cancer, CA-125 for ovarian cancer, and AFP for hepatocellular carcinoma etc., and the glycans attached to them have been shown to be altered during oncogenesis [2-4]. With glycosylation being particularly sensitive to malignant transformation, glycosylation-based biomarkers hold great promise to improve the sensitivity and specificity of current protein-based biomarkers and may eventually contribute to disease early diagnosis and better treatment [5]. Therefore, comprehensive profiling of protein glycosylation is prerequisite to better understand its role in these pathological and physiological processes.

Over the past decade, substantial progress has been made to obtain detailed information of protein glycosylation, including glycan structures, glycosylation site and its occupancy, and protein sequence. Traditionally, there are two very different approaches to study protein glycosylation: (1) the 'glycomics' approach, which focuses on the glycan structures after glycan release from proteins and other sugar-containing moieties and (2) the 'glycoproteomics' approach, which explores the localization and structural elucidation of glycans, as well as protein sequence. In this review, we will focus on MS-based glycoproteomics approach. A lot of advances have been made in this field, and here we highlight recent methodologies development and their application toward the comprehensive understanding of the function of protein glycosylation.

2. Glycopeptides enrichment

Comprehensive profiling of the glycoproteome from a complex biological sample is still challenging due to the high dynamic range of proteins and the micro- and macro-heterogeneity of glycosylation. Isolating glycopeptides from complex samples by an appropriate enrichment method is the most efficient way to reduce the sample complexity and achieve an in-depth glycoproteome analysis.

2.1 Hydrazide chemistry enrichment

Hydrazide chemistry (HC) enrichment is based on the formation of covalent bonds between the $NaIO_4$ oxidized glycans and the hydrazide groups on the hydrazide beads. The advantage of HC method is its high enrichment specificity, as normally 90% of the enriched peptides are glycopeptides [6]. In the original workfow, the glycopeptides captured on hydrazide beads were deglycosylated and released by the treatment of PNGase F for glycosylation site analysis [7, 8].

Recently, the capture and release steps were modified to allow glycopeptides to be released without losing the glycan. Specifically, samples were treated with moderate $NaIO_4$ to selectively oxidize the terminal sialic acid of the glycan to generate an aldehyde while leaving the other parts intact. Next, the sialylated glycopeptides were captured by the hydrazide beads and released through acid hydrolysis of the glycosidic bond of sialic acid by trifluoroacetic acid (TFA) [9, 10]. This method enabled sialylated glycopeptides to be analyzed, but, unfortunantly, the degree of sialylation information was lost. [11]. To preserve the sialylation information, Nishimura *et al.* employed ice-cold 1M hydrochloride to cleave hydrazone bond between the sialic acid and hydrazide beads, allowing the sialic acid to remain on the glycan [12].

2.2 Lectin affinity chromatography

Lectin affinity chromatography (LAC) is another popular enrichment method for protein glycosylation analysis and has been approved by FDA for cancer glycoprotein biomarker detection [13]. Several well characterized lectins had been used for glycopeptide enrichment, which is based on the specific affinity of lectin to glycans with specific structure motif [14]. Multiple lectins can be combined to improve the glycoproteome coverage [15-17], and adding other enrichment methods sequentially after LAC enrichment would further improve the enrichment specificity [18]. Moreover, lectin enrichment has also been incorporated in serial online reactors to allow simultaneous online proteolysis and glycopeptide enrichment, which is useful for the analysis of minute samples [19].

2.3 Hydrophilic interaction chromatography

Enrichment of glycopeptides by hydrophilic interaction chromatography (HILIC) is based upon glycopeptides being more hydrophilic than non-glycopeptides due to the attached glycans [20]. To

increase the hydrophilicity difference between glycopeptides and non-glycopeptides, ion pairing reagents like TFA can be used [21]. Compared with the HC and LAC method, HILIC method is more versatile and thus can provide a more comprehensive glycoproteome profile [22]. The disadvantage of HILIC is its poor enrichment specificity, which means new HILIC materials with stronger hydrophilic functional group need to be developed to improve the specificity [23-26]. The enrichment specificity can also be further improved by combing the HILIC with other enrichment methods such as LAC [27, 28]. New devices, which integrate the HILIC materials in micro-column or tip, have been developed to minimize the sample loss during enrichment procedure and facilitate the detection of low abundant glycopeptide [29-33].

3. Characterization and quantitation of glycopeptides

Qualitative characterization of glycopeptides includes two aspects: glycosylation site profiling and site-specific intact glycopeptide analysis. A typical MS-based glycoproteomics was shown in **Fig. 1**.

3.1 Glycosylation site profiling

3.1.1 N-glycosylation site profiling

Since N-glycosylation site profiling was originally performed by deglycosylaiton with peptide-N-glycosidase F (PNGase F) or endo-β-N-acetylglucosaminidase F&H (endo F&H), some studies have focused on improving the deglycosylation efficiency. Zou *et al.* found that when glycosylation site profiling was performed by HC method, glycopeptides with an N-terminal serine/threonine can be oxidized on both the N-termini and glycans; thus, this type of glycopeptides cannot be released by PNGase F treatment due to being covalently coupling to the

hydrazide beads through the N-termini. To overcome this problem, they utilized a peptide N-terminal protection strategy to block the primary amine groups on peptides, which avoided the adjacent amino alcohols on peptide N-termini being oxidized. The results showed that this strategy successfully prevented the oxidation of peptide N-termini and significantly improved the coverage of glycoproteome [34]. Recently, the same group found that releasing the glycopeptides captured on hydrazide beads by PNGase F deglycosylation was inefficient due to steric hindrance in the heterogeneous condition. Thus, they developed a hydroxylamine-assisted PNGase F deglycosylation method which used the hydroxylamine to efficiently cleave hydrazone bonds by transamination and release intact glycopeptides. As deglycosylation of the released glycopeptides was performed in homogeneous condition, the recovery rate of deglycosylated peptides was improved significantly [35]. Other paper by Zhang *et al.* found that N-terminal glycosylated peptides are difficult to be deglycosylated due to limitation of PNGase F enzymatic specificity, which requires that the glycosylation site is linked to an amino acid at both the N- and C- termini. To overcome this drawback, they developed a N-terminal site-selective succinylation strategy by incorporating an amide bond to mimic an amino acid at the peptide N-termini, which greatly improved N-glycosylation sites coverage [36].

Other researchers mainly resorted to combining different sample preparation techniques, enrichment methods, and fractionation strategies to improve the glycoproteome coverage. Mann *et al.* developed an N-glyco-FASP sample preparation approach, where the lectin column in conventional method was replaced with ultrafiltration units, to decrease the glycopeptide loss. In this method, the glycopeptides were enriched by binding to lectins on the top of a filter, which

greatly reduced the sample loss and improved the glycosite coverage. The robustness of this approach was successfully demonstrated in the large-scale glycosylation site profiling in mouse plasma and four different tissues where 6367 N-glycosylation sites identified. Combining different enrichment methods is also an effective approach to increase the glycosite coverage [37]. Recently, Qian *et al.* combined two widely used glycopeptide enrichment methods, HC and HILIC, for N-glycosylation site analysis in the secretome of two human hepatocellular carcinoma (HCC) cell lines. A total of 1,212 unique N-glycosylation sites from 611 N-glycoproteins were confidently identified. Overall, the N-glycosylation site overlap of two methods was only 28.4% [38]. Zou *et al.* performed a similar strategy which combined the click maltose-HILIC and the HC method to comprehensively map the N-glycosylation sites of human liver tissue. Altogether, 14,480 N-glycopeptides, corresponding to 2,210 N-glycoproteins and 4,783 N-glycosylation sites, were identified [39]. As another example, Yang *et al.* combined seven protease treatments (trypsin, trypsin coupled with Lys-C (Try&Lys), trypsin coupled with Glu-C (Try&Glu), Lys-C, Glu-C, chymotrypsin and pepsin), four different enrichment techniques (HILIC, ZIC-HILIC, HC, and TiO_2 chromatography), and two different fractionation strategies (SCX and high-pH RP), which aided in identifying a total of 13492 N-glycopeptides, corresponding to 8386 N-glycosylation sites on 3982 proteins in the mouse brain. Considering the efficiency and simplicity, a workflow combined used the trypsin, Try+Lys and Try+Glu for protein digestion, HILIC and ZIC-HILIC for the glycopeptide enrichment, and 1D-RPLC-MS/MS for N-glycopeptide detection can also acquire a comparable glycosite coverage [40].

3.1.2 O-glycosylation site profiling

Mapping of O-glycosylation is also a thriving field. Among the different types of O-glycosylation, the O-GlcNAcylation and O-GalNAcylation are the most widely studied [41, 42]. Profiling of the O-glycosylation sites is even more difficult compared to N-glycosylation due to a lack of a consensus sequon and the lack of an enzyme that can effectively deglycosylate the O-linked glycans.

To date, the most successful approach for the profiling of O-GlcNAcylation was metabolic and enzymatic labeling, which incorporates an azide containing group to the O-GlcNAc moiety [43]. Then, the derivatized O-GlcNAc is enriched by an alkynyl biotin or photocleavage tag containing alkynyl beads to be analyzed by LC-MS/MS. By using this highly specific strategy, tens to hundreds of O-GlcNAcylation sites can be mapped [44-47]. The drawback of this approach its low labeling efficiency, leading to the limited coverage of O-GlcNAcylation sites [48]. Besides the enzymatic and metabolic labeling methods, Burlingame *et al.* developed a lectin weak affinity chromatography (LWAC) strategy to enrich O-GlcNAc peptides with wheat germ agglutinin (WGA) lectin [49]. The same group lately optimized this LWAC strategy and identified over 1750 sites of O-GlcNAcylation from murine synaptosomes [50]. Due to the particularly low abundance, low hydrophilicity of the O-GlcNAcylation peptides, and severe interference from other N/O-glycopeptides, isolating O-glycopeptides from a complex sample by HILIC enrichment was originally thought to be ineffective. However, after combining PNGase F, sialidase and O-glycosidase to selectively cleave and remove most of the N/O-linked glycans in glycoproteins, Shen *et al.* was able to eliminate the interference of other N/O-glycopeptides while still preserving the O-GlcNAcylation modified peptides. Benefiting from the improved enrichment specificity of the O-GlcNAc peptides, a total of 474 O-GlcNAc peptides from 457 proteins were identified from

human urinary sample. In comparison, performing HILIC enrichment without the deglycosylation step only identified 107 O-GlcNAc proteins, and an immunoprecipitation (IP) approach using an anti-O-GlcNAc antibody only profiled 31 O-GlcNAc proteins [51].

For O-GalNAcylation, the glycan structures are of higher diversity than O-GlcNAcylation. To facilitate the MS identification of these glycopeptide sequences and their attached sites, Medzihradszky *et al.* utilized exoglycosidase digestion to partially deglycosylate O-GalNAcylation peptides and reduce the complexity of glycan structures and was able to identify 124 O-GalNAcylation sites in 51 O-GalNAcylation proteins from human serum samples [52]. Besides this *in vitro* approach, Clausen *et al.* developed an alternative method called "simple cell" strategy *in vivo*, which utilized a zinc-finger nuclease gene targeting to block the O-GalNAcylation elongation pathway to generate short glycan homogenous O-GalNAcylation. This strategy allowed directly enrichment by the LWAC method followed by MS/MS detection of O-GalNAcylation peptides from different cell lines [53]. Recently, they extended this approach to characterize samples from 12 human cell lines and profiled almost 3000 O-GalNAcylation sites in over 600 O-GalNAcylation glycoproteins, which represented the first map of the human O-glycoproteome [54].

3.2 Site-specific characterization of intact glycopeptides

Along with rapid development of glycosylation site profiling, two major breakthroughs have significantly facilitated the site-specific characterization of intact glycopeptides. These breakthroughs are the advancement of MS/MS dissociation methods towards acquiring both the glycan and peptide backbone fragments and the development of new search engines to decipher the MS/MS spectra of intact glycopeptides.

3.2.1 Comparison of dissociation methods for intact glycopeptides

The dissociation modes for peptide analysis mainly include collision-induced dissociation (CID), beam-type CID (occurs in triple quadrupole (QQQ) and quadrupole time-of-flight (Q-TOF) instruments, and the so called high-energy collisional dissociation (HCD) in Thermo-Fisher™ instruments), and electron transfer/capture dissociation (ETD/ECD). Each of these methods alone can't provide a full picture of the glycopeptide structure. [55]. CID prefers to break glycosidic bonds, and it generates strong characteristic ions of peptides bearing different numbers of glycans after stepwise release of peripheral monosaccharides (Y ions) (**Fig. 2**). It provides abundant information for glycan structure deciphering but poor peptide backbone identification. As for HCD dissociation, in low collision energy, a series of Y ions are preferentially generated, which is similar to CID; while in high collision energy, the peptide backbone fragmentation can also be generated with a decreased intensity of Y ions. ETD/ECD mode mainly fragments the peptide backbone while leaving the glycan intact (**Fig. 2**), which is suitable for the localization of glycosylation sites with a wealth of peptide fragments [56].

3.2.2 Intact glycopeptides analysis by combing different dissociation methods

As no single dissociation method is available to produce a complete picture of intact glycopeptides, combining the complementary fragment information from multiple dissociation modes is an effective strategy to decipher the intact glycopeptides. Larson *et al.* combined CID and ECD to analyze desialylated glycopeptides, where CID-MS^2 spectra of glycopeptides were used for the glycan characterization and the subsequent CID-MS^3 spectra of selected CID-MS^2 fragment ions for peptide sequence identification. Moreover, ECD as a complementary peptide fragmentation

mode was used for the characterization of O-glycosylation sites, where 58 N- and 63 O-glycopeptides from 53 glycoproteins were identified and 40 of the 57 putative O-glycosylation sites were accurately localized [57]. However, the requirement of prior knowledge of the targeted peptides to be selected for MS^3 and the longer duty cycle due to ETD reaction time in ETD-MS^2 limit its capacity compared to HCD- and CID-MS^2, and currently only tens to hundreds of intact glycopeptides can be profiled from complex biological samples using this strategy [58].

Zou & Figeys *et al.* developed an alternative complementary method to interpret intact glycopeptides that is sequentially analyzing the deglycosylated peptides and intact glycopeptides by CID and HCD, respectively. A total of 811 N-glycosylation sites from 567 glycoproteins were identified from HEK293T membrane proteins, and 177 intact N-glycopeptides were also identified by manually integrating the CID and HCD spectra. The number of identified intact glycopeptides was much smaller than the number of identified N-glycosites, which can be attributed to the low ionization efficiency of intact glycopeptides and manual interpretation of the complicated MS/MS spectra [59]. Recently, the same group developed a fully-automated software platform for the high-throughput characterization of intact N-glycopeptides. They used the strong correlation of retention time to effectively remove the random matches and were able to control the probability of random matches within 1%. In total, 2249 intact glycopeptides, representing 1769 site-specific N-glycans on 453 glycosylation sites, were identified [60]. Yang *et al.* developed a similar strategy which profiled 1145 non-redundant glycopeptides from 225 core peptides and 95 glycoproteins from human serum samples [61].

3.2.3 Intact glycopeptides analysis by integrated dissociation methods

Rather than implementing two dissociation methods to get the complementary structure information in two separate LC-MS/MS runs, it would be beneficial if hybrid fragmentation spectra were acquired in a single run. To this end, the evolution of new dissociation modes has shed light on solving this problem. Among them, the stepped collision energy HCD (step-HCD), beam-type CID with high energy, and electron-transfer/higher-energy collision dissociation (EThcD) stand out.

As different HCD collision energies could generate complementary fragments, performing step-HCD (e.g. 30 ±10%) will give a more complete intact glycopeptide structure information in a single spectrum. Qian *et al.* first applied the step-HCD to analyze partially deglycosylated core-fucosylated glycopeptides in mouse liver tissue and HeLa cell samples and found that the overall performance increased by 7-fold [62]. Recently, this method has been wildly used in intact glycopeptides analysis and outstanding results were obtained [63, 64]. Current MS instruments can only provide a three-step collision energy in one spectrum, and more flexible collision energy settings could definitely improve intact glycopeptide analysis.

Under typical beam-type CID conditions, ions produced from dissociation of the peptide backbone are in low abundance. Zaia *et al.* found abundant peptide backbone fragments could be generated by increasing the collision energy, along with oxonium ions and intact peptide ions with varying numbers of saccharide units attached. They successfully used this approach for intact glycopeptide analysis from several standard N-glycoproteins [65]. Recently, Sung *et al.* utilized the same strategy in complex sample analysis and profiled 36 intact glycopeptides of 26 glycoproteins in a HeLa cell sample [66]. Ye & Zou *et al.* continually explored this strategy in intact O-GalNAcylation peptide

analysis and established an automated workflow for O-GalNAcylation peptide MS^2 spectra interpretation, with an identification of 407 intact O-GalNAcylation peptides from 93 glycoproteins in human serum sample [67].

Integrating the HCD and ETD in one spectrum (**Fig. 2**), EThcD also enables both glycan and peptide fragments to be acquired [68]. Li *et al.* first optimized the parameters of EThcD for intact glycopeptides analysis, and found that the efficiency of the dissociation was greatly improved by using charge-dependent optimized ETD reaction times [28]. Large scale experiments in rat carotids collected over the course of restenosis progression resulted in over 2000 N-glycopeptide identifications [69]. Qian *et al.* found that EThcD provided more complete fragmentation information on O-GalNAcylation peptides and a more confident site localization of O-GalNAcylation than HCD. By combining multiple enzyme digestions and multidimensional separation, they identified 173 O-glycosylation sites, 499 non-redundant intact O-glycopeptides, and 6 glycan compositions originating from 49 O-glycoprotein groups from normal human serum [70].

3.2.4 Database search for intact glycopeptides analysis

The biggest challenge for intact glycopeptide characterization is the accurate interpretation of the spectra. Based on different strategies, several new search engines have been developed for intact glycopeptide identification, such as GlycoMaster DB [71], GPQuest [72], I-GPA [73], Byonic [74], SweetNET [75], pGlyco [76], pGlyco 2.0 [63], etc. Particularly, pGlyco 2.0 conducted a comprehensive FDR evaluation at all three levels of glycans, peptides and glycopeptides, greatly improving the accuracy of intact glycopeptide identification (**Fig. 3**). Some other analytical tools, like MAGIC and SugarQb, which translate the intact glycopeptide spectra and enable them to be analyzed using

current peptide search engines, have also been developed [64, 66, 77]. Due to space limitations and lack of standard glycopeptide spectral datasets, a fair comparison between different software has not been performed [78]. Comprehensive evaluation of the current software regarding the coverage of the glycoprotome and quality control of the identification results would provide valuable insights for future software design [79]. Large-scale glycoproteomics research would especially benefit greatly from the improvement of automated glycopeptide identification, due to the large volumes of data generated.

3.3 Quantitation of glycopeptides

Quantitation of protein glycosylation could be performed at the glycan, glycopeptide or glycoprotein level based on the target molecule, and at relative or absolute quantitation levels based on the strategies used. Absolute quantitation is often conducted by employing targeted MS approach. As the theme of the current review is non-targeted bottom-up glycoproteomics, the quantitation strategy discussed here will focus on quantitation at the glycopeptide level.

3.3.1 Label-free quantitation

The label-free approach has been regularly used in proteomics studies to measure protein abundance changes, and it offers the advantage of a simple workflow, low cost and high proteome coverage [80]. Normalization is needed to overcome the MS response variations in different samples and reliable quantitation results could be obtained by normalizing the data to the total ion abundance [81, 82]. However, it could be problematic for glycopeptide analysis due to the low ionization efficiency of glycopeptides, which means small changes of nonglycosylated interferences could lead to large variability in quantitative assays. To overcome this problem,

Rebecchi *et al.* developed a new normalization strategy based on the intensity of all glycopeptides and a two-tiered quantitative analysis to discriminate between glycosylation changes of a given protein and glycoprotein's concentration changes [83]. Additionally, the large volumes of data produced by label-free experiments need rigorous statistical assessment for accurate data processing and interpretation, which requires effective algorithm models and software tools to be developed [84]. Mayampurath *et al.* developed a novel ANOVA-based mixed effects model for label-free glycopeptide quantitation and demonstrated its effectiveness by applying this method to biomarker discovery in human serum [85]. To facilitate simultaneous identification and label-free quantitation of glycopeptides, Park *et al.* developed an automated Integrated GlycoProteome Analyzer (I-GPA) platform and successfully quantified 598 N-glycopeptides from human plasma sample [86].

3.3.2 Label-based quantitation

Compared with label-free methods, the greatest advantage of stable isotope labeling is that different samples are mixed together and analyzed simultaneously, which largely reduces instrument time and run-to-run variations. In general, stable isotope labeling can be classified into three major categories: metabolic labeling, chemical labeling and enzymatic labeling.

The most commonly used metabolic labeling in quantitative proteomics is the stable isotope labeling by amino acids in cell culture (SILAC), which incorporates stable isotope-encoded essential amino acids into living cells [87]. The main advantage of SILAC is that it allows different samples to be combined at the intact cell level, minimizing the possible quantitation error introduced by the sample preparation process [88]. By incorporating a glycopeptide enrichment step,

the regular SILAC workflow can be easily modified for quantitative glycoproteomics studies. After treatment with PNGase F, the total glycosylation expression changes at each site can be quantified through comparison of light and heavy-labeled deglycopeptides, and a number of studies have successfully utilized this approach to quantify glycosylation changes on hundreds of N-glycosites [89-93]. Furthermore, Parker *et al.* utilized the SILAC approach to quantify the intact glycopeptides without PNGase F treatment, which enabled the changes in N-glycosylation microheterogeneity to be revealed [94]. By combining glycopeptide enrichment using hydrazide chemistry with SILAC, Taga *et al.* conducted a quantitative analysis of O-glycosylation and showed increased glycosylation of collagen in Osteogenesis Imperfecta [95].

However, the disadvantage of metabolic labeling is that some biological systems are not suited to efficient metabolic labeling and the cost is relatively high [96]. To this end, chemical labeling approaches have been developed to label proteins or peptides extracted from tissues/cells with stable isotope-incorporated tags. In early 2003, Zhang *et al.* used stable isotope labeling by succinic anhydride after glycoprotein capture by hydrazide beads for identification and quantitation of N-glycopeptides [97]. However, succinic anhydride labeling method requires repeated labeling to achieve reaction completeness and side reactions may happen during the process [98]. To overcome this problem, Sun *et al.* developed an approach that enables sequential glycopeptide enrichment and dimethyl labeling on hydrazide beads, which showed high quantitation accuracy over a 2 order of dynamic range [99]. However, both succinic anhydride and dimethyl labeling has a limited capability for quantitative analysis across different samples; hence, isobaric tags have been developed to allow for multiplexing capability, such as 10-plex tandem mass tag (TMT),[100, 101] 8-

plex isobaric tags for relative and absolute quantitation (iTRAQ),[102, 103] and 12-plex N,N-dimethyl leucine (DiLeu) isobaric tags [104, 105] etc. Employing a 6-plex TMT labeling strategy, Kroksveen *et al.* conducted a quantitative glycoproteomics analysis between 21 subjects in relapsing-remitting multiple sclerosis group and 21 subjects in neurological control group, and successfully quantified 1700 deglycopeptides with 235 deglycopeptides showing significant differences between disease group and control group [106]. Notably, Braga *et al.* conducted a global comparative proteomic study, as well as changes in N-glycosylation, phosphorylation, and Lys-acetylation with a 4-plex iTRAQ in parallel [107]. The reason that multiple quantitative PTMs analysis could be conducted in parallel is because both proteome and PTMs analysis shared the same upper stream steps and samples could be split into aliquots and subject to different PTMs-targeted enrichment methods after isobaric labeling. Such capacity allows multiple PTMs to be analyzed from a limited amount of sample and largely facilitates the study of cross-talking between different PTMs. Besides chemical labeling, enzymatic labeling was also developed by incorporating ^{18}O into the peptides during the enzyme-catalyzed digestion process [108]. Later, Liu *et al.* developed a tandem ^{18}O stable isotope labeling strategy for quantitation of N-Glycoproteome by combining ^{18}O labeling in the C-terminal carboxylic acid during proteolytic process and another ^{18}O labeling in the asparagine residue during deglycosylation process by PNGase F hydrolysis [109].

4. Application in MS-based glycoproteomics in complex biological sample

4.1 In-depth glycoproteome profiling in complex biological sample

4.1.1 Human Serum

In-depth glycoproteome profiling have been extensively conducted in different biology systems, including body fluids, cells and tissues, etc. Serving as an indicator of physiological and pathological states alteration in the body system, serum/plasma is the most common clinical specimen for disease diagnosis. The majority of serum proteins are glycosylated as many proteins are secreted in glycosylated form, with an estimated 50% after removing high abundance proteins [110]. To facilitate the detection of low abundant glycoproteins, Sparbier *et al.* utilized magnetic lectins (ConA, LCA, WGA) beads and boronic acid beads for the enrichment on both protein and peptide level, resulting in 95 N-glycosylation sites from 193 N-glycoproteins [111]. Nevertheless, the coverage is still not desirable mainly due to the serum extreme complexity and highly dynamic concentration spanning over 10 orders of magnitude [112]. To further decrease sample complexity, various approaches have been applied, including immunoaffinity depletion of high-abundance serum proteins (albumin, IGG etc.), sequential enrichment strategies (lectins, HILIC etc.), off-line fractionation (HpH, SCX) and 2D-LC, which yielded more than 600 N-glycosylation sites from over 300 N-glycoproteins [113, 114]. Faced with the challenges of sample complexity brought by various glycoforms, some studies focused on a subset of total glycopeptides such as core-fucosylated peptides which could be enriched by highly specific binding afforded by lectin LcH [115-117]. Park *et al.* developed a novel automated Integrated GlycoProteome Analyzer (I-GPA) with FDR control for fast and confident intact N-glycopeptide identifications, and successfully identified 619 intact N-glycopeptides with an FDR below 1% from human serum [86]. Compared to N-glycosylation, O-glycosylation in serum is less studied mainly due to its lack of consensus motif and diversity of core structures. Recently, Zhang *et al.* have developed a systemic strategy that

combined multiple enzyme digestion, multidimensional separation and EThcD fragmentation, and identified 499 non-redundant intact O-glycopeptides in serum, covering singly, doubly and triply O-glycosylated peptides [70].

4.1.2 Cell culture

Other than human serum, the glycoproteome of different cell types have also been extensively explored. Cell culture has helped us gain valuable insights into various biological processes and disease-related pathological alterations, and has contributed enormously in drug discovery [118]. Adding glycoproteome data to the cellular models would help us gain a better understanding of the inherent complexity in biological systems [119]. Notably, the Levery group developed a robust SimpleCells approach for an O-GalNAc study, and successfully mapped human O-GalNAc glycoproteome with almost 3000 glycosites from over 600 O-glycoproteins in 12 human cell lines from different organs [54, 120]. Although SimpleCells approach has shown extraordinary performance in terms of O-glycosite mapping, it falls short in intact glycopeptide analysis due to glycan truncated during the process. To this end, Bertozzi group developed an IsoTaG strategy for intact glycopeptide characterization, and 1375 intact N-glycopeptides and 2159 intact O-glycopeptides were successfully identified from 15 human tissue-derived cell lines [121, 122]. Later, this approach was also applied for human T-cells O-GlcNAcylation analysis, with over 2000 O-GlcNAcylation peptides identified [123]. Some other studies focused on the glycoproteome on the cell surface, which are crucial in cell-cell communication and cell-environment interaction [124, 125]. In order to selectively capture surface glycoproteins, in 2009, Wollscheid *et al.* developed a powerful unbiased cell surface-capturing (CSC) technology through covalently labeling cell surface N-

glycan moieties [126]. Since then, this approach has been widely used for cell surface N-glycoproteome profiling including embryonic stem cells [127], induced pluripotent stem cells (iPSCs) [128], gastric adenocarcinoma cells [129], hepatocellular carcinoma cells [130], and hundreds of surface glycoproteins have been identified. Another rich source of glycoproteins come from the secreted proteins, or secretome, as many proteins undergo glycosylation prior to secretion [131]. For secreted glycoproteins analysis, conditioned media from serum-free cell culture is usually collected, followed by extraction of the secreted proteins, and then is subject to a typical glycoproteomics workflow. Li *et al.* have extensively conducted glycoproteome profiling of hepatocellular carcinoma cell lines and have mapped 1,213 unique N-glycosites from 611 N-glycoproteins [38, 132]. Cell component analysis revealed that these N-glycoproteins were primarily localized to the extracellular space and plasma membrane, indicating important role of N-glycosylation in the secretory pathway. The study of secreted glycoproteome of other commonly used cell lines such as human embryonic kidney (HEK) cells [133], Chinese hamster ovary cells (CHO) and endothelial cells [134], and some microorganisms such as green alga [135] and filamentous fungi [136] have also been conducted, which provide valuable insights into the secretory pathway and their responses to the environmental stimuli.

4.1.3 Animal tissues and plants

Due to their easy accessibility, the glycoproteome of animal tissues and plants have also been comprehensively profiled. Among them, mouse or rat brain is perhaps the most thoroughly explored tissue. By employing different enrichment strategies, including lectin, HILIC, hydrazide chemistry and TiO_2, Zhang *et al.* have successfully mapped 3446 unique glycosylation sites from

1597 N-glycoproteins in mouse brain, and 65% of the identified N-glycoproteins are membrane or extracellular proteins [22]. To take a step further, Fang *et al.* further increased the coverage by optimizing protease treatments and fractionation strategies and identified 8386 glycosylation sites on 3982 N-glycoproteins, representing the largest N-glycosylation site dataset in mouse brain ever reported [40]. Site-specific N-glycoproteome study in rat brain has also been conducted by utilizing a combined glycomics and glycoproteomics approach, resulting in the identifications of 863 unique intact N-glycopeptides [137]. The N-glycosylation sites mapping studies in other mouse/rat tissues such as liver, kidney, heart, plasma, stomach, ovary etc. revealed a tissue-specific expression pattern of N-glycosylation, indicating the close relation between glycosylation and the specialized function of different organs/tissues [37, 138, 139]. Compared with many number of glycoproteome studies in mammalian, the glycoproteome studies in plants are quite limited,despite an increased interest in deciphering the plants glycoproteome [140-142]. So far, hundreds of N-glycosites have been mapped in rice [143], cereal crop Brachypodium distachyon L. [113], tomato [17], flowering plant Arabidopsis [144] etc., providing valuable insights into the biological role of this ubiquitous protein modification in different plant species.

4.1.4 Microorganisms

As one of the most popular models for basic biological research, yeast has also gained plenty of researchers' interest in the glycoscience field. Breidenbach *et al.* started out mapping the N-glycosites in yeast, yielding a total of 133 N-glycosites spanning 58 glycoproteins, which were mainly distributed in the yeast ER, plasma membrane, vacuole, and cell wall [145]. It has been a puzzle for researchers that O-GlcNAcylation was found in all eukaryotic cells except yeast until

Halim *et al.* discovered and mapped nucleocytoplasmic O-mannose glycoproteome in yeast in 2015, which opened new avenues for O-glycosylation based biological events exploration in yeast [146]. Latter, Neubert *et al.* successfully mapped 2300 O-mannosylation sites from 500 O-glycoproteins from whole yeast cell, and one interesting finding was that these O-mannosylation sites were in the proximity of N-glycosylation sites, indicating their potential interplay [147]. The glycoproteome of some common bacteria [148-150] and viruses [151, 152] have also been mapped, affording a molecular foundation for further understanding glycosylation-assisted physiological processes.

4.2 Comparative MS-based glycoproteomics in complex samples

4.2.1 Disease biomarker discovery

Previously, glycosylation-based biomarkers studies relied on lectin staining or 2D gel electrophoresis approaches to measure the total glycosylation changes or the total glycoprotein changes, which suffer from low-throughput, limited sensitivity, and limited site-specific glycosylation information [153, 154]. With the advancement of glycoproteomics methodologies, glycosylation level changes can be pinpointed on a specific site and further microheterogeneity differences can be revealed through intact glycopeptide analysis in a high-throughput manner.

In a quantitative proteome and glycoproteome study of relapsing-remitting multiple sclerosis and neurological controls, Kroksveen *et al.* identified 96 altered deglycopeptides where their associated protein abundance was not affected, indicating the alterations were due to glycosylation occupancy changes instead of changes at the protein level [106]. Similarly, Shah and co-worker utilized an integrated proteomics and glycoproteomics approach to explore the mechanism of

castration resistance for androgen-deprivation therapy in prostate cancer [155]. This integrated omics approach not only allowed the detection of changes in glycosylation occupancy and microheterogeneity, but also identified associated altered fucosyltransferase and fucosidase expression. To take it a step further, after the initial finding of the increased terminal galactosylation and up-regulation of B4GalT5 galactosyltransferases upon TNF-Alpha-Induced insulin resistance in Adipocytes through an integrated proteomics and glycoproteomics approach, Parker *et al.* showed that the knockdown of B4GalT5 down-regulated the terminal galactosylation, confirming the involvement of B4GalT5 in the TNF-alpha-regulated N-glycome [94]. Instead of analyzing the whole glycoproteome, some other studies focused on a specific type of glycopeptides (fucosylated, sialylated etc.) to improve the coverage depth. Tan *et al.* employed an LCA enrichment approach to selectively enrich core-fucosylated glycopeptides, and were able to identify 613 core-fucosylated peptides and 8 of them exhibited a significant difference between pancreatic cancer and controls [156]. Due to close crosstalk between cells and the extracellular space, the secreted glycoproteins in extracellular space is another rich source for biomarker discovery. Li *et al.* conducted a glycoproteomics study in the secretome of human hepatocellular carcinoma metastatic (HCC) cell lines, and two glycoproteins FAT1 or DKK3 was proposed as novel prognostic biomarkers of HCC after validation with western blot and tissue array immunohistochemistry (IHC) [132]. Specifically, extracellular vesicles (EVs) in the secretory system been exploited as an attractive source for biomarker discovery. Very recently, Chen *et al.* identified 1,453 unique deglycosylated glycopeptides from 556 glycoproteins in plasma-derived EVs, among which 20 were verified significantly higher in breast cancer patients [157]. Additionally, 5 of these

glycoprotein candidates were later successfully validated in patient and healthy individuals through a novel polymer-based reverse phase glycoprotein array (polyGPA) platform.

4.2.2 Biological process exploration

With glycosylation playing a key role in many biological processes, comparative glycoproteomics could reveal the dynamic changes and further shed light upon its functions along these processes. In a recent study, Kang *et al.* employed the quantitative glycoproteomics approach to explore the molecular mechanism underlying the increased insulin secretion of normal pancreatic islet β-cells (PBCs) in response to elevated blood glucose levels [158]. Their results showed that altered sialylation of surface glycoproteins, such as integrins, integrin ligands, semaphorins and plexins was involved in the process of glucose-stimulated insulin secretion (GSIS). In order to uncover the glycomarkers in the neuronal differentiation process, Tyleckova *et al.* have successfully quantified hundreds of N-glycoproteins at onset and upon neuronal differentiation, as well as in mature hNT neurons using the cell surface capture (CSC) technology, and validated the glycosylation alterations of several cell adhesion glycoproteins using selected reaction monitoring (SRM) strategy [159]. Glycosylation has been known to affect the development of central nervous system (CNS) and defective glycosylation has also been shown to impair development and neurological function [160]. To this end, Palmisano *et al.* conducted a glycoproteomics study to monitor the glycosylation changes associated with cell signaling during mouse brain development using the postnatal mice from day 0 until maturity at day 80 [161]. Their results confirmed the role of sialylation in organ development and provided the first extensive global view of dynamic changes N-glycosylation during mouse brain development. A comprehensive N-glycoproteomics

was also conducted to investigate the role of N-glycosylation during the de-etiolation process, which is one of the most dramatic developmental processes known in plants [162]. The study has shown 186 N-glycosylation sites from 162 N-glycoproteins were significantly regulated over the course of the 12 hour de-etiolation period, indicating the important role of N-glycosylation during de-etiolation process. Besides the biological process without disturbance, the biological processes that are the result of environmental stimuli, such as infection, haves also been investigated by some studies. Braga *et al.* explored the modulation of N-glycosylation in grape by Lobesia botrana pathogen infection and demonstrated the importance of N-glycosylation in plant response to biotic stimulus through the glycosylation changes of disease-resistance response glycoprotein DDR206 [107]. In another study into regulation of protein N-glycosylation in human macrophages and their secreted microparticles (MPs) upon Mycobacterium tuberculosis infection, Hare *et al.* showed an increased complex-type glycosylation and concomitant down-regulation of paucimannosylation of macrophages upon infection [163].

5. Concluding remarks

With rapid advancements in various methodologies, including improved enrichment methods, novel MS/MS fragmentation techniques, powerful workflows, and advanced bioinformatics, MS-based glycoproteomics is gaining more attention and has been increasingly applied into different biological systems. Unprecedented glycoproteome depth has been achieved in different complex samples, providing valuable molecular basis for further structure-function study of glycosylation. Due to the advances in hybrid fragmentation methods and maturing search engines for intact glycopeptide analysis, site-specific glycoproteomics will eventually become a routine and practical

approach for large-scale glycosylation analysis, which could help decipher the long-time puzzle of glycosylation microheterogeneity. Furthermore, quantitative glycoproteomics could reveal the alterations in glycosylation microheterogeneity, providing explicit molecule targets for follow-up studies in different physiological and pathological processes.

References:

1. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.
2. Jankovic, M. M.; Milutinovic, B. S., Glycoforms of CA125 antigen as a possible cancer marker. *Cancer Biomarkers* **2008,** 4, (1), 35-42.
3. van Gisbergen, K. P.; Aarnoudse, C. A.; Meijer, G. A.; Geijtenbeek, T. B.; van Kooyk, Y., Dendritic cells recognize tumor-specific glycosylation of carcinoembryonic antigen on colorectal cancer cells through dendritic cell–specific intercellular adhesion molecule-3–grabbing nonintegrin. *Cancer Research* **2005,** 65, (13), 5935-5944.
4. Noda, K.; Miyoshi, E.; Uozumi, N.; Yanagidani, S.; Ikeda, Y.; Gao, C. x.; Suzuki, K.; Yoshihara, H.; Yoshikawa, M.; Kawano, K., Gene expression of α1-6 fucosyltransferase in human hepatoma tissues: A possible implication for increased fucosylation of α-fetoprotein. *Hepatology* **1998,** 28, (4), 944-952.
5. Drake, P. M.; Cho, W.; Li, B.; Prakobphol, A.; Johansen, E.; Anderson, N. L.; Regnier, F. E.; Gibson, B. W.; Fisher, S. J., Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clinical chemistry* **2010,** 56, (2), 223-236.
6. Huang, G.; Sun, Z.; Qin, H.; Zhao, L.; Xiong, Z.; Peng, X.; Ou, J.; Zou, H., Preparation of hydrazine functionalized polymer brushes hybrid magnetic nanoparticles for highly specific enrichment of glycopeptides. *Analyst* **2014,** 139, (9), 2199-2206.
7. 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. *Nature Biotechnology* **2003,** 21, (6), 660-666.
8. Sun, B.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X.; Lin, B.; Hood, L., Shotgun Glycopeptide Capture Approach Coupled with Mass Spectrometry for Comprehensive Glycoproteomics. *Molecular & Cellular Proteomics* **2007,** 6, (1), 141-149.
9. Kurogochi, M.; Amano, M.; Fumoto, M.; Takimoto, A.; Kondo, H.; Nishimura, S.-I., Reverse Glycoblotting Allows Rapid-Enrichment Glycoproteomics of Biopharmaceuticals and Disease-Related Biomarkers. *Angewandte Chemie International Edition* **2007,** 46, (46), 8808-8813.
10. Nilsson, J.; Ruetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G., Enrichment of glycopeptides for glycan structure and attachment site identification. *Nature methods* **2009,** 6, (11), 809-11.
11. Halim, A.; Rüetschi, U.; Larson, G.; Nilsson, J., LC–MS/MS Characterization of O-Glycosylation Sites and Glycan Structures of Human Cerebrospinal Fluid Glycoproteins. *Journal of Proteome Research* **2013,** 12, (2), 573-584.
12. Kurogochi, M.; Matsushista, T.; Amano, M.; Furukawa, J.-i.; Shinohara, Y.; Aoshima, M.; Nishimura, S.-I., Sialic Acid-focused Quantitative Mouse Serum Glycoproteomics by Multiple Reaction Monitoring Assay. *Molecular & Cellular Proteomics* **2010,** 9, (11), 2354-2368.
13. Badr, H. A.; AlSadek, D. M. M.; Darwish, A. A.; ElSayed, A. I.; Bekmanov, B. O.;

Khussainova, E. M.; Zhang, X.; Cho, W. C. S.; Djansugurova, L. B.; Li, C.-Z., Lectin approaches for glycoproteomics in FDA-approved cancer biomarkers. *Expert Review of Proteomics* **2014,** 11, (2), 227-236.
14. Ruiz-May, E.; Catala, C.; Rose, J. K., N-glycoprotein enrichment by lectin affinity chromatography. *Methods Mol Biol* **2014,** 1072, 633-43.
15. Ruiz-May, E.; Hucko, S.; Howe, K. J.; Zhang, S.; Sherwood, R. W.; Thannhauser, T. W.; Rose, J. K. C., A Comparative Study of Lectin Affinity Based Plant N-Glycoproteome Profiling Using Tomato Fruit as a Model. *Molecular & Cellular Proteomics* **2014,** 13, (2), 566-579.
16. Totten, S. M.; Kullolli, M.; Pitteri, S. J., Multi-Lectin Affinity Chromatography for Separation, Identification, and Quantitation of Intact Protein Glycoforms in Complex Biological Mixtures. In *Proteomics: Methods and Protocols*, Comai, L.; Katz, J. E.; Mallick, P., Eds. Springer New York: New York, NY, 2017; pp 99-113.
17. Gbormittah, F. O.; Lee, L. Y.; Taylor, K.; Hancock, W. S.; Iliopoulos, O., Comparative Studies of the Proteome, Glycoproteome, and N-Glycome of Clear Cell Renal Cell Carcinoma Plasma before and after Curative Nephrectomy. *Journal of Proteome Research* **2014,** 13, (11), 4889-4900.
18. Li, Y.; Shah, P.; De Marzo, A. M.; Van Eyk, J. E.; Li, Q.; Chan, D. W.; Zhang, H., Identification of Glycoproteins Containing Specific Glycans Using a Lectin-Chemical Method. *Analytical Chemistry* **2015,** 87, (9), 4683-4687.
19. Yang, J. S.; Qiao, J.; Kim, J. Y.; Zhao, L.; Qi, L.; Moon, M. H., Online Proteolysis and Glycopeptide Enrichment with Thermoresponsive Porous Polymer Membrane Reactors for Nanoflow Liquid Chromatography-Tandem Mass Spectrometry. *Analytical Chemistry* **2018,** 90, (5), 3124-3131.
20. Hägglund, 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. *Journal of Proteome Research* **2004,** 3, (3), 556-566.
21. Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M., Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Analytical Chemistry* **2010,** 82, (13), 5598-5609.
22. Zhang, C.; Ye, Z.; Xue, P.; Shu, Q.; Zhou, Y.; Ji, Y.; Fu, Y.; Wang, J.; Yang, F., Evaluation of different N-glycopeptide enrichment methods for N-glycosylation sites mapping in mouse brain. *Journal of proteome research* **2016,** 15, (9), 2960-2968.
23. Xiong, Z.; Zhao, L.; Wang, F.; Zhu, J.; Qin, H.; Wu, R. a.; Zhang, W.; Zou, H., Synthesis of branched PEG brushes hybrid hydrophilic magnetic nanoparticles for the selective enrichment of N-linked glycopeptides. *Chemical Communications* **2012,** 48, (65), 8138-8140.
24. Xiong, Z.; Qin, H.; Wan, H.; Huang, G.; Zhang, Z.; Dong, J.; Zhang, L.; Zhang, W.; Zou, H., Layer-by-layer assembly of multilayer polysaccharide coated magnetic nanoparticles for the selective enrichment of glycopeptides. *Chemical Communications* **2013,** 49, (81), 9284-9286.
25. Huang, G.; Xiong, Z.; Qin, H.; Zhu, J.; Sun, Z.; Zhang, Y.; Peng, X.; ou, J.; Zou, H., Synthesis of zwitterionic polymer brushes hybrid silica nanoparticles via controlled polymerization for highly efficient enrichment of glycopeptides. *Analytica Chimica Acta* **2014,** 809, 61-68.

26. Jiao, F.; Gao, F.; Wang, H.; Deng, Y.; Zhang, Y.; Qian, X.; Zhang, Y., Polymeric hydrophilic ionic liquids used to modify magnetic nanoparticles for the highly selective enrichment of N-linked glycopeptides. *Scientific Reports* **2017,** 7, (1), 6984.
27. Sugahara, D.; Kaji, H.; Sugihara, K.; Asano, M.; Narimatsu, H., Large-scale identification of target proteins of a glycosyltransferase isozyme by Lectin-IGOT-LC/MS, an LC/MS-based glycoproteomic approach. *Sci Rep* **2012,** 2, 680.
28. Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L., Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD). *Analyst* **2018**.
29. Selman, M. H. J.; Hemayatkar, M.; Deelder, A. M.; Wuhrer, M., Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Analytical Chemistry* **2011,** 83, (7), 2492-2499.
30. Zhu, J.; Wang, F.; Chen, R.; Cheng, K.; Xu, B.; Guo, Z.; Liang, X.; Ye, M.; Zou, H., Centrifugation Assisted Microreactor Enables Facile Integration of Trypsin Digestion, Hydrophilic Interaction Chromatography Enrichment, and On-Column Deglycosylation for Rapid and Sensitive N-Glycoproteome Analysis. *Analytical Chemistry* **2012,** 84, (11), 5146-5153.
31. Liu, J.; Wang, F.; Lin, H.; Zhu, J.; Bian, Y.; Cheng, K.; Zou, H., Monolithic Capillary Column Based Glycoproteomic Reactor for High-Sensitive Analysis of N-Glycoproteome. *Analytical Chemistry* **2013,** 85, (5), 2847-2852.
32. Dedvisitsakul, P.; Jacobsen, S.; Svensson, B.; Bunkenborg, J.; Finnie, C.; Hägglund, P., Glycopeptide Enrichment Using a Combination of ZIC-HILIC and Cotton Wool for Exploring the Glycoproteome of Wheat Flour Albumins. *Journal of Proteome Research* **2014,** 13, (5), 2696-2703.
33. Jiang, H.; Yuan, H.; Qu, Y.; Liang, Y.; Jiang, B.; Wu, Q.; Deng, N.; Liang, Z.; Zhang, L.; Zhang, Y., Preparation of hydrophilic monolithic capillary column by in situ photo-polymerization of N-vinyl-2-pyrrolidinone and acrylamide for highly selective and sensitive enrichment of N-linked glycopeptides. *Talanta* **2016,** 146, 225-230.
34. Huang, J. F.; Qin, H. Q.; Sun, Z.; Huang, G.; Mao, J. W.; Cheng, K.; Zhang, Z.; Wan, H.; Yao, Y. T.; Dong, J.; Zhu, J.; Wang, F. J.; Ye, M. L.; Zou, H. F., A peptide N-terminal protection strategy for comprehensive glycoproteome analysis using hydrazide chemistry based method. *Scientific Reports* **2015,** 5, 10164.
35. Huang, J.; Wan, H.; Yao, Y.; Li, J.; Cheng, K.; Mao, J.; Chen, J.; Wang, Y.; Qin, H.; Zhang, W.; Ye, M.; Zou, H., Highly Efficient Release of Glycopeptides from Hydrazide Beads by Hydroxylamine Assisted PNGase F Deglycosylation for N-Glycoproteome Analysis. *Analytical Chemistry* **2015,** 87, (20), 10199-10204.
36. Weng, Y.; Sui, Z.; Jiang, H.; Shan, Y.; Chen, L.; Zhang, S.; Zhang, L.; Zhang, Y., Releasing N-glycan from Peptide N-terminus by N-terminal Succinylation Assisted Enzymatic Deglycosylation. *Sci Rep* **2015,** 5, 9770.
37. Zielinska, D. F.; Gnad, F.; Wiśniewski, J. R.; Mann, M., Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **2010,** 141, (5), 897-

907.
38. Li, X.; Jiang, J.; Zhao, X.; Wang, J.; Han, H.; Zhao, Y.; Peng, B.; Zhong, R.; Ying, W.; Qian, X., N-glycoproteome analysis of the secretome of human metastatic hepatocellular carcinoma cell lines combining hydrazide chemistry, HILIC enrichment and mass spectrometry. *PloS one* **2013,** 8, (12), e81921.
39. Zhu, J.; Sun, Z.; Cheng, K.; Chen, R.; Ye, M.; Xu, B.; Sun, D.; Wang, L.; Liu, J.; Wang, F.; Zou, H., Comprehensive Mapping of Protein N-Glycosylation in Human Liver by Combining Hydrophilic Interaction Chromatography and Hydrazide Chemistry. *Journal of Proteome Research* **2014,** 13, (3), 1713-1721.
40. Fang, P.; Wang, X.-j.; Xue, Y.; Liu, M.-q.; Zeng, W.-f.; Zhang, Y.; Zhang, L.; Gao, X.; Yan, G.-q.; Yao, J.; Shen, H.-l.; Yang, P.-y., In-depth mapping of the mouse brain N-glycoproteome reveals widespread N-glycosylation of diverse brain proteins. *Oncotarget* **2016,** 7, (25), 38796-38809.
41. Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y., Advances in mass spectrometry driven O-glycoproteomics. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015,** 1850, (1), 33-42.
42. Ma, J.; Hart, G. W., Analysis of Protein O-GlcNAcylation by Mass Spectrometry. *Current Protocols in Protein Science* **2017,** 87, (1), 24.10.1-24.10.16.
43. Worth, M.; Li, H.; Jiang, J., Deciphering the Functions of Protein O-GlcNAcylation with Chemistry. *ACS Chemical Biology* **2017,** 12, (2), 326-335.
44. Khidekel, N.; Ficarro, S. B.; Clark, P. M.; Bryan, M. C.; Swaney, D. L.; Rexach, J. E.; Sun, Y. E.; Coon, J. J.; Peters, E. C.; Hsieh-Wilson, L. C., Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **2007,** 3, (6), 339-348.
45. 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. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** 108, (20), 8146-51.
46. Parker, B. L.; Gupta, P.; Cordwell, S. J.; Larsen, M. R.; Palmisano, G., Purification and Identification of O-GlcNAc-Modified Peptides Using Phosphate-Based Alkyne CLICK Chemistry in Combination with Titanium Dioxide Chromatography and Mass Spectrometry. *Journal of Proteome Research* **2010,** 10, (4), 1449-1458.
47. Alfaro, J. F.; Gong, C. X.; Monroe, M. E.; Aldrich, J. T.; Clauss, T. R. W.; Purvine, S. O.; Wang, Z. H.; Camp, D. G.; Shabanowitz, J.; Stanley, P.; Hart, G. W.; Hunt, D. F.; Yang, F.; Smith, R. D., Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** 109, (19), 7280-7285.
48. Thompson, J. W.; Griffin, M. E.; Hsieh-Wilson, L. C., Chapter Four - Methods for the Detection, Study, and Dynamic Profiling of O-GlcNAc Glycosylation. In *Methods in Enzymology*, Imperiali, B., Ed. Academic Press: 2018; Vol. 598, pp 101-135.
49. Vosseller, K.; Trinidad, J. C.; Chalkley, R. J.; Specht, C. G.; Thalhammer, A.; Lynn, A. J.;

Snedecor, J. O.; Guan, S.; Medzihradszky, K. F.; Maltby, D. A.; Schoepfer, R.; Burlingame, A. L., O-Linked N-Acetylglucosamine Proteomics of Postsynaptic Density Preparations Using Lectin Weak Affinity Chromatography and Mass Spectrometry. *Molecular & Cellular Proteomics* **2006,** 5, (5), 923-934.
50. Trinidad, J. C.; Barkan, D. T.; Gulledge, B. F.; Thalhammer, A.; Sali, A.; Schoepfer, R.; Burlingame, A. L., Global Identification and Characterization of Both O-GlcNAcylation and Phosphorylation at the Murine Synapse. *Molecular & Cellular Proteomics* **2012,** 11, (8), 215-229.
51. Shen, B.; Zhang, W.; Shi, Z.; Tian, F.; Deng, Y.; Sun, C.; Wang, G.; Qin, W.; Qian, X., A novel strategy for global mapping of O-GlcNAc proteins and peptides using selective enzymatic deglycosylation, HILIC enrichment and mass spectrometry identification. *Talanta* **2017,** 169, (Supplement C), 195-202.
52. Darula, Z.; Sherman, J.; Medzihradszky, K. F., How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Molecular & cellular proteomics : MCP* **2012,** 11, (7), O111 016774.
53. Steentoft, C.; Vakhrushev, S. Y.; Vester-Christensen, M. B.; Schjoldager, K. T. B. G.; Kong, Y.; Bennett, E. P.; Mandel, U.; Wandall, H.; Levery, S. B.; Clausen, H., Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* **2011,** 8, (11), 977-982.
54. Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T. B. G.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; Marcos-Silva, L.; Gupta, R.; Paul Bennett, E.; Mandel, U.; Brunak, S.; Wandall, H. H.; Levery, S. B.; Clausen, H., Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **2013,** 32, (10), 1478-1488.
55. Hu, H.; Khatri, K.; Klein, J.; Leymarie, N.; Zaia, J., A review of methods for interpretation of glycopeptide tandem mass spectral data. *Glycoconjugate Journal* **2016,** 33, (3), 285-296.
56. 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. *Molecular & Cellular Proteomics* **2016**.
57. Halim, A.; Nilsson, J.; Rüetschi, U.; Hesse, C.; Larson, G., Human Urinary Glycoproteomics; Attachment Site Specific Analysis of N- and O-Linked Glycosylations by CID and ECD. *Molecular & Cellular Proteomics* **2012,** 11, (4).
58. Mayampurath, A.; Yu, C.-Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H., Computational Framework for Identification of Intact Glycopeptides in Complex Samples. *Analytical Chemistry* **2014,** 86, (1), 453-463.
59. Chen, R.; Seebun, D.; Ye, M.; Zou, H.; Figeys, D., Site-specific characterization of cell membrane N-glycosylation with integrated hydrophilic interaction chromatography solid phase extraction and LC–MS/MS. *Journal of Proteomics* **2014,** 103, (0), 194-203.
60. Cheng, K.; Chen, R.; Seebun, D.; Ye, M.; Figeys, D.; Zou, H., Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *Journal of Proteomics* **2014,** 110, (0), 145-154.
61. Liu, M.; Zhang, Y.; Chen, Y.; Yan, G.; Shen, C.; Cao, J.; Zhou, X.; Liu, X.; Zhang, L.; Shen,

H.; Lu, H.; He, F.; Yang, P., Efficient and Accurate Glycopeptide Identification Pipeline for High-Throughput Site-Specific N-Glycosylation Analysis. *Journal of Proteome Research* **2014,** 13, (6), 3121-3129.
62. Cao, Q.; Zhao, X.; Zhao, Q.; Lv, X.; Ma, C.; Li, X.; Zhao, Y.; Peng, B.; Ying, W.; Qian, X., Strategy Integrating Stepped Fragmentation and Glycan Diagnostic Ion-Based Spectrum Refinement for the Identification of Core Fucosylated Glycoproteome Using Mass Spectrometry. *Analytical Chemistry* **2014,** 86, (14), 6804-6811.
63. Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; Tu, H. J.; Chi, H.; Sun, R. X.; Cao, Y.; Dong, M. Q.; Jiang, B. Y.; Huang, J. M.; Shen, H. L.; Wong, C. C. L.; He, S. M.; Yang, P. Y., pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun* **2017,** 8, (1), 438.
64. Bollineni, R. C.; Koehler, C. J.; Gislefoss, R. E.; Anonsen, J. H.; Thiede, B., Large-scale intact glycopeptide identification by Mascot database search. *Scientific Reports* **2018,** 8, (1), 2117.
65. Khatri, K.; Staples, G. O.; Leymarie, N.; Leon, D. R.; Turiák, L.; Huang, Y.; Yip, S.; Hu, H.; Heckendorf, C. F.; Zaia, J., Confident Assignment of Site-Specific Glycosylation in Complex Glycoproteins in a Single Step. *Journal of Proteome Research* **2014,** 13, (10), 4347-4355.
66. Lynn, K.-S.; Chen, C.-C.; Lih, T. M.; Cheng, C.-W.; Su, W.-C.; Chang, C.-H.; Cheng, C.-Y.; Hsu, W.-L.; Chen, Y.-J.; Sung, T.-Y., MAGIC: An Automated N-Linked Glycoprotein Identification Tool Using a Y1-Ion Pattern Matching Algorithm and in Silico MS2 Approach. *Analytical Chemistry* **2015,** 87, (4), 2466-2473.
67. Qin, H.; Cheng, K.; Zhu, J.; Mao, J.; Wang, F.; Dong, M.; Chen, R.; Guo, Z.; Liang, X.; Ye, M.; Zou, H., Proteomics Analysis of O-GalNAc Glycosylation in Human Serum by an Integrated Strategy. *Analytical Chemistry* **2017,** 89, (3), 1469-1476.
68. Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S., Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **2012,** 84, (22), 9668-9673.
69. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.-W.; Kent, K. C.; Li, L., Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *Journal of The American Society for Mass Spectrometry* **2017,** 28, (9), 1751-1764.
70. Zhang, Y.; Xie, X.; Zhao, X.; Tian, F.; Lv, J.; Ying, W.; Qian, X., Systems analysis of singly and multiply O-glycosylated peptides in the human serum glycoproteome via EThcD and HCD mass spectrometry. *Journal of proteomics* **2018,** 170, 14-27.
71. He, L.; Xin, L.; Shan, B.; Lajoie, G. A.; Ma, B., GlycoMaster DB: Software To Assist the Automated Identification of N-Linked Glycopeptides by Tandem Mass Spectrometry. *Journal of Proteome Research* **2014,** 13, (9), 3881-3895.
72. Toghi Eshghi, S.; Shah, P.; Yang, W.; Li, X.; Zhang, H., GPQuest: A Spectral Library Matching Algorithm for Site-Specific Assignment of Tandem Mass Spectra to Intact N-glycopeptides. *Analytical Chemistry* **2015,** 87, (10), 5181-5188.

73. Park, G. W.; Kim, J. Y.; Hwang, H.; Lee, J. Y.; Ahn, Y. H.; Lee, H. K.; Ji, E. S.; Kim, K. H.; Jeong, H. K.; Yun, K. N.; Kim, Y.-S.; Ko, J.-H.; An, H. J.; Kim, J. H.; Paik, Y.-K.; Yoo, J. S., Integrated GlycoProteome Analyzer (I-GPA) for Automated Identification and Quantitation of Site-Specific N-Glycosylation. *Scientific Reports* **2016,** 6, 21175.
74. Bern, M.; Kil, Y. J.; Becker, C., Byonic: Advanced Peptide and Protein Identification Software. *Current Protocols in Bioinformatics* **2012,** 40, (1), 13.20.1-13.20.14.
75. Nasir, W.; Toledo, A. G.; Noborn, F.; Nilsson, J.; Wang, M.; Bandeira, N.; Larson, G., SweetNET: A Bioinformatics Workflow for Glycopeptide MS/MS Spectral Analysis. *Journal of Proteome Research* **2016,** 15, (8), 2826-2840.
76. Zeng, W.-F.; Liu, M.-Q.; Zhang, Y.; Wu, J.-Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; Chi, H.; Sun, R.-X.; Wong, C. C. L.; He, S.-M.; Yang, P., pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Scientific Reports* **2016,** 6, 25102.
77. Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattinger, A.; Dürnberger, G.; Dusberger, F.; Elling, U.; Mach, L.; Mechtler, K.; Penninger, J. M., Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **2017,** 549, 538.
78. Lakbub, J. C.; Su, X.; Zhu, Z.; Patabandige, M. W.; Hua, D.; Go, E. P.; Desaire, H., Two New Tools for Glycopeptide Analysis Researchers: A Glycopeptide Decoy Generator and a Large Data Set of Assigned CID Spectra of Glycopeptides. *Journal of Proteome Research* **2017,** 16, (8), 3002-3008.
79. Lee, L. Y.; Moh, E. S. X.; Parker, B. L.; Bern, M.; Packer, N. H.; Thaysen-Andersen, M., Toward Automated N-Glycopeptide Identification in Glycoproteomics. *Journal of Proteome Research* **2016,** 15, (10), 3904-3915.
80. Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C., Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *Journal of proteome research* **2012,** 11, (3), 1582-1590.
81. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & cellular proteomics* **2005,** 4, (10), 1487-1502.
82. Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W.-j.; Webb-Robertson, B.-J. M.; Smith, R. D.; Lipton, M. S., Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *Journal of proteome research* **2006,** 5, (2), 277-286.
83. Rebecchi, K. R.; Wenke, J. L.; Go, E. P.; Desaire, H., Label-free quantitation: a new glycoproteomics approach. *Journal of the American Society for Mass Spectrometry* **2009,** 20, (6), 1048-1059.
84. Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; Van Sluyter, S. C.; Haynes, P. A., Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **2011,** 11, (4), 535-553.
85. Mayampurath, A.; Song, E.; Mathur, A.; Yu, C.-y.; Hammoud, Z.; Mechref, Y.; Tang, H.,

Label-free glycopeptide quantification for biomarker discovery in human sera. *Journal of proteome research* **2014,** 13, (11), 4821-4832.
86. Park, G. W.; Kim, J. Y.; Hwang, H.; Lee, J. Y.; Ahn, Y. H.; Lee, H. K.; Ji, E. S.; Kim, K. H.; Jeong, H. K.; Yun, K. N., Integrated GlycoProteome Analyzer (I-GPA) for automated identification and quantitation of site-specific N-glycosylation. *Scientific reports* **2016,** 6, 21175.
87. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics* **2002,** 1, (5), 376-386.
88. Abarca, M.; Bragulat, M.; Castellá, G.; Cabañes, F.; Bacha, H.; Maaroufi, K.; Achour, A.; Hammami, M.; Belli, N.; Pardo, E., Functional and quantitative proteomics using SILAC in cancer research. *Junta Directiva* **1994,** 60, 31.
89. Ji, Y.; Wei, S.; Hou, J.; Zhang, C.; Xue, P.; Wang, J.; Chen, X.; Guo, X.; Yang, F., Integrated proteomic and N-glycoproteomic analyses of doxorubicin sensitive and resistant ovarian cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. *Oncotarget* **2017,** 8, (8), 13413.
90. Yang, X.; Wang, Z.; Guo, L.; Zhu, Z.; Zhang, Y., Proteome-wide analysis of N-glycosylation stoichiometry using SWATH technology. *Journal of proteome research* **2017,** 16, (10), 3830-3840.
91. Smeekens, J. M.; Xiao, H.; Wu, R., Global Analysis of Secreted Proteins and Glycoproteins in Saccharomyces cerevisiae. *Journal of proteome research* **2016,** 16, (2), 1039-1049.
92. Geiger, T.; Cox, J.; Ostasiewicz, P.; Wisniewski, J. R.; Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* **2010,** 7, (5), 383.
93. Deeb, S. J.; Cox, J.; Schmidt-Supprian, M.; Mann, M., N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Molecular & cellular proteomics* **2014,** 13, (1), 240-251.
94. Parker, B. L.; Thaysen-Andersen, M.; Fazakerley, D. J.; Holliday, M.; Packer, N. H.; James, D. E., Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-alpha-induced insulin resistance in adipocytes. *Molecular & Cellular Proteomics* **2016,** 15, (1), 141-153.
95. Taga, Y.; Kusubata, M.; Ogawa-Goto, K.; Hattori, S., Site-specific quantitative analysis of overglycosylation of collagen in osteogenesis imperfecta using hydrazide chemistry and SILAC. *Journal of proteome research* **2013,** 12, (5), 2225-2232.
96. Tolonen, A. C.; Haas, W.; Chilaka, A. C.; Aach, J.; Gygi, S. P.; Church, G. M., Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Molecular systems biology* **2011,** 7, (1), 461.
97. 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. *Nature biotechnology* **2003,** 21, (6), 660.
98. Wang, S.; Regnier, F. E., Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography. *Journal of Chromatography A* **2001,** 924, (1-2), 345-357.

99. Sun, Z.; Qin, H.; Wang, F.; Cheng, K.; Dong, M.; Ye, M.; Zou, H., Capture and dimethyl labeling of glycopeptides on hydrazide beads for quantitative glycoproteomics analysis. *Analytical chemistry* **2012,** 84, (20), 8452-8456.
100. Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry* **2003,** 75, (8), 1895-1904.
101. Werner, T.; Sweetman, G.; Savitski, M. F. l.; Mathieson, T.; Bantscheff, M.; Savitski, M. M., Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Analytical chemistry* **2014,** 86, (7), 3594-3601.
102. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics* **2004,** 3, (12), 1154-1169.
103. Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H., 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **2007,** 7, (20), 3651-3660.
104. Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L., N, N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. *Analytical chemistry* **2010,** 82, (7), 2817-2825.
105. Frost, D. C.; Greer, T.; Li, L., High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Analytical chemistry* **2014,** 87, (3), 1646-1654.
106. Kroksveen, A. C.; Guldbrandsen, A.; Vaudel, M.; Lereim, R. R.; Barsnes, H.; Myhr, K.-M.; Torkildsen, Ø.; Berven, F. S., In-depth cerebrospinal fluid quantitative proteome and deglycoproteome analysis: Presenting a comprehensive picture of pathways and processes affected by multiple sclerosis. *Journal of proteome research* **2016,** 16, (1), 179-194.
107. Melo-Braga, M. N.; Verano-Braga, T.; León, I. R.; Antonacci, D.; Nogueira, F. C.; Thelen, J. J.; Larsen, M. R.; Palmisano, G., Modulation of protein phosphorylation, N-glycosylation and Lys-acetylation in grape (Vitis vinifera) mesocarp and exocarp owing to Lobesia botrana infection. *Molecular & Cellular Proteomics* **2012,** 11, (10), 945-956.
108. López-Ferrer, D.; Ramos-Fernández, A.; Martínez-Bartolomé, S.; García-Ruiz, P.; Vázquez, J., Quantitative proteomics using 16O/18O labeling and linear ion trap mass spectrometry. *Proteomics* **2006,** 6, (S1).
109. Liu, Z.; Cao, J.; He, Y.; Qiao, L.; Xu, C.; Lu, H.; Yang, P., Tandem 18O stable isotope labeling for quantification of N-glycoproteome. *Journal of proteome research* **2009,** 9, (1), 227-236.
110. Yang, Z.; Hancock, W. S., Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *Journal of chromatography A* **2004,** 1053, (1-2), 79-88.

111. Sparbier, K.; Asperger, A.; Resemann, A.; Kessler, I.; Koch, S.; Wenzel, T.; Stein, G.; Vorwerg, L.; Suckau, D.; Kostrzewa, M., Analysis of glycoproteins in human serum by means of glycospecific magnetic bead separation and LC-MALDI-TOF/TOF analysis with automated glycopeptide detection. *Journal of biomolecular techniques: JBT* **2007,** 18, (4), 252.
112. Anderson, N. L.; Anderson, N. G., The human plasma proteome history, character, and diagnostic prospects. *Molecular & cellular proteomics* **2002,** 1, (11), 845-867.
113. Zhang, M.; Chen, G.-X.; Lv, D.-W.; Li, X.-H.; Yan, Y.-M., N-linked glycoproteome profiling of seedling leaf in Brachypodium distachyon L. *Journal of proteome research* **2015,** 14, (4), 1727-1738.
114. Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D., Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *Journal of proteome research* **2005,** 4, (6), 2070-2080.
115. Ma, C.; Zhang, Q.; Qu, J.; Zhao, X.; Li, X.; Liu, Y.; Wang, P. G., A precise approach in large scale core-fucosylated glycoprotein identification with low-and high-normalized collision energy. *Journal of proteomics* **2015,** 114, 61-70.
116. Hägglund, P.; Matthiesen, R.; Elortza, F.; Højrup, P.; Roepstorff, P.; Jensen, O. N.; Bunkenborg, J., An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *Journal of proteome research* **2007,** 6, (8), 3021-3031.
117. Jia, W.; Lu, Z.; Fu, Y.; Wang, H.-P.; Wang, L.-H.; Chi, H.; Yuan, Z.-F.; Zheng, Z.-B.; Song, L.-N.; Han, H.-H., A strategy for precise and large scale identification of core fucosylated glycoproteins. *Molecular & Cellular Proteomics* **2009,** 8, (5), 913-923.
118. Zhang, D.; Luo, G.; Ding, X.; Lu, C., Preclinical experimental models of drug metabolism and disposition in drug discovery and development. *Acta Pharmaceutica Sinica B* **2012,** 2, (6), 549-561.
119. Breker, M.; Schuldiner, M., The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. *Nature Reviews Molecular Cell Biology* **2014,** 15, (7), 453.
120. Vakhrushev, S. Y.; Steentoft, C.; Vester-Christensen, M. B.; Bennett, E. P.; Clausen, H.; Levery, S. B., Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Molecular & Cellular Proteomics* **2013,** 12, (4), 932-944.
121. 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* **2015,** 12, (6), 561.
122. Woo, C. M.; Felix, A.; Byrd, W. E.; Zuegel, D. K.; Ishihara, M.; Azadi, P.; Iavarone, A. T.; Pitteri, S. J.; Bertozzi, C. R., Development of IsoTaG, a chemical glycoproteomics technique for profiling intact N-and O-glycopeptides from whole cell proteomes. *Journal of proteome research* **2017,** 16, (4), 1706-1718.
123. Woo, C. M.; Lund, P. J.; Huang, A. C.; Davis, M. M.; Bertozzi, C. R.; Pitteri, S., Mapping and quantification of over 2,000 O-linked glycopeptides in activated human T cells with isotope-targeted glycoproteomics (IsoTaG). *Molecular & Cellular Proteomics* **2018**, mcp. RA117. 000261.
124. Von Heijne, G., The membrane protein universe: what's out there and why bother?

Journal of internal medicine **2007,** 261, (6), 543-557.
125. Wollscheid, B.; von Haller, P. D.; Yi, E.; Donohoe, S.; Vaughn, K.; Keller, A.; Nesvizhskii, A. I.; Eng, J.; Li, X.-j.; Goodlett, D. R., Lipid raft proteins and their identification in T lymphocytes. In *Membrane Dynamics and Domains*, Springer: 2004; pp 121-152.
126. Wollscheid, B.; Bausch-Fluck, D.; Henderson, C.; O'brien, R.; Bibel, M.; Schiess, R.; Aebersold, R.; Watts, J. D., Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature biotechnology* **2009,** 27, (4), 378.
127. Gundry, R. L.; Riordon, D. R.; Tarasova, Y.; Chuppa, S.; Bhattacharya, S.; Juhasz, O.; Wiedemeier, O.; Milanovich, S.; Noto, F. K.; Tchernyshyov, I., A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Molecular & Cellular Proteomics* **2012,** 11, (8), 303-316.
128. Mallanna, S. K.; Cayo, M. A.; Twaroski, K.; Gundry, R. L.; Duncan, S. A., Mapping the cell-surface N-glycoproteome of human hepatocytes reveals markers for selecting a homogeneous population of iPSC-derived hepatocytes. *Stem cell reports* **2016,** 7, (3), 543-556.
129. Li, K.; Sun, Z.; Zheng, J.; Lu, Y.; Bian, Y.; Ye, M.; Wang, X.; Nie, Y.; Zou, H.; Fan, D., In-depth research of multidrug resistance related cell surface glycoproteome in gastric cancer. *Journal of proteomics* **2013,** 82, 130-140.
130. Mi, W.; Jia, W.; Zheng, Z.; Wang, J.; Cai, Y.; Ying, W.; Qian, X., Surface glycoproteomic analysis of hepatocellular carcinoma cells by affinity enrichment and mass spectrometric identification. *Glycoconjugate journal* **2012,** 29, (5-6), 411-424.
131. Tjalsma, H.; Bolhuis, A.; Jongbloed, J. D.; Bron, S.; van Dijl, J. M., Signal peptide-dependent protein transport inBacillus subtilis: a genome-based survey of the secretome. *Microbiology and Molecular Biology Reviews* **2000,** 64, (3), 515-547.
132. Li, X.; Jiang, J.; Zhao, X.; Zhao, Y.; Cao, Q.; Zhao, Q.; Han, H.; Wang, J.; Yu, Z.; Peng, B., In-depth analysis of secretome and N-glycosecretome of human hepatocellular carcinoma metastatic cell lines shed light on metastasis correlated proteins. *Oncotarget* **2016,** 7, (16), 22031.
133. Sugahara, D.; Tomioka, A.; Sato, T.; Narimatsu, H.; Kaji, H., Large-scale identification of secretome glycoproteins recognized by Wisteria floribunda agglutinin: A glycoproteomic approach to biomarker discovery. *Proteomics* **2015,** 15, (17), 2921-2933.
134. Yin, X.; Bern, M.; Xing, Q.; Ho, J.; Viner, R.; Mayr, M., Glycoproteomic analysis of the secretome of human endothelial cells. *Molecular & Cellular Proteomics* **2013**, mcp. M112. 024018.
135. Ruiz-May, E.; Sørensen, I.; Fei, Z.; Zhang, S.; Domozych, D. S.; Rose, J. K., The Secretome and N-Glycosylation Profiles of the Charophycean Green Alga, Penium margaritaceum, Resemble Those of Embryophytes. *Proteomes* **2018,** 6, (2), 14.
136. Rubio, M. V.; Zubieta, M. P.; Cairo, J. P. L. F.; Calzado, F.; Leme, A. F. P.; Squina, F. M.; Prade, R. A.; Damásio, A. R. L., Mapping N-linked glycosylation of carbohydrate-active enzymes in the secretome of Aspergillus nidulans grown on lignocellulose. *Biotechnology for biofuels* **2016,** 9, (1), 168.
137. Parker, B. L.; Thaysen-Andersen, M.; Solis, N.; Scott, N. E.; Larsen, M. R.; Graham, M.

E.; Packer, N. H.; Cordwell, S. J., Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *Journal of proteome research* **2013,** 12, (12), 5791-5800.
138. Tian, Y.; Kelly-Spratt, K. S.; Kemp, C. J.; Zhang, H., Mapping tissue-specific expression of extracellular proteins using systematic glycoproteomic analysis of different mouse tissues. *Journal of proteome research* **2010,** 9, (11), 5837-5847.
139. Medzihradszky, K. F.; Kaasik, K.; Chalkley, R. J., Tissue-specific glycosylation at the glycopeptide level. *Molecular & Cellular Proteomics* **2015,** 14, (8), 2103-2110.
140. Ruiz-May, E.; Thannhauser, T. W.; Zhang, S.; Rose, J. K. C., Analytical technologies for identification and characterization of the plant N-glycoproteome. *Frontiers in plant science* **2012,** 3, 150.
141. Strasser, R., Plant protein glycosylation. *Glycobiology* **2016,** 26, (9), 926-939.
142. Song, W.; Henquet, M. G.; Mentink, R. A.; van Dijk, A. J.; Cordewener, J. H.; Bosch, D.; America, A. H.; van der Krol, A. R., N-glycoproteomics in plants: perspectives and challenges. *Journal of proteomics* **2011,** 74, (8), 1463-1474.
143. Ying, J.; Zhao, J.; Hou, Y.; Wang, Y.; Qiu, J.; Li, Z.; Tong, X.; Shi, Z.; Zhu, J.; Zhang, J., Mapping the N-linked glycosites of rice (Oryza sativa L.) germinating embryos. *PloS one* **2017,** 12, (3), e0173853.
144. Song, W.; Mentink, R. A.; Henquet, M. G.; Cordewener, J. H.; van Dijk, A. D.; Bosch, D.; America, A. H.; van der Krol, A. R., N-glycan occupancy of Arabidopsis N-glycoproteins. *Journal of proteomics* **2013,** 93, 343-355.
145. Breidenbach, M. A.; Palaniappan, K. K.; Pitcher, A. A.; Bertozzi, C. R., Mapping yeast N-glycosites with isotopically recoded glycans. *Molecular & Cellular Proteomics* **2012,** 11, (6), M111. 015339.
146. Halim, A.; Larsen, I. S. B.; Neubert, P.; Joshi, H. J.; Petersen, B. L.; Vakhrushev, S. Y.; Strahl, S.; Clausen, H., Discovery of a nucleocytoplasmic O-mannose glycoproteome in yeast. *Proceedings of the National Academy of Sciences* **2015,** 112, (51), 15648-15653.
147. Neubert, P.; Halim, A.; Zauser, M.; Essig, A.; Joshi, H. J.; Zatorska, E.; Larsen, I. S. B.; Loibl, M.; Castells-Ballester, J.; Aebi, M., Mapping the O-mannose glycoproteome in Saccharomyces cerevisiae. *Molecular & Cellular Proteomics* **2016,** 15, (4), 1323-1337.
148. Shaw, T. I.; Zhang, M., HIV N-linked glycosylation site analyzer and its further usage in anchored alignment. *Nucleic acids research* **2013,** 41, (W1), W454-W458.
149. Go, E. P.; Irungu, J.; Zhang, Y.; Dalpathado, D. S.; Liao, H.-X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H., Glycosylation Site-Specific Analysis of HIV Envelope Proteins (JR-FL and CON-S) Reveals Major Differences in Glycosylation Site Occupancy, Glycoform Profiles, and Antigenic Epitopes' Accessibility. *Journal of proteome research* **2008,** 7, (4), 1660-1674.
150. Bagdonaite, I.; Nordén, R.; Joshi, H. J.; King, S. L.; Vakhrushev, S. Y.; Olofsson, S.; Wandall, H. H., Global mapping of O-glycosylation of varicella zoster virus, human cytomegalovirus, and Epstein-Barr virus. *Journal of Biological Chemistry* **2016,** 291, (23), 12014-12028.
151. Smith, G. T.; Sweredoski, M. J.; Hess, S., O-linked glycosylation sites profiling in

Mycobacterium tuberculosis culture filtrate proteins. *Journal of proteomics* **2014,** 97, 296-306.
152. Boysen, A.; Palmisano, G.; Krogh, T. J.; Duggin, I. G.; Larsen, M. R.; Møller-Jensen, J., A novel mass spectrometric strategy "BEMAP" reveals Extensive O-linked protein glycosylation in Enterotoxigenic Escherichia coli. *Scientific reports* **2016,** 6, 32016.
153. Fodero, L. R.; Sáez-Valero, J.; Barquero, M. S.; Marcos, A.; McLean, C. A.; Small, D. H., Wheat germ agglutinin - binding glycoproteins are decreased in Alzheimer's disease cerebrospinal fluid. *Journal of neurochemistry* **2001,** 79, (5), 1022-1026.
154. Hashimoto, Y.; Taniguchi, M.; Kitaura, M.; Nakamura, Y.; Jimbo, D.; Urakami, K., Alteration of Concanavalin A Binding Glycoproteins in Cerebro spinal Fluid and Serum of Alzheimer's Disease Patients. *Yonago Acta medica* **2008,** 51, 1-9.
155. Shah, P.; Wang, X.; Yang, W.; Eshghi, S. T.; Sun, S.; Hoti, N.; Chen, L.; Yang, S.; Pasay, J.; Rubin, A., Integrated proteomic and glycoproteomic analyses of prostate cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. *Molecular & Cellular Proteomics* **2015,** 14, (10), 2753-2763.
156. Tan, Z.; Yin, H.; Nie, S.; Lin, Z.; Zhu, J.; Ruffin, M. T.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M., Large-scale identification of core-fucosylated glycopeptide sites in pancreatic cancer serum using mass spectrometry. *Journal of proteome research* **2015,** 14, (4), 1968-1978.
157. Chen, I.-H.; Aguilar, H. A.; Paez Paez, J. S.; Wu, X.; Pan, L.; Wendt, M. K.; Iliuk, A. B.; Zhang, Y.; Tao, W. A., An analytical pipeline for discovery and verification of glycoproteins from plasma-derived extracellular vesicles as breast cancer biomarkers. *Analytical chemistry* **2018**.
158. Kang, T.; Jensen, P.; Huang, H.; Christensen, G. L.; Billestrup, N.; Larsen, M. R., Characterization of the molecular mechanisms underlying glucose stimulated insulin secretion from isolated pancreatic β-cells using post-translational modification specific proteomics (PTMomics). *Molecular & Cellular Proteomics* **2018,** 17, (1), 95-110.
159. Tyleckova, J.; Valekova, I.; Zizkova, M.; Rakocyova, M.; Marsala, S.; Marsala, M.; Gadher, S. J.; Kovarova, H., Surface N-glycoproteome patterns reveal key proteins of neuronal differentiation. *Journal of proteomics* **2016,** 132, 13-20.
160. Freeze, H. H., Understanding human glycosylation disorders: biochemistry leads the charge. *Journal of Biological Chemistry* **2013,** 288, (10), 6936-6945.
161. Palmisano, G.; Parker, B. L.; Engholm-Keller, K.; Lendal, S. E.; Kulej, K.; Schulz, M.; Schwämmle, V.; Graham, M. E.; Saxtorph, H.; Cordwell, S. J., A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Molecular & cellular proteomics* **2012,** 11, (11), 1191-1202.
162. Bu, T.-t.; Shen, J.; Chao, Q.; Shen, Z.; Yan, Z.; Zheng, H.-y.; Wang, B.-c., Dynamic N-glycoproteome analysis of maize seedling leaves during de-etiolation using Concanavalin A lectin affinity chromatography and a nano-LC–MS/MS-based iTRAQ approach. *Plant cell reports* **2017,** 36, (12), 1943-1958.
163. Hare, N. J.; Lee, L. Y.; Loke, I.; Britton, W. J.; Saunders, B. M.; Thaysen-Andersen, M.,

Mycobacterium tuberculosis infection manipulates the glycosylation machinery and the N-glycoproteome of human macrophages and their microparticles. *Journal of proteome research* **2016,** 16, (1), 247-263.

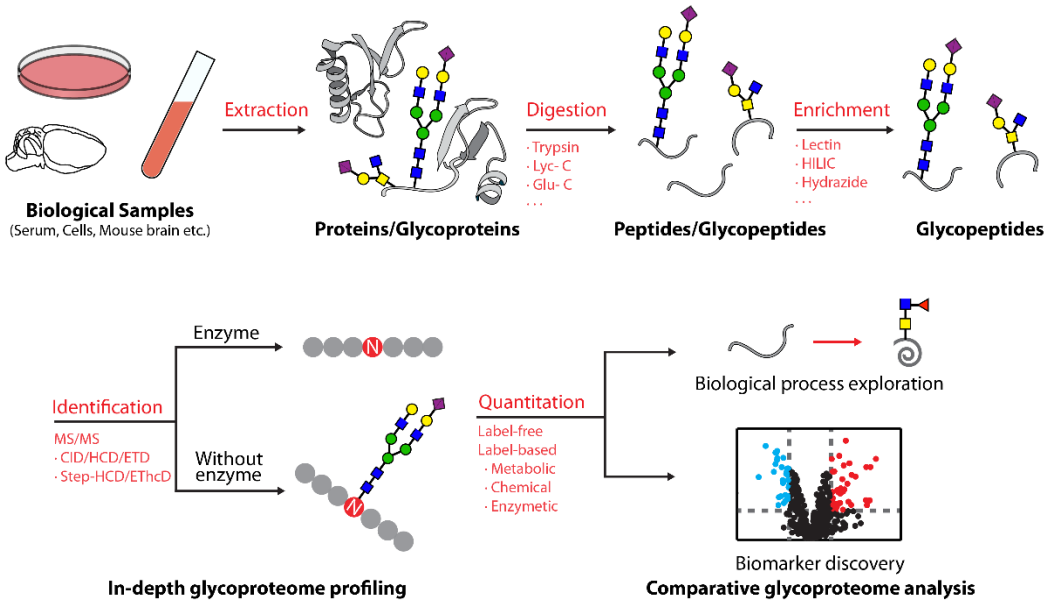


Figure 1. A typical workflow for MS-based glycoproteomics in different complex biological samples.

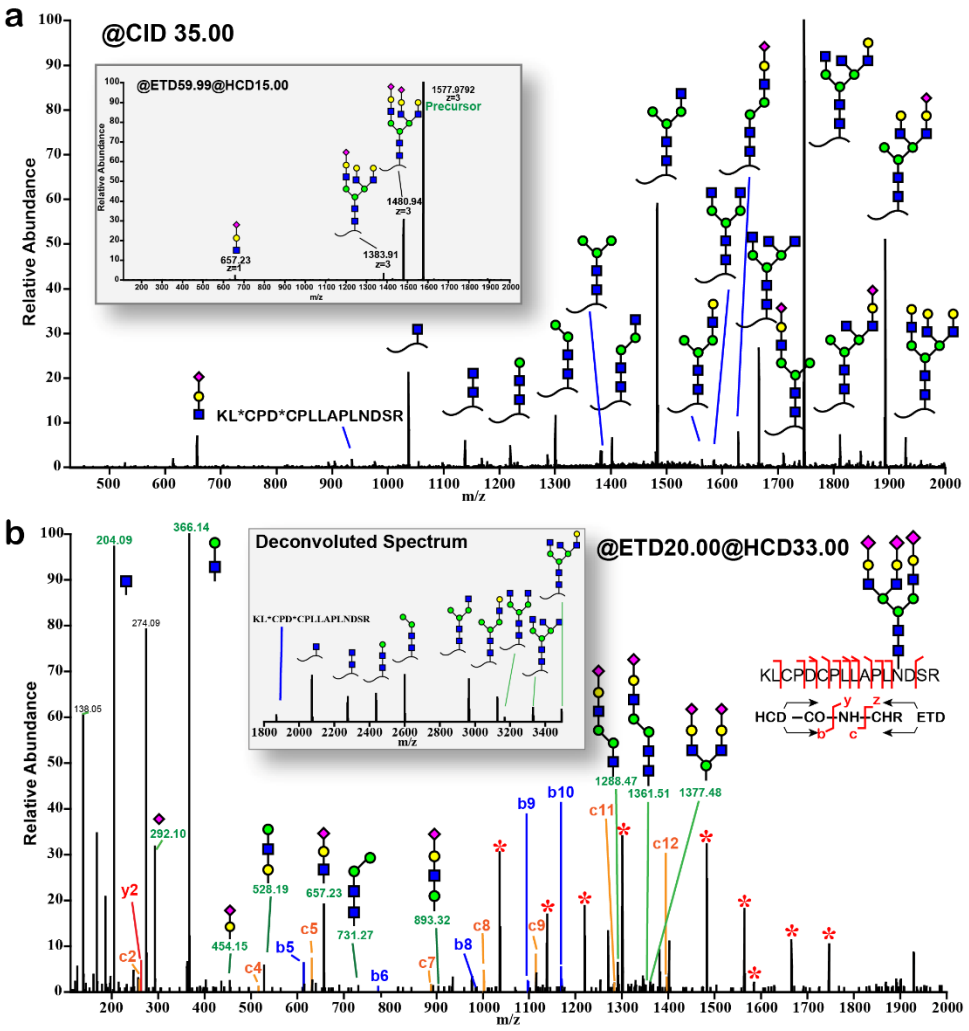


Figure 2. MS/MS of 3+ charge state precursor ion at m/z 1577.9 of bovine fetuin triantennary N-glycopeptide KLCPDCPLLAPLNDSR (AA 126–141). Alternating between CID/ETD/EThcD resulted in different sets of ions. (a) CID and ETD spectra (inset). Asterisk (*) in the peptide sequence indicates carbamidomethylation. (b) EThcD spectrum. Starred peaks (*) in the spectra were deconvoluted and annotated in the inset. [69]

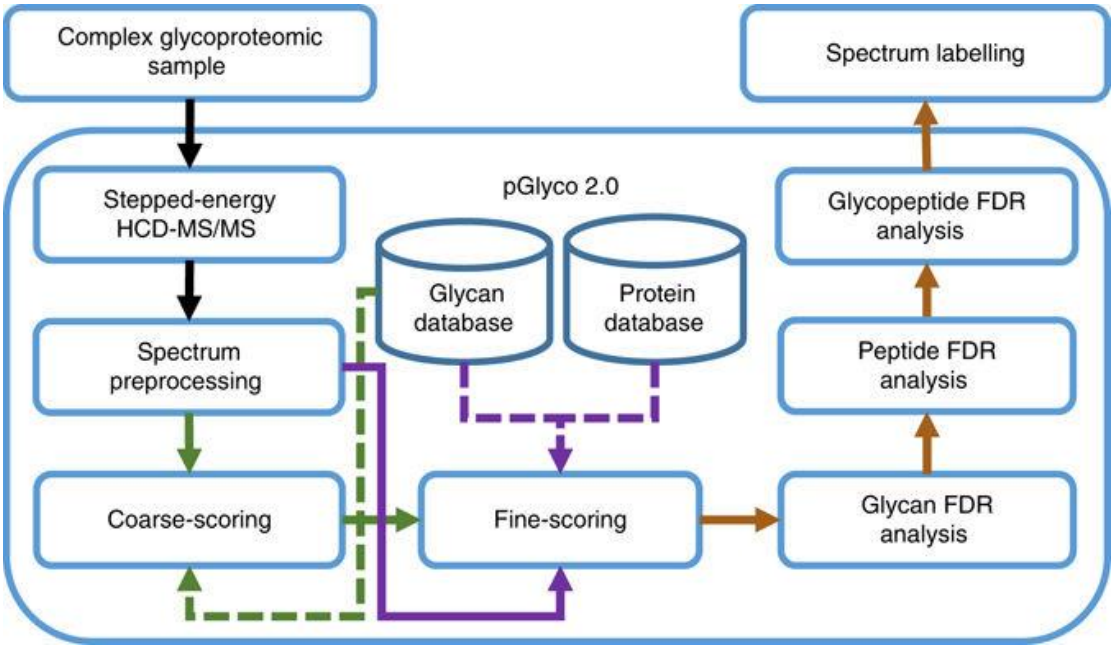
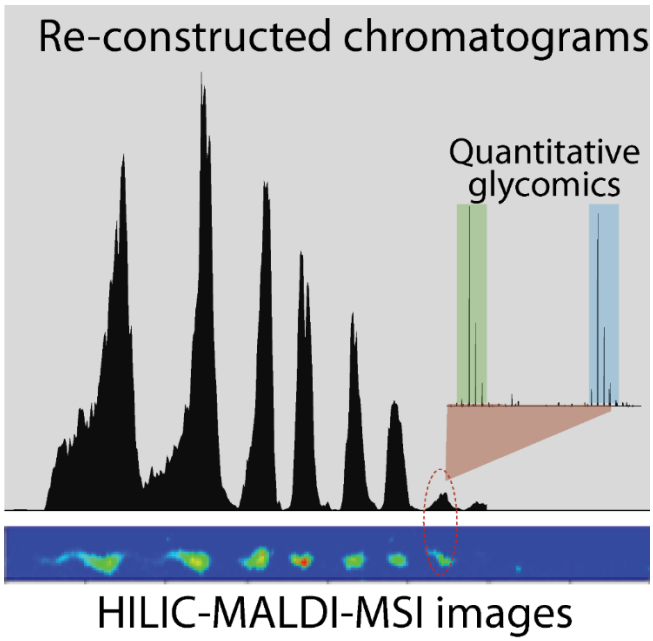


Figure 3. Design of a dedicated software pGlyco 2.0 for intact glycopeptide interpretation. [61]

Chapter 4

Development of a Hydrophilic Interaction Liquid Chromatography Coupled with atrix-Assisted Laser Desorption/Ionization-Mass Spectrometric Imaging Platform for N-glycan Relative Quantitation Using Stable-isotope Labeled Hydrazide Reagents



- Adapted from **Chen, Z.**, Zhong, X., Tie, C., Chen, B., Zhang, X., & Li, L. (2017). Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents. Analytical and bioanalytical chemistry, 409(18), 4437-4447.

- **Author contribution:** study was designed by **Chen, Z.**, Zhong, X., Zhang, X., Li, L.; experiment was performed by Chen, Z. and Tie, C., Chen, B.; manuscript was written by Chen, Z., and edited by Zhong, X. and Li, L.

Abstract

In this work, the capability of newly developed platform-hydrophilic interaction liquid chromatography (HILIC) coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging (MALDI-MSI) for quantitative analysis of N-glycans has been demonstrated. As a proof-of-principle experiment, heavy and light stable-isotope labeled hydrazide reagents labeled maltodextrin ladder were used to demonstrate the feasibility of the HILIC-MALDI-MSI platform for reliable quantitative analysis of N-glycans. MALDI-MSI analysis by Orbitrap mass spectrometer enabled high-resolution and high-sensitivity detection of N-glycans eluted from HILIC column, allowing the reconstruction of LC chromatograms as well as accurate mass measurements for structural inference. MALDI-MSI analysis of the collected LC traces showed that the chromatographic resolution was preserved. The N-glycans released from human serum was used to demonstrate the utility of this novel platform in quantitative analysis of N-glycans from a complex sample. Benefiting from the minimized ion suppression provided by HILIC separation, comparison between MALDI-MS and the newly developed platform HILIC-MALDI-MSI revealed that HILIC-MALDI-MSI provided higher N-glycan coverage as well as better quantitation accuracy in the quantitative analysis of N-glycans released from human serum.

Keywords: Glycomics, N-glycans, MALDI imaging, Quantitation, HILIC, Hydrazide reagents.

Introduction

Glycosylation, the most diverse and complex form of protein post-translational modifications (PTMs), has been proven to play essential role in many key biological processes including cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction and endocytosis [1]. In fact, glycosylation has been widely detected in different kinds of proteins such as growth factors, cytokines, immune receptors and enzymes [2]. It is estimated that over 70% of all human proteins are glycosylated [3]. Aberrant glycosylation has been implicated in various diseases, most strikingly in a class of diverse diseases collectively referred to as congenital disorders of glycosylation [4]. Furthermore, it has been well-established that glycosylation pattern alters significantly in cancer cells [5, 6]. Thus, accurate and reliable glycan structure charaterization and quantitation is crucial for glycan biomarker discovery and understanding the roles of glycosylation in biological processes.

Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) have become very powerful and widely used analytical tools for structure and quantitative analysis of N-glycans. As a complementary ionization method to ESI, MALDI generates only singly charged ions which simplifies the spectrum and the signal of a specific analyte are “focused” and “enhanced” on these singly charged ions. As MALDI is more tolerant to salts and other contaminants when proper matrix or matrix additives was applied [7-10], less sample clean-up is needed, thus increasing sample recovery rate and decreasing sample preparation time. However, MALDI-MS analyses of complex glycan samples are often hampered by the complexity of the sample due to ion suppression effect [11], leading to low glycan coverage, poor MS/MS

fragmentation quality and increased quantitation inaccuracy. One possible solution is to couple offline liquid chromatography (LC) separation with MALDI-MS to decrease the complexity of the analyte mixtures by chromatographic separation before MS analysis, which could provide improved sensitivity and dynamic range. In addition, decoupling the separation from MS analysis provides the opportunity to independently optimize LC separation performance and the MS performance. Another benefit for LC coupled to MALDI is that non-volatile salts such as ion pairing agent could be used to improve the chromatographic separation capability without negative impact on the MS signal.

In fact, the powerful analytical potential of both commercial and in-house built devices for LC-MALDI MS coupling have been demonstrated for both proteomics and glycomics studies [12-17]. Due to the strong retention of polar compounds, hydrophilic interaction liquid chromatography (HILIC) has been widely used for the separation of both native [18-21] and derivatized [22-28] glycans in glycomics studies. Offline coupling of HILIC with automated MALDI-MS analysis has been established for glycan structure characterization [16, 17]. However, one disadvantage is that discrete spotting of the eluate on the MALDI plate causes loss of chromatographic resolution. To overcome this weakness, our lab [29, 30] has previously developed LC-MALDI-MSI platform to demonstrate its ability for enhanced proteomics and peptidomics studies. In contrast to the offline LC-MALDI-MS coupling schemes, continuous real-time LC flow is collected on a MALDI plate and subject to MSI analysis after applying matrix. In this way, chromatographic resolution from LC separation dimension could be almost fully preserved, which leads to minimized ion suppression and enhanced signal. Comparable LC chromatograms can be re-constructed, allowing visualization of

separation of analytes. Herein, taking advantage of the powerful separation ability of HILIC for glycans, we developed a novel HILIC-MALDI-MSI platform to demonstrate its capability for carbohydrate analysis or glycomics study.

The present study is the first demonstration of HILIC-MALDI MSI as a novel platform for relative quantitative analysis of N-glycans using stable-isotope labeled hydrazide reagents. Maltodextrin ladder was used to demonstrate the feasibility of this approach for accurate relative quantitative analysis. To demonstrate the practical utility of the developed approach in complex biological samples, we conducted relative quantitative analysis of the N-glycans enzymatically released from human serum. The data revealed that the approach is reproducible and reliable. A comparison between the commonly used MALDI-MS platform and the newly developed HILIC-MALDI-MSI platform revealed that HILIC-MALDI-MSI outperformed MALDI-MS analysis in both N-glycan coverage and quantitation accuracy. Collectively, our study suggests that HILIC-MALDI-MSI assisted by the hydrazide labeling reagent is a promising new approach for quantitative N-glycan profiling with significant potential in glycan biomarker discovery and comparative glycomic studies.

Experimental section

Chemicals and materials

Acetic acid, acetonitrile, ammonium acetate and ammonium bicarbonate were obtained from Fisher Scientific (Pittsburgh, PA). Formic acid (FA), N,N-Dimethylaniline (DMA), maltooctose, maltodextrin and ribonuclease B (RNase B) from bovine pancreas were purchased from Sigma-

Aldrich (St. Louis, MO). 2,5-Dihydroxybenzonic acid (99%, DHB) was from Acros Organics (Geel, Belgium). Dithiothreitol (DTT) and PNGase F were from Promega (Madison, WI). Microcon filters YM-30 was purchased from Merck Millipore (Billerica, MA). Human serum powder, LC/MS grade isopropanol (IPA), acetonitrile (ACN), methanol (MeOH) and water were from Fisher Scientific (Pittsburgh, PA). The hydrazide labeling reagents, light tag 2-hydrazino-4,6-bis-(diethylamino)-s-triazine (HDEAT) and heavy tag 2-hydrazino-4,6-bis-(d10-diethylamino)-s-triazine (d_{20}-HDEAT), were synthesized according to a protocol described previously [31].

Release of N-glycans

The filter-aided N-glycan separation (FANGS) strategy was modified based on the previous protocol [32]. Briefly, 100 μg of human plasma or RNase B dissolved in 100 μL digestion buffer (50 mM ammonium bicarbonate) was loaded onto a 30 kDa molecular weight cutoff filter. Then 100 μL of 20 mM DTT in digestion buffer was added to the filter. Filters were incubated in water bath alternating between 100 ℃ (boiling water) and 25 ℃ (room temperature) for 15 seconds each for four cycles (2 minutes total time) to denature protein. Then the samples were concentrated onto the filter by centrifugation at 14, 000 xg for 40 min. The samples were washed with 100 μL of digestion buffer by centrifugation at 14, 000 xg for 20 min three times to remove small molecular weight contaminants. After that, the filter was transferred to a fresh collection vial, and 4 μL of PNGase F (1 IUB milli-unit/mL) and 96 μL of digestion buffer was added. The reaction was incubated at 37 ℃ for 18 h to enzymatically cleave N-glycans. N-Glycans were eluted by centrifugation at 14, 000 xg for 20 min. To ensure complete elution of N-glycans from the filter,

two more washes with 100 μL digestion buffer were collected. Purified N-glycans were dried down in SpeedVac and re-dissolved in 100 μL 50% MeOH 0.1% formic acid and incubated in 37 ℃ water bath for 1 hour to allow deamidation. The samples were then dried down again and stored at -20 ℃ before further derivatization.

Derivatization of glycans

To each reaction vial, 230 μg of HDEAT and d_{20}-HDEAT was aliquoted. Glycan samples (100 μg maltooctose, 100 μg maltodextrin and human serum N-glycans from 100 μg protein) were reconstituted in 200 μL 1% acetic acid, 70% IPA solution and transferred to the reaction vial. Then the reaction was incubated at 37 ℃ water bath for 2 hours. To quench the reaction, the labeled samples were dried down in SpeedVac. The light and heavy labeled maltooctose were mixed at different ratios of 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4 and 1:5 for the following MALDI-MS analysis. The heavy and light labeled maltodextrin and human serum N-glycans were mixed at a ratio of 1:1 and re-dissolved in HILIC gradient initial condition for HILIC-MALDI-MSI analysis without further purification. The heavy and light labeled human serum N-glycans were also subject to MALDI-MS analysis.

Nano-flow hydrophilic interaction liquid chromatography (HILIC)

The LC separation was performed with a homemade HILIC column on a Waters nanoACQUITY UPLC system. Samples were loaded and separated on a 75 μm ×40 cm homemade capillary column packed with 3 μm poly(2-hydroxyethyl aspartamide) (PHEA) (Poly LC, Columbia, MD). Mobile phase A and B were 10 mM ammonium acetate in water (pH 4.7) and 100% ACN, respectively. The flow rate was set at 0.3 μL/min. After dissolved in the initial gradient

condition (25% A for labeled human serum N-glycans, 10% A for labeled maltodextrin), 1 μL sample was injected. The gradient for labeled human serum N-glycans was from 25% A to 80% A over 55 min. The gradient for labeled maltodextrin was set as follows : 10-35% A (0–20 min), 35-40% A (20–21 min), 40–40% A (21-30 min), 40–45% A (30–31 min), 45–45% A (31–40 min), 45–55% A (40–41 min), 55-55% A (41-46 min), 55-70% A (46-47 min), and 70-70% A (47-55 min).

HILIC-MALDI interface and matrix application

The LC flow was collected directly on a ground stainless steel MALDI plate, which was fixed on a syringe pump (Pump 11 Elite, Harvard Apparatus, Holliston, MA, USA) and moved along with the pump.. The tip of the column directly touches the MALDI plate to deposit LC traces with an angle of 45 degree. Compared to the original capillary, the pooled tip with a smaller diameter at the very distal end could minimize the effluent diffusion. The LC eluent from 15 min to 40 min eluent was collected for labeled human serum N-glycans. After that, matrix (100 mg/mL DHB, 2% DMA in 50% ACN) was sprayed on the dried traces using a TM-Sprayer from HTX Technologies (Carrboro, NC). The parameters were carefully adjusted and set as follows: matrix flow rate: 0.25 mL/min; nozzle velocity: 1000 mm/min; temperature: 85 degree Celsius; gas pressure: 10 psi; line spacing: 3 mm; number of layers: 2; dry time: 1 min.

Data acquisition and processing

The MS images of collected LC traces were acquired with a MALDI-Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA) in positive ion mode. The instrument methods were set using Xcalibur (Thermo Scientific, Waltham, MA, USA). Labeled maltodextrin

was acquired in a mass range of *m/z* 500-4000, and labeled human serum N-glycans were acquired in the mass range of *m/z* 1000-3500 with a resolution of 60,000. A laser energy of 30 μJ was used in all spectra collection. The LC traces to be imaged and the raster step size were controlled using the LTQ Tune software (Thermo Scientific, Waltham, MA, USA). To generate images, the spectra were collected at 150 μm intervals in both the x and y dimensions. The images of analyte ions were extracted centered at the *m/z* with a window of 10 ppm using ImageQuest (Thermo Scientific, Waltham, MA, USA). ImageQuest enables construction of MS images by reconstituting the x and y coordinates of the spectra in the acquired image file with their original locations within the LC traces. Images was assigned an intensity based color scale for optimal visualization. The same intensity scales were used for the MS images of different analytes from the same image file. For each analyte, the averaged intensity of the monoisotopic peak in the extracted ion image was used for relative quantitation. The composition of human serum N-glycans was identified by matching the accurate mass (<5 ppm) with the established human serum N-glycan library [33, 34]. LC chromatograms was re-constructed using Excel and Origin based on the extracted ion intensities information exported using Xcalibur. Namely, the x coordinate of each MS spectrum was converted into the retention time, while the MS intensities at the same x coordinate but different y coordinate were summed up. In this way, the actual peak intensity of a specific ion can be plotted against the retention time.

Results and Discussion

Isotopic hydrazide labeling reagent

Zhang group [31] has synthesized the hydrazino-s-triazine based labeling reagents and previous study [35, 36] showed that the hydrazide labeling reagent HDEAT enabled a 40-fold detection enhancement in ESI mode. The labeling reagent contains three basic tertiary amine and is thus readily protonated, the labeled high-mannose N-glycans released from RNase B are almost exclusively observed as protonated species as shown in **Fig. S1-A in MALDI source.** While for the N-glycans released from human serum, which includes all kinds of high-mannose, complex, fucosylated and sialylated glycans, sodiated peaks dominated in the mass spectrum as shown in **Fig. S1-B**, probably due to the sodium ions brought by glass bottles used for human serum storage. In our study, the intensity of the most intense peak was used for relative quantitation. Stable-isotope labeling appears to be one of the most popular and effective strategy in the relative quantification of N-glycans. An identifiable mass difference was generated by incorporation of different isotopic species onto chemically similar analytes of interest, allowing simultaneous MS analysis of multiple samples. By introducing twenty deuterium atoms into light tag HDEAT, a heavy tag d_{20}-HDEAT was synthesized. With a 20 Da mass shift between the light and heavy tag, the possible isotopic distribution can be avoided for a glycan with molecular weight up to 10,000 Da, ensuring quantitation precision, simplifying data processing as well as accurate glycan identification. The structure information of the isotopic hydrazide labeling reagent and its reaction scheme with N-glycans is in shown **Fig. 1b & 1c.** The overall workflow including labeling with heavy and light tags, HILIC separation, matrix applying, MALDI imaging and MS1-based quantitation was shown in Fig. 1a.

Method validation of isotopic labeling strategy for quantitation in MALDI mode

The hydrazide labeling efficiency was evaluated using maltooctose as standards with different labeling ratio. As we could see from the **Fig. S2**, when the labeling reagent VS. maltooctose molar ratio reaches 20, the labeling efficiency reaches above 90%. The labeling reaction almost went to complete when the labeling ratio was increased to 50. The labeling efficiency was determined based on the sum of the monoisotpopic peak of proton, sodium and potassium adducts of labeled and unlabeled standard maltooctose. Previously, the reliability and accuracy of quantitation of the isotopic labeling reagent HDEAT and d_{20}-HDEAT has been demonstrated in ESI mode [37]. In our study, Maltooctose and high-mannose N-glycans released from RNase B were used to evaluate their feasibility and reliability in MALDI mode. Our results shows that accurate quantitation can be achieved.

Even though the synthesized d_{20}-HDEAT achieved rather high purity above 90% benefiting from the commercially available high deuterated (99.6%) d_{11}-diethylamine, there is still 8.2% of d_{19}-HDEAT in the final product [37]. Thus, the intensity of d_{20}-HDEAT labeled glycans peak need to be corrected to ensure accurate quantitation. As shown in the example of labeled maltodextrin$_{(3)}$ in **Fig. S3,** instead of being labeled exclusively by d_{20}-HDEAT, part of the samples were labeled by d_{19}-HDEAT. On the other hand, the isotopic peak of d_{19}-HDEAT labeled glycan overlaps with the monoisotopic peak of d_{20}-HDEAT labeled glycan at resolution of 60,000 used in our study. Taking these two aspects into consideration, we can get the equation for the corrected peak intensity of heavy labeled glycan as shown in equation (1) and (2). The correction factor varies from each N-glycan, thus we calculated the factor based on the N-glycan composition with resolution set at 60,000 using the isotope distribution calculation tools on the website

http://www.chemcalc.org/.

Intensity of heavy labeled glycan monoisotopic peak = Intensity of d_{20}-HDEAT labeled glycan monoisotopic peak + Correction factor×Intensity of d_{19}-HDEAT labeled monoisotopic peak (1)

Correction factor=1-Abundance of 2^{nd} isotopic peak of Intensity of d_{19}-HDEAT labeled monoisotopic peak / Abundance of 1^{st} isotopic peak of d_{19}-HDEAT labeled monoisotopic peak (2)

Light and heavy labeled maltooctose was used to evaluate the quantification accuracy of the isotopic quantification in different ratio. A total of 9 samples were prepared in different ratios: 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4 and 1:5. The quantitation errors were well within 10 % when the light/heavy (L:H) ratio varied from 1:1 to 1:5. To verify that there was no essential difference between the light and heavy tag, the accuracy of a reverse labeling, namely light/heavy ratio from 5:1 to 1:1 was also evaluated. The results showed that comparable coefficients of variations within 10 % was obtained. **Fig. 2A** plots experimental $\log_2$(L:H) vs. theoretical $\log_2$(L:H) ratio and a weighted least squares regression was calculated. One example of the mass spectrum of light and heavy labeled maltooctose at the ratio of 5:1 and 1:5 was shown in **Fig. 2B.** The linear range of quantification spans a 5-fold change in glycan abundance in both directions. Labeling accuracy and efficiency was also evaluated using the high-mannose N-glycans released from RNase B. The mass spectrum of heavy and light labeled (ratio 1:1) N-glycans released from RNase B was shown in Fig. 3A. A percentage difference within 15% compared with the theoretical ratio was achieved as shown in **Fig. 3B** using three technical replicates. When labeling the two samples in parallel

with light and heavy tags, the labeling efficiency difference plays a vital role in the accuracy of quantitation. Thus, we evaluated the labeling efficiency in two parallel labeling experiments. The results in **Fig. 3C** showed pretty consistent labeling efficiency for the same N-glycan with a maximum difference of 4 % and an average difference of 3.4% (95% confidence interval, 3.2%–3.7%).

Reconstruction of HILIC chromatograms from MALDI-MSI results

Maltodextrin ladder, a mixture of oligomers, was used for the methodology evaluation. MALDI-Orbitrap mass spectrometer was used to image the collected LC effluent trace. On the basis of a HILIC pre-run, a time span from 14 to 35 min was set as the effluent and imaging data collection window, which covered an effective distance of 7.6 cm on the MALDI sample plate. The relative coordinates for each spectrum taken along the sample trace were recorded during acquisition and were converted to LC retention times. The MALDI plate mounted on a syringe pump moved at a speed of 3.3 mm/min. Thus, a 1.5 min HILIC peak, a typical peak width of LC separation achieved using our homemade column, can generate a length of 4.95 mm (3.3 mm/min ×1.5 min) trace on the plate. As the raster increment was set at 0.15 mm for both x and y coordinates, up to 33 (4.95 mm ÷0.15 mm) data points could be collected for each chromatographic peak imprinted on the MALDI plate, which was sufficient for re-constructing LC chromatograms. **Fig. S4** shows a typical reconstructed peak in our study, which contains 35 data points across the analyte peak. A complete re-construction of eight extracted maltodextrin peaks and the base peak imaging results are shown in **Fig. 4**. From the figure, we can see spatial distribution of the extracted ion images on the collected LC trace on the MALDI plate correlate well with the re-constructed

peak in terms of retention time. The quantitative results were shown in **Table S1**.

HILIC-MALDI-MSI quantitative analysis of isotopic labeled maltodextrin standards

To verify the feasibility of isotope relative quantitative analysis on the newly developed HILIC-MALDI-MSI platform, maltodextrin was selected as model analytes. Maltodextrin standards at a 1:1 ratio were labeled with HDEAT and d_{20}-HDEAT respectively and combined for HILIC-MALDI-MSI analysis. The chromatograms were re-constructed based on the imaging results shown in **Fig. 4B**,which shows that a baseline separation of maltodextrin oligomers (n= 2-8) was achieved on the homemade HILIC column (**Fig. 4A)**. Retention time shift has often been an issue for deuterium encoded tag in reversed phase LC runs due to isotopic effect, which may negatively affect the quantitation accuracy. The extracted ion chromatograms shown in **Fig. S5** indicated that negligible retention time shift (<5s) was observed on our homemade HILIC column. The almost identical image distribution and retention time of the reconstructed peak based on the extracted ion images of the light and heavy labeled N-glycans shown in **Fig. 5A & B** also support this conclusion. The previous study in our lab [38] using CE-MALDI-MSI for quantitation of peptides using isotopic formaldehyde labeling has shown the peak pair ratios of the extracted ion intensity from MALDI MS imaging can achieve reliable and accurate quantitation results (with CV <15%). Hence, the intensity of the monoisotopic peak from extracted ion image was used for quantitation, and the results shows that a good quantitation accuracy was achieved, which was detailed in **Table S1**. An example of the peak pair intensity ratios from the extracted ion images of isotopically labeled maltodextrin (Degree of polymerization=4) is shown in **Fig. 5C**.

Improved quantitative N-glycan profiling on HILIC-MALDI-MSI platform

To demonstrate the feasibility of the currently developed approach in complex sample, N-glycans enzymatically cleaved from the human serum were labeled by HDEAT and d_{20}-HDEAT respectively and mixed at a 1:1 ratio and subjected to HILIC-MALDI-MSI analysis. A total of 35 N-glycans were detected in protonated and sodiated form, with sodiated peak dominating the spectra as detailed in **Table S2**. Base peak images from MSI results in **Fig. 6** show that the signal of the detected N-glycans spread out along the entire trace of the MALDI plate with a span of 9.0 cm, significantly reducing ion suppression, which would help to produce enhanced ion signals for targeted analytes. The extracted ion images of two high-mannose and two sialylated N-glycans were also shown in **Fig. 6**. The spatial separation of the N-glycans on the MALDI plate confirms our previous observation that the temporal resolution of the LC separation was well-preserved. MALDI-MS analysis was used as a comparison to evaluate the capability of HILIC-MALDI-MSI platform for quantitative analysis of N-glycans in complex sample. MALDI-MS analysis was conducted using the same amount of sample. As shown in **Fig. 7A**, although three N-glycans (H5N4S2, H5N4F1S2, H7N6) show a slight smaller quantitation error on MALDI-MS platform, probably due to the "sweet spots" sampled, an overall better quantitation accuracy was achieved and 5 more N-glycans were detected on HILIC-MALDI-MSI platform. A further examination of the intensity of each analyte ion revealed that the peak intensity was ~10-fold higher on the HILIC-MALDI-MSI platform as shown in Fig. 7B, due to less ion suppression after separation.

These results indicated that the newly developed HILIC-MALDI-MSI platform provided improved performance for quantitative profiling of N-glycans. In fact, initially quantitative analysis was viewed as irreproducible and implausible in MALDI mode, because crystallization

does not yield a uniform distribution of the analyte and the measured ion intensity may vary for a given amount of analyte loaded onto the MALDI target [39]. The intensity of an analyte extracted from MALDI mass spectrum is highly dependent on the spot that is irradiated by the laser. The presence of so-called "sweet spots" is the main cause, which often results in poor shot-to-shot reproducibility and poor quantitation accuracy. Various approaches including matrix-comatrix system [40, 41] or other improved sample preparation methods [42, 43] have been developed to overcome this issue. In our study, instead of irradiating just a couple of spots of the crystallized sample, mass spectrometric imaging irradiates almost the entire samples via step-wise rastering in an area on the MALDI plate, which functions as "scanning" or "mapping" every spot in the whole area. In this way, the variations caused by "sweet spot" can be reduced. Furthermore, numerous studies [44-46] have discussed that suppression effects can also distort the quantitation accuracy when complex biological samples are analyzed due to different analytes' charge competition, basicity, hydrophobicity and relative abundances. To this end, benefiting from mass spectrometric imaging analysis, the HILIC chromatographic resolution used in our study could be well preserved, largely reducing the ion suppression effects, thus improving quantitation accuracy.

Isobaric tags, such as aminoxy TMT [47] and QUANTITY [48], have been extensively developed and applied for quantitative glycomics in ESI source, with reporter ions readily generated in an automated data dependent association (DDA) fashion. However, tandem MS experiments in MALDI source needs prior knowledge of the analytes to build up an inclusion list of precursor ions and far less automated compared to the tandem analysis in ESI source, which makes MS1-based quantitation more feasible in MALDI source. In the present study, isotopic tags HDEAT and

d20-HDEAT with 20 Da mass difference were utilized, which largely facilitated the MS1 ion intensity based quantitation. With various effective approaches developed such as MALDI-MS, LC-MALDI-MS and LC-ESI-MS for quantitative glycomics study, our current developed approach enriches the toolbox by pushing MALDI-based quantitative glycomics one step forward utilizing mass spectrometric imaging to thoroughly sample the analytes on a typical MALDI plate. Benefiting from the almost 10 fold intensity improvement, increased N-glycans quantitation accuracy and coverage of HILIC-MALDI-MSI has been demonstrated compared to regularly used MALDI-MS in the present study, but still there are many facets of this platform worth being further explored. For example, how will it perform in characterization and quantitation of glycan isomers when a porous graphic carbon (PGC) column is used? To what extent can the glycan separation, especially for glycan isomer, benefit from using high concentration of volatile salts or even non-volatile salts such as ion pairing agent? Besides, a systematic performance comparison with LC-ESI will be quite interesting to see how they complement with each other, in terms of glycan separation efficiency, glycan coverage, glycan isomer characterization and efforts needed in sample preparation. All of these evaluations are needed and will certainly help us to put HILIC-MALDI-MSI platform in a proper place in the quantitative glycomics toolbox. However, these evaluations are beyond the scope of current study. To conclude, faced with the various challenges in quantitative glycomics study, we believe more platforms will certainly contribute to a deeper glycomics study.

Conclusion

Our study explored the utility of a novel HILIC-MALDI-MSI platform for quantitative

glycomics analysis for the first time. MALDI-MSI analysis of the collected LC traces showed that the chromatographic resolution was preserved, making the most use of HILIC separation to minimize ion suppression. Accurate and reliable relative quantitation with stable isotope labeled hydrazide reagents was demonstrated using a variety of samples ranging from maltooctose, maltodextrin standards, N-glycans released from RNase B to more complex human serum sample. Compared to direct MALDI-MS, HILIC-MALDI-MSI provided higher N-glycan coverage as well as better quantitation accuracy due to unbiased sampling of all the spots by mass spectrometric imaging and minimized ion suppression after HILIC separation. Overall, the HILIC-MALDI MSI system utilizing hydrazide labeling reagents is robust with highly reproducible results, providing great potential in comparative glycomics studies.

Acknowledgments

This research was supported in part by the National Institutes of Health grants R21AG055377 and NIH R01 DK071801. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges a Vilas Distinguished Achievement Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

1. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.
2. Marth, J. D.; Grewal, P. K., Mammalian glycosylation in immunity. *Nature Reviews Immunology* **2008,** 8, (11), 874-887.
3. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1999,** 1473, (1), 4-8.
4. Jaeken, J., Congenital disorders of glycosylation. *Ann. N. Y. Acad. Sci.* **2010,** 1214, (1), 190-198.
5. Fuster, M. M.; Esko, J. D., The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature Reviews Cancer* **2005,** 5, (7), 526-542.
6. Blomme, B.; Van Steenkiste, C.; Callewaert, N.; Van Vlierberghe, H., Alteration of protein glycosylation in liver diseases. *J. Hepatol.* **2009,** 50, (3), 592-603.
7. Xu, S.; Ye, M.; Xu, D.; Li, X.; Pan, C.; Zou, H., Matrix with high salt tolerance for the analysis of peptide and protein samples by desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.* **2006,** 78, (8), 2593-2599.
8. Ohta, Y.; Iwamoto, S.; Kawabata, S.-i.; Tanimura, R.; Tanaka, K., Salt Tolerance Enhancement of Liquid Chromatography-Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry Using Matrix Additive Methylenediphosphonic Acid. *Mass Spectrometry* **2014,** 3, (1), A0031-A0031.
9. Börnsen, K. O., Influence of salts, buffers, detergents, solvents, and matrices on MALDI-MS protein analysis in complex mixtures. *Mass Spectrometry of Proteins and Peptides: Mass Spectrometry of Proteins and Peptides* **2000**, 387-404.
10. Fu, Y.; Xu, S.; Pan, C.; Ye, M.; Zou, H.; Guo, B., A matrix of 3, 4-diaminobenzophenone for the analysis of oligonucleotides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Nucleic Acids Res.* **2006,** 34, (13), e94-e94.
11. Annesley, T. M., Ion suppression in mass spectrometry. *Clin. Chem.* **2003,** 49, (7), 1041-1044.
12. Mueller, D. R.; Voshol, H.; Waldt, A.; Wiedmann, B.; Van Oostrum, J., LC-MALDI MS and MS/MS—an efficient tool in proteome analysis. In *Subcellular Proteomics*, Springer: 2007; pp 355-380.
13. Sparbier, K.; Asperger, A.; Resemann, A.; Kessler, I.; Koch, S.; Wenzel, T.; Stein, G.; Vorwerg, L.; Suckau, D.; Kostrzewa, M., Analysis of glycoproteins in human serum by means of glycospecific magnetic bead separation and LC-MALDI-TOF/TOF analysis with automated glycopeptide detection. *Journal of Biomolecular Techniques* **2007,** 18, (4), 252.
14. Hattan, S. J.; Parker, K. C., Methodology utilizing MS signal intensity and LC retention time for quantitative analysis and precursor ion selection in proteomic LC-MALDI analyses. *Anal. Chem.* **2006,** 78, (23), 7986-7996.
15. Perlman, D. H.; Huang, H.; Dauly, C.; Costello, C. E.; McComb, M. E., Coupling of protein HPLC to MALDI-TOF MS using an on-target device for fraction collection, concentration,

digestion, desalting, and matrix/analyte cocrystallization. *Anal. Chem.* **2007,** 79, (5), 2058-2066.
16. Maslen, S.; Sadowski, P.; Adam, A.; Lilley, K.; Stephens, E., Differentiation of isomeric N-glycan structures by normal-phase liquid chromatography-MALDI-TOF/TOF tandem mass spectrometry. *Anal. Chem.* **2006,** 78, (24), 8491-8498.
17. Maslen, S. L.; Goubet, F.; Adam, A.; Dupree, P.; Stephens, E., Structure elucidation of arabinoxylan isomers by normal phase HPLC–MALDI-TOF/TOF-MS/MS. *Carbohydr. Res.* **2007,** 342, (5), 724-735.
18. Zaia, J., Mass spectrometry and glycomics. *OMICS: J. Integrative Biol.* **2010,** 14, (4), 401-418.
19. Dixon, R. B.; Bereman, M. S.; Petitte, J. N.; Hawkridge, A. M.; Muddiman, D. C., One-year plasma N-linked glycome intra-individual and inter-individual variability in the chicken model of spontaneous ovarian adenocarcinoma. *Int. J. Mass spectrom.* **2011,** 305, (2), 79-86.
20. Bereman, M. S.; Williams, T. I.; Muddiman, D. C., Development of a nanoLC LTQ orbitrap mass spectrometric method for profiling glycans derived from plasma from healthy, benign tumor control, and epithelial ovarian cancer patients. *Anal. Chem.* **2008,** 81, (3), 1130-1136.
21. Bereman, M. S.; Young, D. D.; Deiters, A.; Muddiman, D. C., Development of a Robust and High Throughput Method for Profiling N-Linked Glycans Derived from Plasma Glycoproteins by NanoLC− FTICR Mass Spectrometry. *J. Proteome Res.* **2009,** 8, (7), 3764-3770.
22. Wuhrer, M.; Koeleman, C. A.; Deelder, A. M.; Hokke, C. H., Normal-phase nanoscale liquid chromatography-mass spectrometry of underivatized oligosaccharides at low-femtomole sensitivity. *Anal. Chem.* **2004,** 76, (3), 833-838.
23. Butler, M.; Quelhas, D.; Critchley, A. J.; Carchon, H.; Hebestreit, H. F.; Hibbert, R. G.; Vilarinho, L.; Teles, E.; Matthijs, G.; Schollen, E., Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. *Glycobiology* **2003,** 13, (9), 601-622.
24. Walker, S. H.; Lilley, L. M.; Enamorado, M. F.; Comins, D. L.; Muddiman, D. C., Hydrophobic derivatization of N-linked glycans for increased ion abundance in electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2011,** 22, (8), 1309-1317.
25. Walker, S. H.; Papas, B. N.; Comins, D. L.; Muddiman, D. C., Interplay of permanent charge and hydrophobicity in the electrospray ionization of glycans. *Anal. Chem.* **2010,** 82, (15), 6636-6642.
26. Bereman, M. S.; Comins, D. L.; Muddiman, D. C., Increasing the hydrophobicity and electrospray response of glycans through derivatization with novel cationic hydrazides. *Chem. Commun.* **2009,** 46, (2), 237-239.
27. Bowman, M. J.; Zaia, J., Comparative glycomics using a tetraplex stable-isotope coded tag. *Anal. Chem.* **2010,** 82, (7), 3023-3031.
28. Xia, B.; Feasley, C. L.; Sachdev, G. P.; Smith, D. F.; Cummings, R. D., Glycan reductive isotope labeling for quantitative glycomics. *Anal. Biochem.* **2009,** 387, (2), 162-170.
29. Zhang, Z.; Jiang, S.; Li, L., Semi-automated liquid chromatography–mass spectrometric imaging platform for enhanced detection and improved data analysis of complex peptides. *J. Chromatogr.* **2013,** 1293, 44-50.

30. Zhang, Z.; Kuang, J.; Li, L., Liquid chromatography-matrix-assisted laser desorption/ionization mass spectrometric imaging with sprayed matrix for improved sensitivity, reproducibility and quantitation. *Analyst* **2013,** 138, (21), 6600-6606.
31. Tie, C.; Zhang, X.-X., A new labelling reagent for glycans analysis by capillary electrophoresis-mass spectrometry. *Anal. Methods* **2012,** 4, (2), 357-359.
32. Hecht, E. S.; McCord, J. P.; Muddiman, D. C., Definitive Screening Design Optimization of Mass Spectrometry Parameters for Sensitive Comparison of Filter and Solid Phase Extraction Purified, INLIGHT Plasma N-Glycans. *Anal. Chem.* **2015,** 87, (14), 7305-7312.
33. Song, T.; Aldredge, D.; Lebrilla, C. B., A Method for In-Depth Structural Annotation of Human Serum Glycans That Yields Biological Variations. *Anal. Chem.* **2015,** 87, (15), 7754-62.
34. Aldredge, D.; An, H. J.; Tang, N.; Waddell, K.; Lebrilla, C. B., Annotation of a serum N-glycan library for rapid identification of structures. *J. Proteome Res.* **2012,** 11, (3), 1958-68.
35. Zhao, M.-Z.; Zhang, Y.-W.; Yuan, F.; Deng, Y.; Liu, J.-X.; Zhou, Y.-L.; Zhang, X.-X., Hydrazino-s-triazine based labelling reagents for highly sensitive glycan analysis via liquid chromatography–electrospray mass spectrometry. *Talanta* **2015,** 144, 992-997.
36. Weng, Y.; Sui, Z.; Jiang, H.; Shan, Y.; Chen, L.; Zhang, S.; Zhang, L.; Zhang, Y., Releasing N-glycan from Peptide N-terminus by N-terminal Succinylation Assisted Enzymatic Deglycosylation. *Sci Rep* **2015,** 5, 9770.
37. Zhao, M.-Z.; Tie, C.; Zhang, Y.-W.; Deng, Y.; Zhang, F.-T.; Zhou, Y.-L.; Zhang, X.-X., Deuterated hydrazino-s-triazine as highly-efficient labelling reagent for glycan relative quantification analysis using electrospray ionization mass spectrometry. *RSC Advances* **2015,** 5, (97), 79317-79322.
38. Wang, J.; Ye, H.; Zhang, Z.; Xiang, F.; Girdaukas, G.; Li, L., Advancing matrix-assisted laser desorption/ionization-mass spectrometric imaging for capillary electrophoresis analysis of peptides. *Anal. Chem.* **2011,** 83, (9), 3462-3469.
39. Duncan, M. W.; Roder, H.; Hunsucker, S. W., Quantitative matrix-assisted laser desorption/ionization mass spectrometry. *Briefings in functional genomics & proteomics* **2008,** 7, (5), 355-370.
40. Gusev, A. I.; Wilkinson, W. R.; Proctor, A.; Hercules, D. M., Improvement of signal reproducibility and matrix/comatrix effects in MALDI analysis. *Anal. Chem.* **1995,** 67, (6), 1034-1041.
41. Cohen, S. L.; Chait, B. T., Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.* **1996,** 68, (1), 31-37.
42. Karas, M.; Ehring, H.; Nordhoff, E.; Stahl, B.; Strupat, K.; Hillenkamp, F.; Grehl, M.; Krebs, B., Matrix ‐ assisted laser desorption/ionization mass spectrometry with additives to 2, 5 ‐ dihydroxybenzoic acid. *Org. Mass Spectrom.* **1993,** 28, (12), 1476-1481.
43. Nordhoff, E.; Schürenberg, M.; Thiele, G.; Lübbert, C.; Kloeppel, K.-D.; Theiss, D.; Lehrach, H.; Gobom, J., Sample preparation protocols for MALDI-MS of peptides and oligonucleotides using prestructured sample supports. *Int. J. Mass spectrom.* **2003,** 226, (1), 163-180.
44. Knochenmuss, R.; Zenobi, R., MALDI ionization: the role of in-plume processes. *Chem. Rev.* **2003,** 103, (2), 441-452.

45. Schlosser, G.; Pocsfalvi, G.; Huszár, E.; Malorni, A.; Hudecz, F., MALDI‐TOF mass spectrometry of a combinatorial peptide library: effect of matrix composition on signal suppression. *J. Mass Spectrom.* **2005,** 40, (12), 1590-1594.
46. Nicola, A. J.; Gusev, A. I.; Proctor, A.; Jackson, E. K.; Hercules, D. M., Application of the fast‐evaporation sample preparation method for improving quantification of angiotensin II by matrix‐assisted laser desorption/ionization. *Rapid Commun. Mass Spectrom.* **1995,** 9, (12), 1164-1171.
47. Zhong, X.; Chen, Z.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L., Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags. *Anal. Chem.* **2015,** 87, (13), 6527-6534.
48. Yang, S.; Wang, M.; Chen, L.; Yin, B.; Song, G.; Turko, I. V.; Phinney, K. W.; Betenbaugh, M. J.; Zhang, H.; Li, S., QUANTITY: an isobaric tag for quantitative glycomics. *Sci. Rep.* **2015,** 5, 17585.

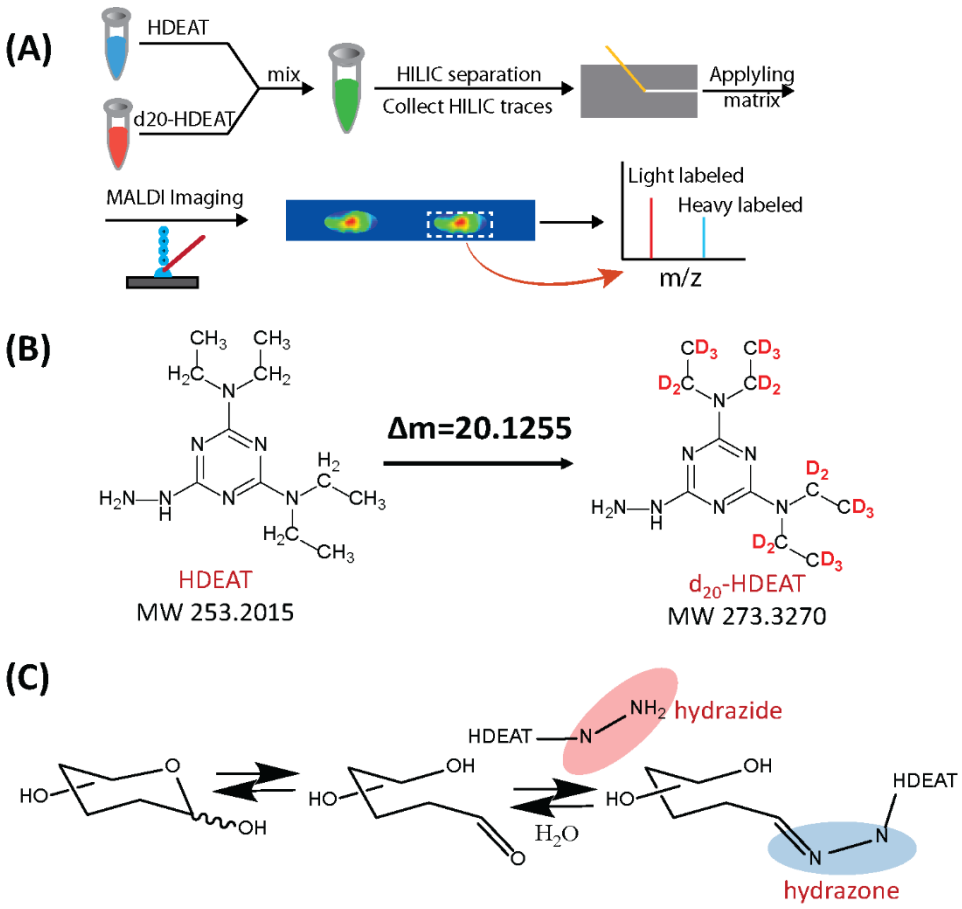


Fig. 1 (A) Overall workflow for quantitative analysis of N-glycans using duplex HDEAT labeling reagents and LC-MALDI-MSI platform. Light and heavy labeled samples were mixed together and separated by HILIC; HILIC traces were collected on MALDI plate and subjected to MSI analysis; the intensity of the analyte peak was extracted from the imaging area and used for quantitation. (B) The Chemical structures of light and heavy version of HDEAT. (C) Labeling reaction, hydrazide group of labeling reagent react with aldehyde group in glycans to form stable hydrazone product.

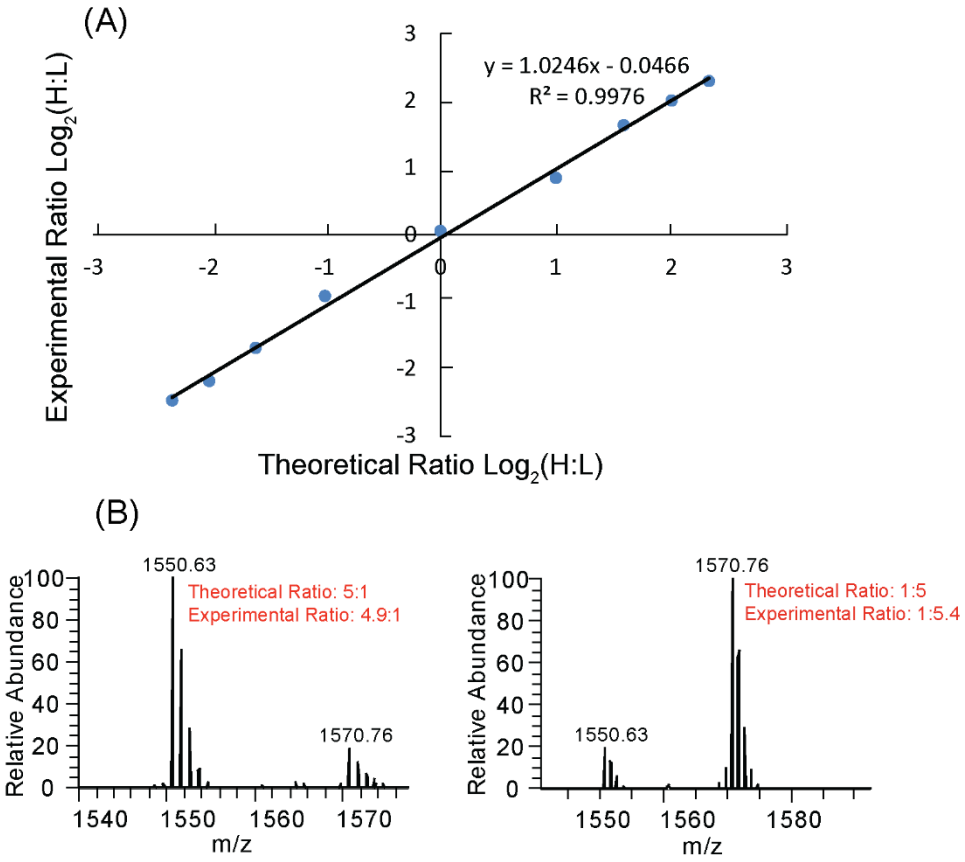


Fig. 2 (A) A plot of theoretical $\log_2$(H:L) vs. experimental $\log_2$(H:L) for the labeling ratios at 5:1, 4:1, 3:1, 2:1, 1:1 and reverse labeling at ratios of 1:2, 1:3, 1:4, 1:5 using maltooctose as standards. A weighted linear least squares regression is plotted for the data. (B) The mass spectra of heavy and light labeled maltooctose at the ratios of 5:1 and 1:5, respectively; the ion peak shown here is in protonated form. Note: The experiments were conducted using MALDI-MS.

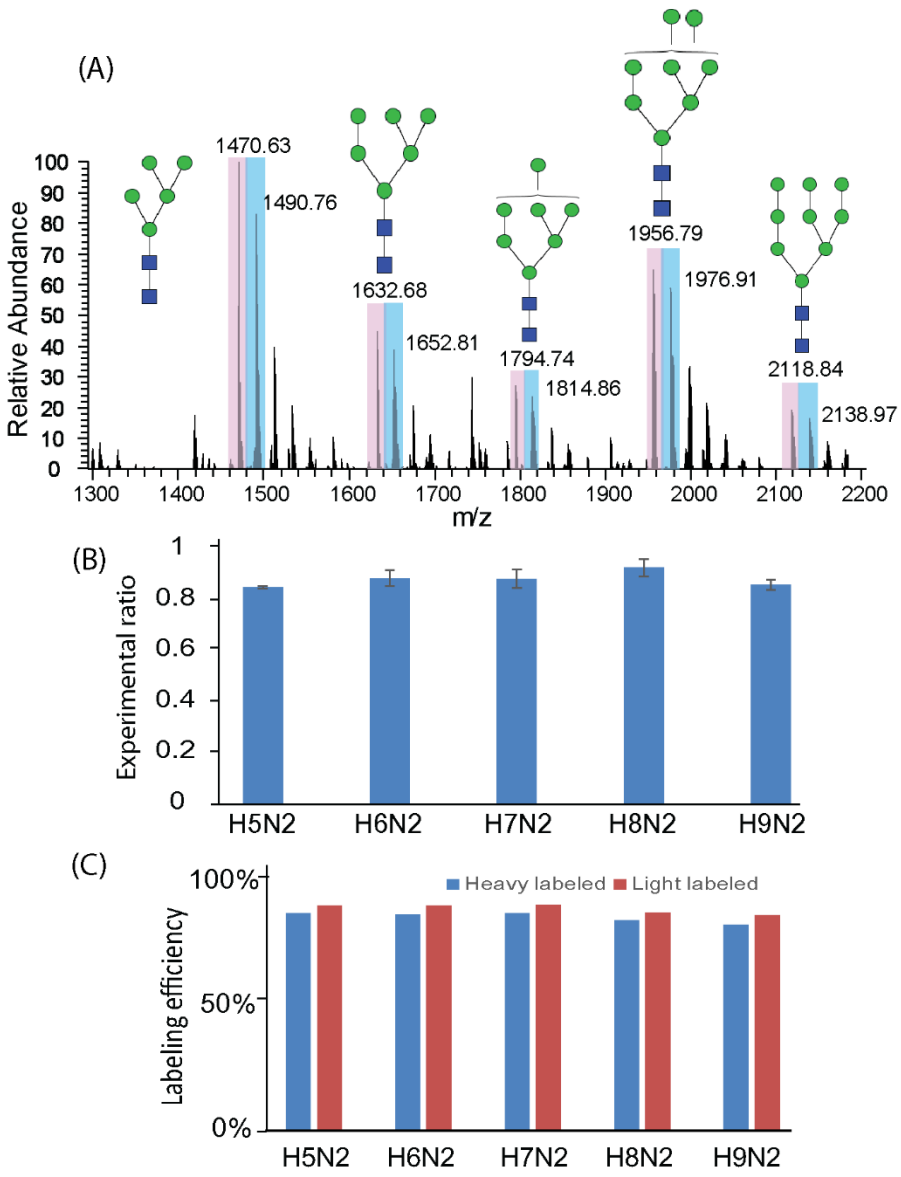


Fig. 3 (A) Mass spectrum of heavy and light labeled (ratio 1:1) N-glycans released from RNase B (Pink shaded peaks are light labeled and blue shaded peaks are heavy isotope labeled). (B) Experimental ratios of heavy and light labeled RNase B N-glycans. The error bars stand for standard deviations in three technical replicates. (C) Labeling efficiency comparison of the heavy and light labeled RNase B N-glycans in the two parallel reaction vials. (H: Hexose, N: N-acetylhexoseamine). Note: The experiments were conducted using MALDI-MS.

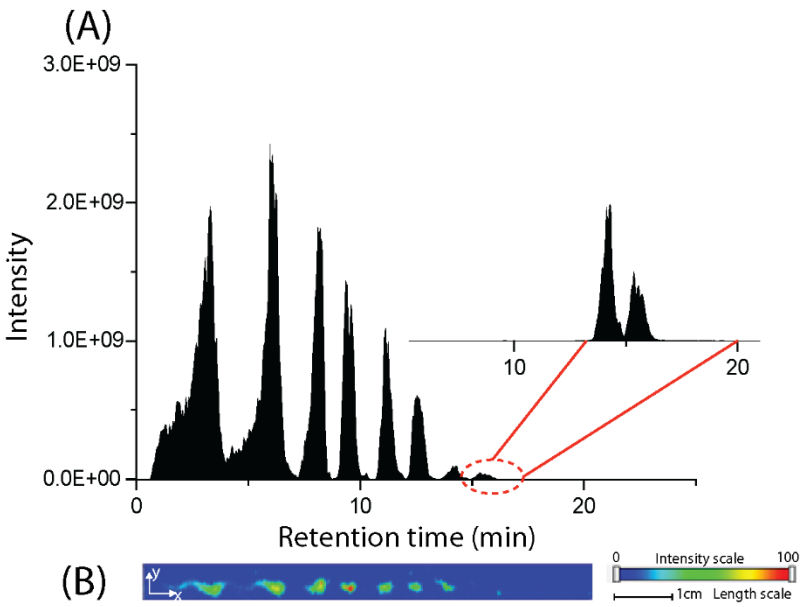


Fig. 4 (A) Re-constructed extracted ion chromatograms of heavy and light labeled maltodextrin labeled at 1:1 ratio based on MSI results. (B) HILIC traces base peak images for heavy and light labeled maltodextrin labeled at 1:1 ratio. Note: The image area of heavy and light labeled maltodextrin overlapped with each other.

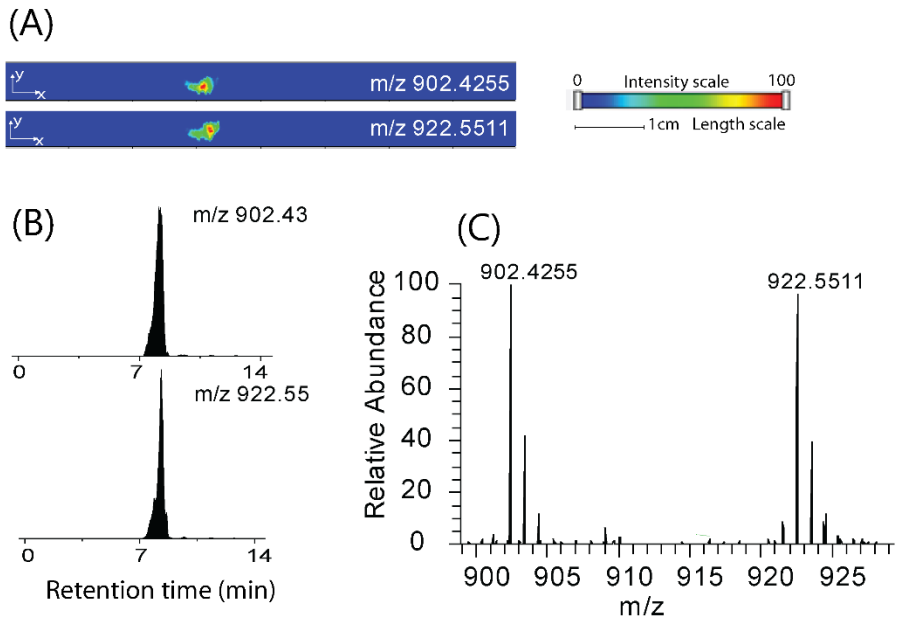


Fig. 5 (A) Extracted ion images of heavy and light labeled maltodextrin$_{(4)}$ at 1:1 ratio based on MSI results. (B) Re-constructed LC peak of heavy and light labeled maltodextrin (DP=4) based on HILIC-MALDI-MSI results. (C) Mass spectrum of heavy and light labeled maltodextrin (DP=4) in protonated form.

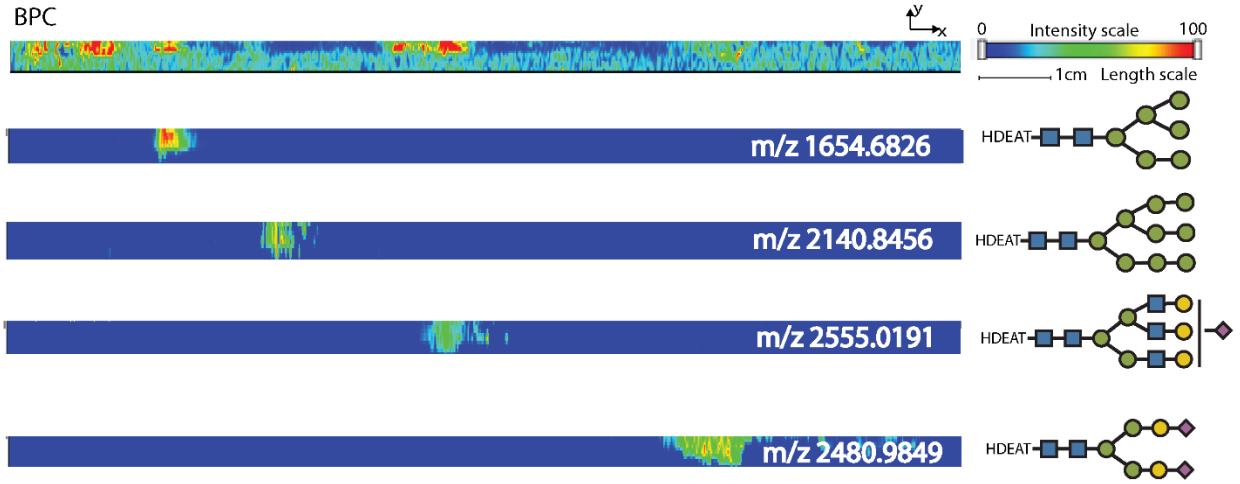


Fig. 6 HILIC traces base peak images for heavy and light labeled N-glycans released from human serum and extracted ion imaging results of four representative light labeled N-glycans. All the ions are in sodiated form. Note: Heavy and light labeled N-glycans overlapped with each other, the extracted images presented here are light labeled N-glycans.

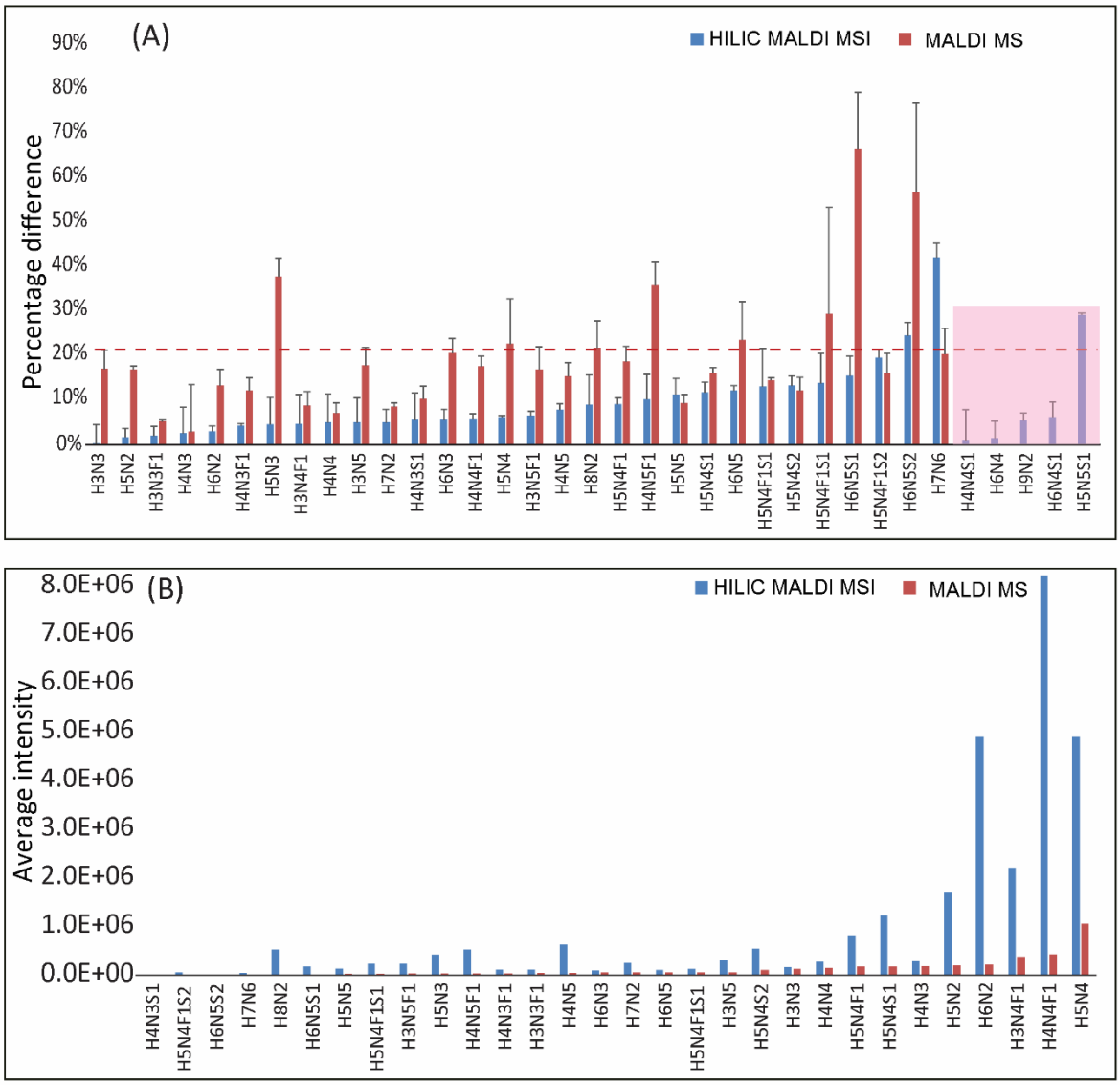


Fig. 7 (A) Quantitation accuracy and N-glycan coverage comparison between HILIC-MALDI-MSI and MALDI-MS using 1:1 labeled N-glycans released from human serum. The pink shaded represents the N-glycans only identified and quantified on HILIC-MALDI-MSI platform. Quantitation accuracy was represented as the percentage difference between experimental ratio and theoretical ratio. The error bar stands for standard deviations of measured ratios from three experiment replicates. (B) Average intensity difference between these two platforms.

Supporting information:

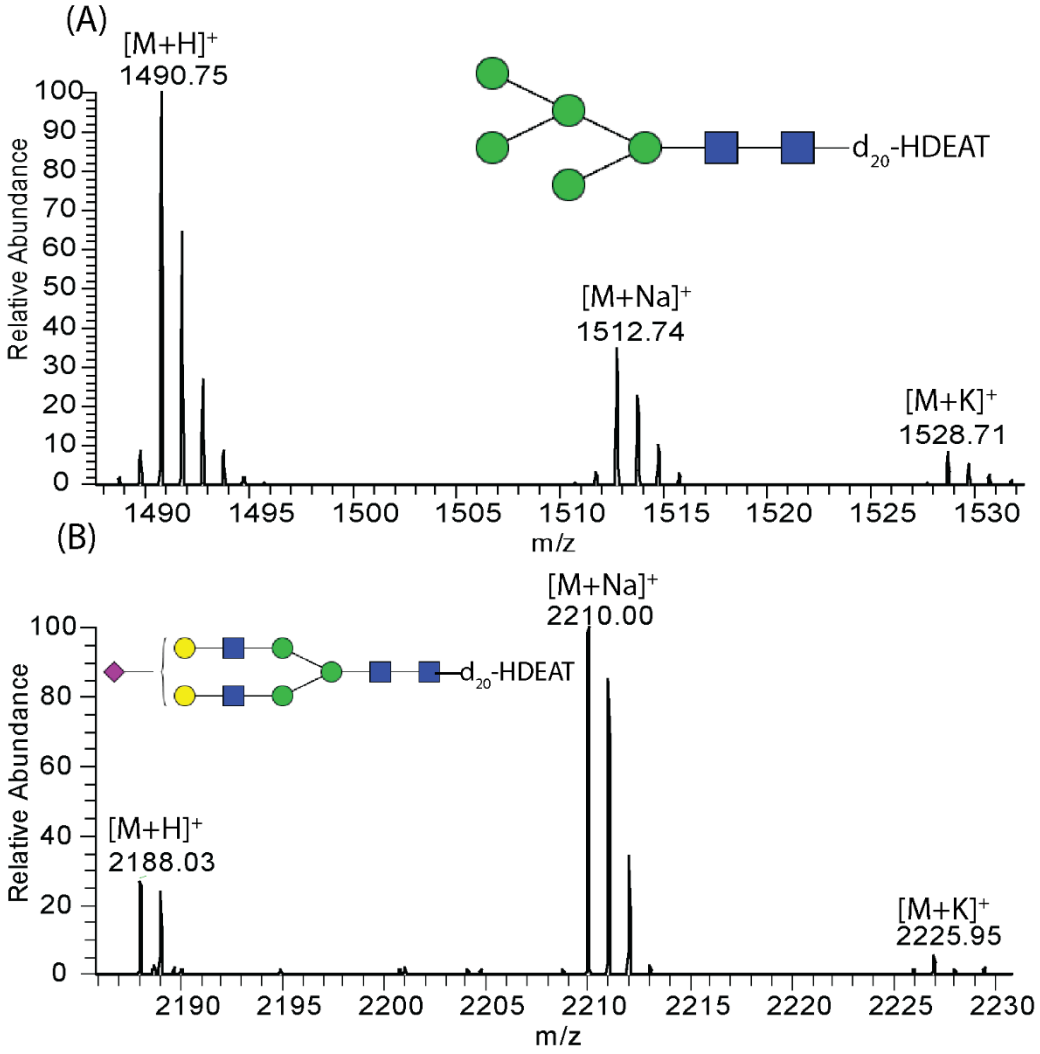


Figure S1. (A) Mass spectrum of d20-HDEAT labeled high-mannose N-glycan (H5N2) released from RNase B. (B) Mass spectrum of d20-HDEAT labeled N-glycan (Hex5N4S1) released from human serum. (H:hexose, N: N-acetylhexoseamine, F: fucose, S: sialic acid).

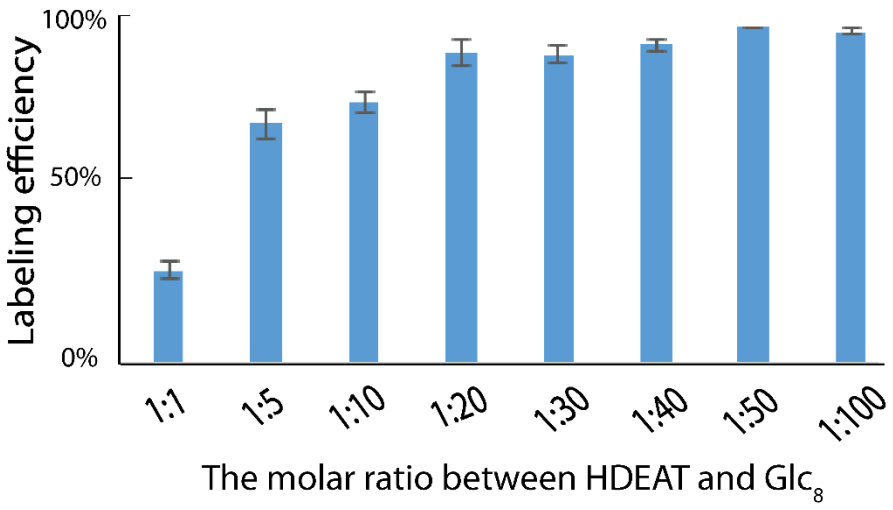


Figure. S2 The labeling efficiency under different reaction ratio between labeling reagent HDEAT and maltooctose

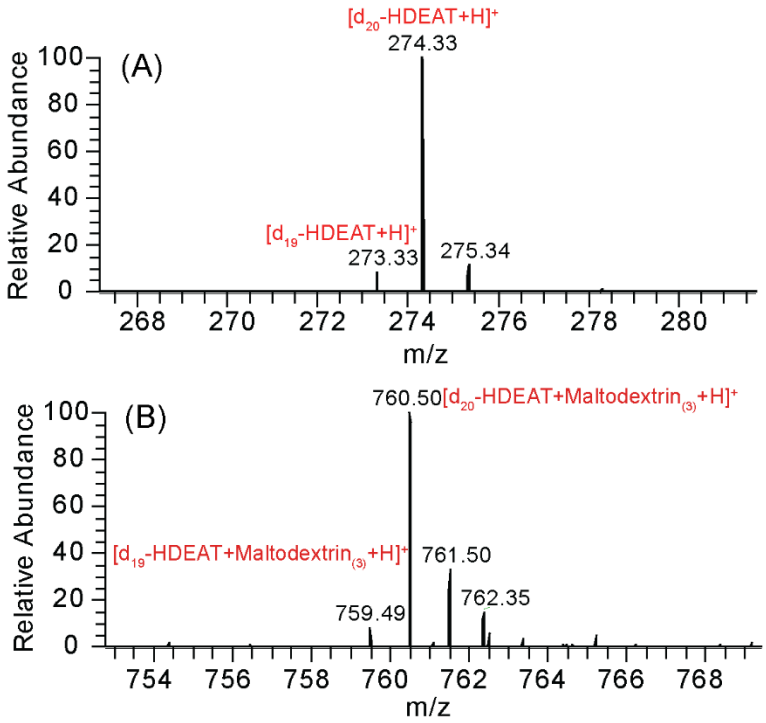


Figure. S3. (A) Mass spectrum of d20-HDEAT. (B) Mass spectrum of d20-HDEAT labeled matodextrin (DP=4), Resolution was set at 60, 000 on Orbitrap**.**

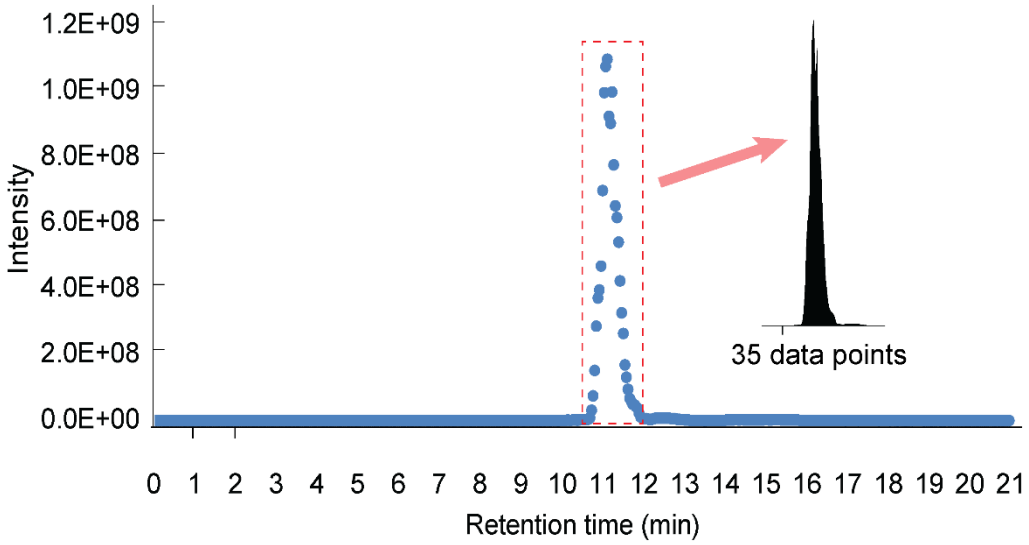


Figure. S4 Re-constructed LC peak for maltodextrin (DP=4) based on HILIC-MALDI-MSI results, the peak includes 35 data points from MSI results.

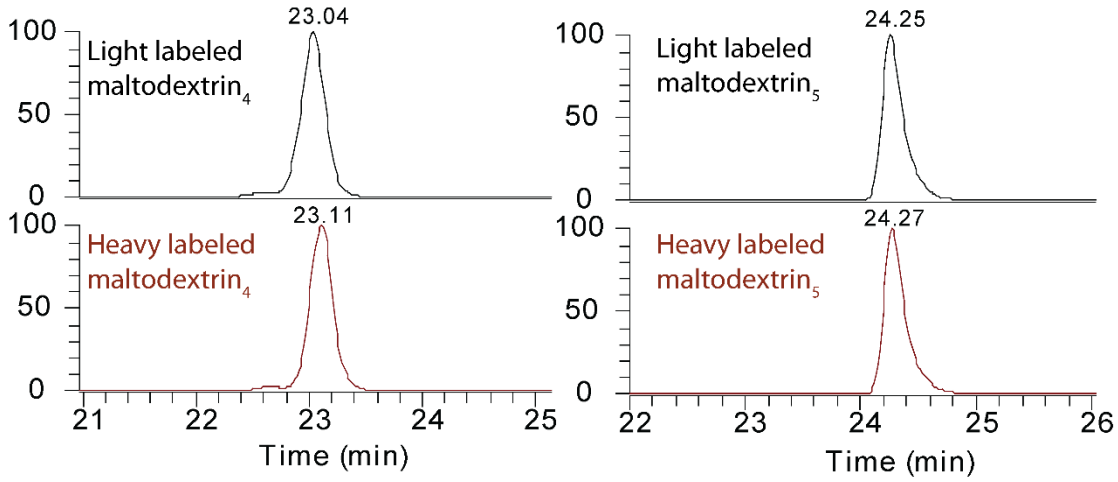


Figure. S5 Extracted ion chromatograms for light and heavy labeled maltodextrin (DP=4) and maltodextrin (DP=5) in protonated form on our homemade HILIC column.

Table S1. Quantitation results of light and heavy labeled maltodextrin at a theoretical 1:1 ratioon HILIC-MALDI-MSI platform.

No.	Maltodextrin ogligomers	m/z of HDEAT labeled maltodextrin in protonated form	m/z of d20-HDEAT labeled maltodextrin in protonated form	Expeirmental ratio	Error
1	Maltodextrin(2)	578.31	598.44	1.13	13%
2	Maltodextrin(3)	740.37	760.49	0.99	-1%
3	Maltodextrin(4)	902.42	922.55	1.08	8%
4	Maltodextrin(5)	1064.47	1084.60	1.12	12%
5	Maltodextrin(6)	1226.53	1246.65	1.06	6%
6	Maltodextrin(7)	1388.58	1408.70	1.07	7%
7	Maltodextrin(8)	1550.63	1570.76	1.02	2%
8	Maltodextrin(9)	1712.68	1732.81	1.12	12%

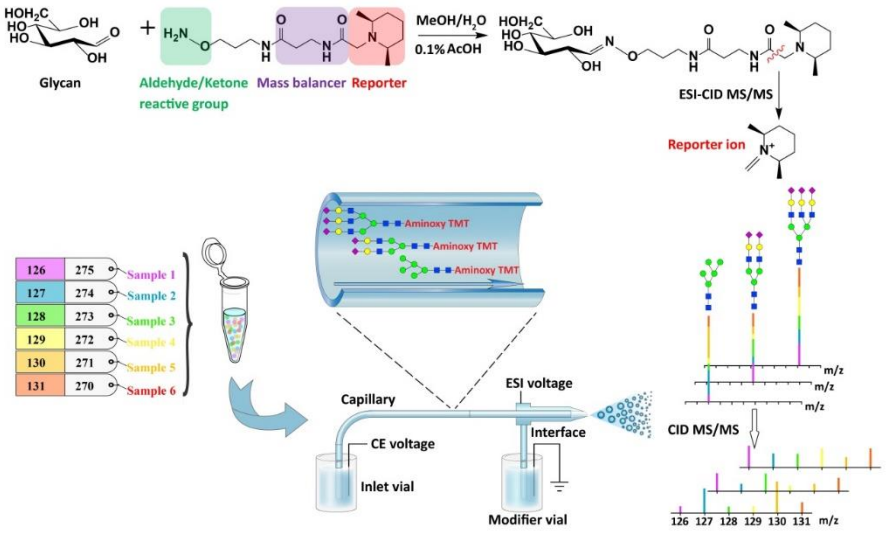
Table S2. Comparison of MS-MALDI and HILIC-MALDI-MSI quantitative results of HDEAT and d20-HDEAT labelled N-glycans released from human serum. (H:hexose, N: N-acetylhexoseamine, F: fucose, S: Sialic acid)

No.	Glycan compositions	m/z of HDEAT Labelled glycans in sodiated form	m/z of d20-HDEAT Labelled glycans in sodiated form	Correction factor	Peak Intensity of HDEAT labeled peak using MALDI-MS	Peak Intensity of d20-HDEAT labeled peak using HILIC-MALDI-MSI	Quantitation error using HILIC-MALDI-MSI	Quantitation Error using MALDI-MS	RSD using MALDI-MS	RSD using HILIC-MALDI-MSI
1	H3N3F1	1517.65	1537.77	0.31	3.4E+04	8.9E+04	0%	17%	4%	4%
2	H5N4F1	2044.83	2064.96	0.08	1.4E+05	6.7E+05	2%	17%	1%	2%
3	H4N4F1	1882.78	1902.90	0.15	3.5E+05	6.7E+06	2%	5%	0%	2%
4	H3N5F1	1923.80	1943.93	0.12	2.8E+04	1.9E+05	3%	-3%	11%	6%
5	H7N2	1816.72	1836.85	0.20	4.0E+04	2.0E+05	-3%	13%	4%	1%
6	H3N5	1777.75	1797.87	0.19	4.7E+04	2.6E+05	-4%	12%	3%	0%
7	H6N5S1	2555.00	2575.13	-0.14	6.6E+03	1.4E+05	5%	38%	4%	6%
8	H4N3	1533.64	1553.77	0.31	1.5E+05	2.5E+05	5%	9%	3%	7%
9	H4N5	1939.80	1959.92	0.12	3.7E+04	5.1E+05	5%	7%	2%	6%
10	H4N5F1	2085.86	2105.98	0.05	3.1E+04	4.3E+05	5%	18%	4%	6%
11	H5N4	1898.77	1918.90	0.15	8.6E+05	4.0E+06	5%	9%	1%	3%
12	H6N3	1857.75	1877.87	0.17	3.9E+04	8.1E+04	6%	10%	3%	6%
13	H5N5	2101.85	2121.98	0.05	1.6E+04	1.1E+05	6%	21%	3%	2%
14	H6N2	1654.67	1674.79	0.27	1.7E+05	4.0E+06	6%	18%	2%	1%
15	H8N2	1978.77	1998.90	0.13	6.3E+03	4.3E+05	6%	23%	10%	0%
16	H5N4F1S1	2539.01	2559.13	-0.15	4.3E+04	1.0E+05	7%	17%	5%	1%

17	H3N3	1371.59	1391.71	0.38	9.9E+04	1.3E+05	8%	15%	3%	1%
18	H4N3F1	1679.70	1699.82	0.24	3.3E+04	9.0E+04	9%	22%	6%	7%
19	H5N4S1	2189.87	2209.99	0.02	1.4E+05	1.0E+06	9%	19%	3%	1%
20	H5N4F1S1	2335.93	2356.05	-0.05	1.6E+04	1.9E+05	10%	36%	5%	6%
21	H5N2	1492.61	1512.74	0.33	1.6E+05	1.4E+06	11%	9%	2%	4%
22	H3N4F1	1720.73	1740.85	0.21	3.1E+05	1.8E+06	-12%	16%	1%	2%
23	H5N3	1695.69	1715.82	0.24	2.9E+04	3.4E+05	12%	24%	9%	1%
24	H4N4	1736.72	1756.85	0.21	1.2E+05	2.3E+05	13%	15%	1%	9%
25	H5N4S2	2480.96	2501.09	-0.11	8.7E+04	4.4E+05	13%	12%	3%	2%
26	H5N4F1S2	2627.02	2647.15	-0.18	2.5E+03	4.1E+04	14%	-30%	24%	7%
27	H4N3S1	1824.74	1844.86	0.18	2.3E+03	3.3E+04	16%	67%	13%	4%
28	H6N5	2263.91	2284.03	-0.02	4.1E+04	8.4E+04	-20%	16%	5%	2%
29	H6N5S2	2846.10	2866.22	-0.27	4.9E+03	1.2E+04	25%	57%	20%	3%
30	H7N6	2629.04	2649.16	-0.18	6.E+03	3.4E+04	-42%	20%	6%	3%
31	H5N5S1	2392.95	2413.07	-0.08	-	4.9E+04	-1%	-	-	7%
32	H9N2	2140.83	2160.95	0.06	-	3.7E+04	-2%	-	-	4%
33	H6N4	2060.83	2080.95	0.08	-	1.2E+05	5%	-	-	2%
34	H4N4S1	2027.82	2047.94	0.09	-	3.1E+04	6%	-	-	3%
35	H6N4S1	2351.92	2372.05	-0.05	-	2.8E+05	-29%	-	-	0%

Chapter 5

Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags



- Adapted with permission from (Zhong, X., **Chen, Z.**, Snovida, S., Liu, Y., Rogers, J. C., & Li, L. (2015). Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags. Analytical chemistry, 87(13), 6527-6534.) Copyright (2015) American Chemical Society.
- **Author contribution:** study was designed by Zhong, X., **Chen, Z.**, Rogers, J.C., Li, L.; experiment was performed by Zhong X., Chen, Z. and Snovida, S., Liu, Y.; manuscript was written by Zhong, X. and Chen, Z., and edited by Zhong, X., Chen, Z., and Li, L.

ABSTRACT

The recently developed carbonyl-reactive aminoxy tandem mass tag (aminoxyTMT) reagents enable multiplexed characterization and quantitative comparison of structurally complex glycans between different biological samples. Compared to some previously reported isotopic labeling strategies for glycans, the use of aminoxyTMT method features a simple labeling procedure, excellent labeling efficiency and reduced spectral complexity at MS^1 level. Presence of the tertiary amine functionality in the reporter region of the aminoxyTMT labels leads to increased ionization efficiency of the labeled glycans, thus improving ESI-MS detection sensitivity. The use of the labeling reagent also makes electrophoretic separation of the labeled neutral and acidic glycans feasible. In this work, we characterized the electrospray ionization (ESI) and collision induced dissociation (CID) behaviour of the aminoxyTMT-labeled neutral and sialyted glycans. For the high-mannose type of *N*-glycans and small sialyted oligosaccharides, CID fragmentation of $[M+Na+H]^{2+}$ provides the most informative MS^2 spectra for both quantitative and qualitative analysis. For complex type of *N*-glycans, MS^3 of the protonated $Y_1(H)$ ion provides an approach for accurate quantification. Online CE-ESI-MS/MS analyses of multiplexed aminoxyTMT-labeled human milk oligosaccharides (HMOs) and different types of *N*-glycans released from glycoprotein standards were demonstrated. Improved resolution and quantification accuracy of the labeled HMO isomers was achieved by coupling CE with traveling wave ion mobility (TWIM)-CID-MS/MS. *N*-glycans released from human serum protein digests were labeled with sixplex aminoxyTMT and subjected to CE-MS analysis, and good accuracy and reproducibility of relative quantification based on the reporter ion peak intensity were achieved through MS^3 of $Y_1(H)$ ion.

INTRODUCTION

Glycosylation of proteins is one of the most important post-translational modifications. Studies have shown that the glycan moieties on the glycoproteins play critical roles in structural modulation, and function as specific binding ligands for endogenous receptors or exogenous agents in many biological processes such as protein trafficking, cell-cell signalling, cellular adhesion[1, 2]. Changes in glycomic profiles have been linked to various diseases[3-7] including immunological disorders, cancer, and cardiovascular problems. These implications urge researchers to develop cutting-edge bioanalytical platforms for quantitative analysis of glycans which would facilitate the elucidation of the diverse biological roles of glycans and their roles in human diseases[8].

Quantitative analysis of native glycans remains challenging due to high complexity and diversity of glycan structures[9], the difficulty of synthesizing glycan standards, and the relatively low response in both optical and mass spectrometric detection methods. Several fluorescence detection methods based on reducing-end labeling strategies have been developed[10, 11] and are routinely applied in conjunction with liquid chromatography (LC)[12] or capillary electrophoresis (CE) separation[13-15]. Reductive amination is the most common method used to attach the glycans with chromophores, such as 2-aminobenzamide (2-AB)[16], 2-aminopyridine (2-PA)[17], 2-aminobenzoic acid (2-AA)[18], and 1-aminopyrene-3, 6, 8-trisulfonic acid (APTS)[19]. Although high separation efficiency and sensitive detection could be achieved with LC or CE combined with fluorescence or laser induced fluorescence detection[6, 10, 20], structural identification is largely dependent on the availability of standards, sequential enzymatic digestion, and highly reproducible LC retention times or CE migration times.

With the recent advances in mass spectrometry, a variety of MS-based glycan quantification strategies have been developed, which also allow more confident assignment of glycan compositions based on mass-to-charge ratios of the intact glycans and fragment ions generated by different types activation methods [21-24]. These include both chemical and metabolic labeling approaches[22]. Metabolic incorporation of ^{15}N into *N*- and *O*-linked glycans for quantitative glycomics has been demonstrated by Orlando et al, which involves the use of amide-^{15}N-Gln in the cell medium to provide the only source of nitrogen for biosynthesis of aminosugars[25]. While metabolic labeling is relatively more expensive and restricted to the investigation of limited types of biological samples, chemical labeling strategies can be applied to glycan samples from different sources. MS^1 precursor ion intensity-based relative quantification is realized by labeling glycans with 'heavy' and 'light' pairs of reagents. These strategies include permethylation of glycan samples with d_0- and d_3-CH_3I,[26] or $^{12}CH_3I$ and $^{13}CH_3I$,[27] enzymatic release of glycans in $H_2{}^{16}O$ and $H_2{}^{18}O$,[28] reductive amination with normal and isotopically encoded tags such as d_0- and d_4-PA[29], d_0 and d_4 2-AA[30], 2-$^{12}[C_6]$-AA and 2-$^{13}[C_6]$-AA[31], $^{12}[C_6]$ aniline and $^{13}[C_6]$-aniline[32], or even tetraplex isotopic tags[33], formation of glycan hydrazones with $^{12}[C_6]$ and $^{13}[C_6]$ 4-phenethylbenzohydrazide[34], or d_0 and d_5 Girard's reagent P[35]. Incorporation of stable isotopes through permethylation can lead to variable mass shift between the differentially labeled glycans, depending on the total number of -OH, -NH, -COOH groups in one glycan, and thus complicating the spectral interpretation. In contrast, isotopic labeling at the reducing end of the glycans results in constant mass shift between the heavy and light labeled pairs, however, spectral interference caused by overlapping of the isotopes of the differentially labeled glycans is observed if the mass difference induced by the isotopic labeling is not large enough.

Originated from the tandem mass tags (TMT) that have been widely applied in quantitative proteomics[36-38], several multiplex isobaric tags, which are usually composed of a reporter group and an aldehyde reactive group linked by a mass balance group, have been developed for high throughput relative quantification of glycans based on reporter ion intensities generated by MS/MS. Compared to isotopic labeling, the MS^1 spectra of isobarically labeled glycans are simplified because the differentially labeled glycans have the same mass. The isobaric aldehyde reactive tags (iARTs) were derived from the deuterium isobaric amine-reactive tags (DiARTs) by modifying the amino reactive group to a primary amine group for reductive amination[39]. Two versions of glycan-specific carbonyl-reactive TMT reagents with hydrazide and alkoxyamine reactive functionalities, which form hydrazones or oximes with reducing sugars, respectively, were evaluated by Hahne et al[40]. One of these tags, the aminoxyTMT that employs the alkoxyamine functional group, has recently become commercially available. The previous study indicated that under MALDI condition, mainly the monosodium adduct ions of the aminoxyTMT-labeled glycans were observed, and the low intensity of the reporter ions generated by post-source dissociation (PSD) compromised the accuracy and dynamic range of relative quantification[40]. However, the performance of the aminoxy-TMT labeled glycans in ESI mode, CID fragmentation of the multiply charged precursor ions, as well as the reporter ion intensity based quantification by ESI-CID MS/MS remain unexplored. In this work, we characterized the aminoxyTMT-labeled different types of glycans using an ESI-Q-ToF instrument and found that the CID MS/MS of multiply charged aminoxyTMT-labeled glycan generated sufficient reporter ions for accurate quantification.

Another aspect that has not been explored previously is the coupling of microscale separation platforms with ESI-MS detection for analysis of multiplex aminoxyTMT-labeled

glycans, which is critical for improving coverage of complex samples with limited amounts. The tertiary amine group in the aminoxyTMT makes the labeled glycan positively charged in acidic buffers and thus provides possibility of electrophoretic separation of the aminoxyTMT-labeled glycans. As an attractive separation technique alternative to hydrophilic interaction chromatography (HILIC) and porous graphitized carbon (PGC) liquid chromatography, CE offers advantages for separation of polar molecules including low-cost of column material, very small amount of sample consumption, fast separation speed and high resolution power. Accompanied by efforts towards developing robust and sensitive CE-ESI-MS over the past decades[41, 42], more and more researchers have incorporated this technique in their toolboxes for analysis of oligosaccharides in biological samples and glycans from different types of glycoconjugates[20, 43-46], such as human milk oligosaccharides[47], *N*-glycans from therapeutic recombinant glycoproteins[48-51], *N*-glycans released from serum[52, 53], oligosaccharides derived from glycosphingolipids[54]. Most of these CE-ESI-MS studies involve separation of native glycans carrying negatively charged functionalities[48, 52, 53, 55] or APTS-labeled glycans[47, 49-51, 53, 55, 56], which have to be coupled with negative ESI detection that is inherently less sensitive and more susceptible to corona discharge compared to positive ESI[57-59]. In contrast, the aminoxyTMT labeling promotes sensitivity in positive ESI of the labeled glycans by introducing hydrophobic and basic groups. Here, we demonstrated, for the first time, online coupling of CE and ESI-CID-MS/MS detection for the analysis of multiplex aminoxyTMT-labeled human milk oligosaccharide (HMO) standards, different types of *N*-glycans released from glycoprotein standards and human serum proteins. In addition, CE coupled with travelling wave form ion mobility mass spectrometry (TWIMS) was explored and proven to be effective of improving the resolution of small isomeric glycans and quantification accuracy. These results show significant promise of applying multiplex

aminoxyTMT labeling along with CE-ESI-MS/MS for high-throughput characterization of more complicated biological systems.

EXPERIMENTAL SECTION

Chemicals and Materials (see supporting information)

N-glycan release and aminoxyTMT labeling of glycans (see supporting information)

Direct infusion ESI-MS/MS and ESI-IMS-MS/MS (see supporting information)

CE-ESI-MS setup (see supporting information)

MS parameters for online CE-ESI-MS/MS and CE-ESI-IM-MS/MS analysis

For CE-MS/MS analysis of three-plex aminoxyTMT-labeled *N*-glycan released from RNase B, full MS scan over the range of *m/z* 400~1500 was acquired to generate the inclusion list for the following data-dependent analysis (DDA). Top 2 DDA with the inclusion list of all the $[M+H+Na]^{2+}$ and $[M+2H+K]^{3+}$ precursor ion *m/z* was performed in a separate CE-MS run using parameters as follows. The MS survey scan range was acquired over the range of *m/z* 400~1400 Da, scan duration was 1 second. +2 and +3 charged ions with intensity arising above a threshold of 1500 were selected for MS/MS scans. The precursor ion isolation window was ~3 *m/z*, MS/MS scan was acquired over the range of *m/z* 100~1900, and the scan duration was 1.5 seconds. The trap collision cell voltage profile was set as the following equation: $CE\ (V) = 25 + (\frac{m}{z} - 400) \times 0.055$. After one MS/MS scan, the current precursor was then excluded from the rest of the run.

For CE-ESI-IMS-MS/MS analysis of differentially labeled HMO standards, traveling wave velocity was set at 700 m/s, and wave height 40 V. Alternative full MS scan and targeted MS/MS

scan of m/z 661.9 were acquired during the CE separation. The CID MS/MS was performed in the transfer collision cell with collision energy of 50 V.

Alternative full MS scan and pseudo MS^3 scan events were applied for CE-MS/MS analysis of aminoxyTMT labeled *N*-glcans released from fetuin and human serum protein digests. The cone voltage was set at 30 V for 0.5 second of full MS scan in the range of *m/z* 400 ~1700, and switched to100 V for 0.5 second of pseudo-MS^3 scan during CE separation. For the pseudo-MS^3 scan,Y_1(H) ion m/z 523.3 generated in the ion source was selected by the quadrupole and fragmented in the trap collision cell with a collision energy of 40 V, and the TOF scan range was *m/z* 100~150.

RESULTS AND DISCUSSION

CID MS/MS fragmentation of aminoxyTMT-labeled glycan standards and relative quantification of isobarically labeled glucose oligomers (see supporting information)

CE-ESI-MS analysis of aminoxyTMT labeled high-mannose *N*-glycans

For improving the glycan coverage and resolving potentially existing isomeric species, a unique platform that couples microscale separation technique with MS detection is usually required for analysis of complex samples with small volume. CE-ESI-MS is a promising tool for analysis of aminoxyTMT labeled glycans because the modification introduces a tertiary amine group at the reducing end, making electrophoretic separation of neutral glycans feasible in acidic environment. When the tertiary amine on the reporter ion group is attached by a proton under acidic environment, the labeled carbohydrate molecule will possess overall positive charge and migrate towards the cathode end of the capillary under electric field.

CE separation of three-plex aminoxyTMT-labeled *N*-glycans released from RNase B (1:1:1 ratio) was conducted using methanol/water/formic acid (v/v/v) 50:49:1 as background electrolyte.

Figure 1 shows the EIEs of the $[M+H+Na]^{2+}$ ions of the $(Man)_{5\sim9}(GlcNAc)_2$. The five labeled high mannose *N*-glycans migrated out of the separation capillary in the order of increasing number of monosaccharide residues. The total amount of glycans loaded onto the column corresponded to the glycans released from approximately 0.2 μg of RNase B. By pooling together the differentially labeled glycans, ion intensities of both MS^1 and MS^2 scans were boosted because the isobarically labeled glycans were detected at the same precursor *m/z*, and most fragment ions (except the reporter ions) originated from different samples were also detected at the same *m/z*.

Similar as the glucose oligomers, the ESI generated $[M+Na+H]^{2+}$ precursor ions of the high-mannose *N*-glycans produce abundant reporter ions for quantification and diverse fragment ions for structural inference. The CID MS/MS spectrum of $[(Man)_5(GlcNAc)_2\text{-TMT}+Na+H]^{2+}$ ion (*m/z* 779.8) is shown in Figure 2 as an example. At trap collision cell voltage of 46 V, the reporter ions dominated the spectrum, and abundant fragment ions associated with the *N*-glycan structure were present as well. Except $Y_{1(H)}$ ion (*m/z* 523.3) and several doubly charged $Y_{(Na+H)}$ ions, most fragment ions were observed as sodium adducts. Sodiated $Y_{(Na)}$ ions (*m/z* 1396.6, 1234.5, 1072.5, 910.4) and doubly charged $Y_{(Na+H)}$ ions (*m/z* 698.8, 617.7) were formed by consecutive loss of mannose residues. $B_{2(Na)}$, $B_{3(Na)}$, $B_{4(Na)}$ ions (*m/z* 509.1, 833.2, 1036.3), as well as some $B/Y_{(Na)}$ ions (*m/z* 874.2, 671.2, 550.1, 347.1) produced from internal cleavage formed by different routes were observed. $C_{3(Na)}$ and $C_3\text{-}2H_{(Na)}$ (*m/z* 851.2, 849.2) ions, and $C_{4(Na)}$ and $C_4\text{-}2H_{(Na)}$ ions (*m/z* 1054.3, 1052.3) were formed by cleavage of the two β1-4 bonds of the chitobiose core respectively. Cross-ring cleavages of the GlcNAc at the reducing terminus produced ${}^{3,5}A_{4(Na)}$ (*m/z* 907.2), ${}^{2,4}A_{5(Na)}$ (*m/z* 1096.3), ${}^{0,2}A_{5(Na)}$ (*m/z* 1156.3), and internal fragment ion ${}^{0,2}A_5/Y_{3\beta(Na)}$ or ${}^{0,2}A_5/Y_{4\alpha(Na)}$ (*m/z* 944.3). The intensity of the $[M\text{-reporter ion-CO}+H+Na]^+$ ion was also prominent.

In addition to $[M+Na+H]^{2+}$, $[M+2H]^{2+}$ and $[M+2H+K]^{3+}$ precursor ions were also abundant in MS^1 scan and they could be subjected to CID MS/MS during the online data-dependant analysis. Figure S3 shows the CID MS/MS spectra of $[(Man)_6(GlcNAc)_2\text{-}TMT+2H]^{2+}$, $[(Man)_6(GlcNAc)_2\text{-}TMT+2H+K]^{3+}$, $[(Man)_6(GlcNAc)_2\text{-}TMT+Na+H]^{2+}$ acquired during CE-MS separation. Although the trap collision energy profile was optimized for $[M+Na+H]^{2+}$ precursor ions, $[M+2H]^{2+}$ and $[M+2H+K]^{3+}$ ions of the high-mannose *N*-glycans also produced decent amount of reporter ions in spite of few fragment ions in the high mass range. Table 1 summarizes the reporter ion intensity ratios calculated from the sum of the reporter ion intensity generated from CID MS/MS of $[M+Na+H]^{2+}$, $[M+2H]^{2+}$, $[M+2H+K]^{3+}$ precursor ions and the error calculated in this way was within 17% for the five high-mannose *N*-glycans.

CE-ESI-TWIMS-CID MS/MS analysis of aminoxyTMT labeled acidic oligosaccharide

Oligosaccharides containing acidic monosaccharide residues such as NeuAc and NeuGc become amphoteric compounds after being labeled with aminxoyTMT reagents. To study the CE separation behaviour of amnixoyTMT labeled acidic glycans, we selected four human milk oligosaccharides (sialyllacto-N-tetraose a, b, c, DSLNT) and subjected them to isobaric labeling and CE-ESI-MS analysis. With methanol/water/formic acid (v/v/v) 50:49:1 as background electrolyte, the aminoxyTMT labeled DSLNT could be separated from the sialyllacto-N-tetraoses because of its lower pKa and larger size, however, only two CE peaks corresponding to the three pentasaccharide isomers were observed. By further investigation, it was confirmed that the labeled sialyllacto-N-tetraose b (LSTb) and sialyllacto-N-tetraose c (LSTc) co-eluted at 34.6 min, followed by sialyllacto-N-tetraose a (LSTa) eluting at 38.1 min (Figure 3A). If the co-migrated isomeric compound ions were subjected to CID MS/MS simultaneously, the observed reporter ion

intensities would only reflect the total amounts of the co-eluted isomers instead of the relative quantity of each individual component in the samples.

Recent studies have shown promises of differentiation of isomeric carbohydrate as metal ion adducts by traveling wave ion mobility mass spectrometry[60, 61]. Taking advantage of the TWIM capability of the Waters Syanpt G2 Q-ToF instrument, we were able to further improve the resolution of aminoxyTMT labeled isomeric HMO standards by integrating a second separation dimension, the ion mobility separation in the gas phase, with CE separation. Although both CE and IMS are electrophoretic based separation techniques, they could be orthogonal separation dimensions because the aqueous phase mobility and gas phase mobility of an analyte molecule are dependent on different parameters that affect 'charge-to-shape ratio' of the molecule in specific environment. In terms of an amphoteric compound during CE separation, the pKa values of the compound and pH of the BGE define the 'average charge' carried by an ensemble of analyte molecules which co-migrated in the aqueous phase and the 'shape' of the molecule refers to the hydrated aqueous phase conformation of the molecule in the buffer. After the analyte molecules are converted to gas-phase ions by electrospray ionization, the number of H^+, Na^+, K^+ and other charge carriers attached to one analyte molecule defines the charges carried by the single analyte molecule ions, and the 'shape' of one analyte ion refers to the gas phase conformation of a specific type of ion adduct of the analyte molecule. As different types of ion adducts of glycans are usually observed under positive ESI mode, we first compared the IMS separation of $[M+H+Na]^{2+}$, $[M+H+K]^{2+}$, $[M+2Na]^{2+}$, $[M+K+Na]^{2+}$ adducts of the three aminoxyTMT labeled isomeric pentasaccharides at a set of DC traveling wave velocity / wave height combinations by direct infusion experiments. Figure S4 shows the arrival time distributions of the labeled LSTa, LSTb and LSTc as different types of ion adduct at wave velocity of 700 m/s and wave height of 40 V.

Interestingly, the $[M+H+Na]^{2+}$ adduct ion of amixoxyTMT labeled LSTc could be differentiated from those of LSTa and LSTb by TWIMS, which provides complementary results to CE separation.

As a demonstration of coupling CE with TWIMS-CID-MS/MS for improved relative quantification accuracy of the aminoxyTMT labeled isomeric oligosaccharides, LSTa , LSTb, LSTc were differentially labeled with aminoxyTMT^6-128 and aminoxyTMT^6-131 in 1:4, 1:1, 3:1 ratio respectively, and mixed together with aminoxyTMT labeled DSNLT, followed by CE-ESI-TWIM-CID-MS/MS analysis. With the optimized CE and IMS conditions, the coelutingaminoxyTMT labeled LSTb and LSTc at 34.6 min were resolved in the ion mobility drift cell, with the drift time of the $[M+Na+H]^{2+}$ ion adducts at 4.35 ms and 4.83 ms respectively (Figure 3B, top panel). It was also noticed the in-source fragmentation produced of $Y_{4\alpha(H+Na)}$ and $Y_{3\beta(H+Na)}$ ions of labeled DSLNT have the same arrival time distribution as the $[M+Na+H]^{2+}$ ions of the earlier eluted aminoxyTMT labeled LSTa and LSTb (Figure 3B). After the two dimensional separation, the three differentially labeled isomers were almost completely resolved and the $[M+Na+H]^{2+}$ (*m/z* 661.9) precursor ions from each isomer were subjected to CID MS/MS in the transfer collision cell separately. The reporter ion intensities obtained at three different CE migration time/IM drift time combinations, shown in Figure 3C, represent the relative quantities of each individual pentasacchride isomers; and the observed reporter ion ratios were close to the expected values because of the improved separation.

CE-ESI- pseudo-MS^3 analysis of aminoxyTMT labeled sialyted complex *N*-glycans

Complex type of *N*-glycans released from bovine fetuin, which contain terminal sialic acid residues, were also investigated by CID MS/MS and CE-MS after being labeled with

aminoxyTMT reagents. By direction infusion CID MS/MS experiments, it was found that the reporter ion yield of the aminoxyTMT labeled complex type of *N*-glycans are generally not as good as that of the high-mannose type of *N*-glycans because the GlcNAc fragment ion *m/z* 138.06 always dominates the CID MS/MS spectra. The representative MS/MS spectra of different ion adducts of the triantennary complex *N*-glycan $(Hex)_6(GlcNAc)_5(NeuAc)_3$ is shown in Figure S5. In order to overcome this limitation, we propose a pseudo-MS^3 based quantification method here for CE-MS applications on Q-TOF MS platform. Since acidic BGE and modifier are used for CE-MS and a tertiary amine group is incorporated into the aminoxyTMT reagent, abundant $Y_1(H)$ ions (*m/z* 523.3), which is common for all types of *N*-glycans labeled with aminoxyTMT reagents, can be easily generated in the ion source from the fully and partially protonated precursor ion adducts by raising the sample cone voltage and then subjected to CID MS/MS in the collision cell for production of reporter ions. In order to obtain both intact glycan mass and the decent reporter ion intensities during online separation, two scan events, full MS scan at low cone voltage (30 V) and targeted MS/MS scan of $Y_1(H)$ ion at high cone voltage (100 V) were acquired alternatively during one CE-MS run.

The pseudo-MS^3 method was applied to the aminoxyTMT^6-128,131 (1:1) labeled *N*-glycans released from bovine fetuin. As shown in Figure 4A, the labeled *N*-glycans from fetuin migrated in the order of neutral, monosialyted, disialyted, trisialyted, tetrasialyted, and pentasialyted glycans (TIE from full MS scans, top panel); the EIE of *m/z* 523.3 (bottom panel) corresponding to the $Y_1(H)$ generated by in-source fragmentation at low cone voltage defined the migration zone of aminoxyTMT labeled *N*-glycans. Figure 4B present the EIEs of the two reporter ions, *m/z* 128.1 and *m/z* 131.1, generated from targeted MS/MS scans of *m/z* 523.3 at high cone voltage (pseudo-MS^3 scans). The migration peak profiles of the two ion channels were almost identical and the ion

intensity ratios in glycan migration zone were close to 1:1. Despite that the pseudo-MS^3 method is only ideal for highly efficient separation techniques coupling with MS and the generated reporter ions intensities have to be deconvoluted in case of coeluting and partially resolved components, this issue can be alleviated if the real MS^3 method, isolating a specific aminoxyTMT labeled precursor ion to produce the common $Y_1(H)$ and $Y_2(H)$ ions and submitting them to further fragmentation, is applied on an MS platform with trap-type of mass analyzer such as LTQ-Orbitrap.

Multiple isomeric structures associated with the same *N*-glycan composition occur frequently in this sample. After examining and excluding the in-source fragmentation generated components, a total of 10 sialyted *N*-glycan compositions, containing at least 23 glycan structures, were putatively identified from the CE-MS experiments (Figure S6). Two representative extracted ion electropherograms, corresponding to the $[M+3H]^{3+}$ ions of the biantennary complex type glycan $(Hex)_5(GlcNAc)_4(NeuAc)_2$ and the triantennary complex type glycan $(Hex)_6(GlcNAc)_5(NeuAc)_3$ respectively, are shown in Figure 4C and 4D. Four isomeric structures associated with the composition $(Hex)_5(GlcNAc)_4(NeuAc)_2$ were resolved by CE and at least three isomeric structures associated with the composition $(Hex)_6(GlcNAc)_5(NeuAc)_3$ were detected. Although efforts have been made to improve the *N*-glycan isomer separation by employing TWIM as the second separation dimension, the limited resolution of TWIM, the more complicated tertiary structures of multiply charged large gas phase *N*-glycan ions and the multiple conformers generated from one isomer ion, are still barriers for application of CE-ESI-IM-MS on *N*-glycan samples.

CE-ESI-pseudo-MS^3 analysis of sixplex aminoxyTMT labeled *N*-glycans released from human serum proteins

N-glycans released from human serum proteins labeled with sixplex aminoxyTMT^6 -126, 127, 128, 129, 131 in a 1:2:1:1:1:2 ratio were used as a complex biological sample to assess the developed platform for glycan quantification. Data-dependant analysis was performed during online separation first to inspect interfering ions in the reporter ion region. Figure S7 shows the MS/MS spectra of $[M+2H]^{2+}$, $[M+2H+K]^{2+}$, $[M+H+Na]^{2+}$ ion adducts of the core-fucosylated complex *N*-glycan $(Hex)_3(GlcNAc)_4Fuc$. It was found that, the reporter ion yield of neutral complex type of *N*-glycans was also low and the HexNAc fragment ion of the formula $[C_6H_8O_2N +H]^+$,[40] *m/z* 126.06 and its isotope *m/z* 127.06 could be just resolved from the two reporter ions *m/z* 126.13 and *m/z* 127. 13 by the TOF-MS under resolution mode. However, the pseudo-MS^3 method is still applicable for the core-fucosylated *N*-glycans since the $Y_1Y_{1\delta}(H)$ ion *m/z* 523.3 was also observed when the protonated precursor ions were activated. The TIE of the sixplex aminxoyTMT labeled human serum *N*-glycans is shown in Figure 5. The hydrophilic peptides retained in the glycan fraction during sample purification migrated out of the capillary column first, followed by the aminoxyTMT labeled neutral *N*-glycans, monosialyted, disialyted, trisialyted, and tetrasialyted *N*-glycans. 46 *N*-glycan compositions were identified within 30 ppm accuracy of at least two precursor adduct ions belonging to the same composition. By manually excluding the in-source fragmentation peaks, 84 isomeric structures belonging to the 46 *N*-glycan compositions were putatively identified (Table S1). The inset of Figure 5 shows the pseudo MS^3 spectra acquired at during the migration time zones of the sixplex aminoxyTMT labeled $(Hex)_3(GlcNAc)_4Fuc$, $(Hex)_5(GlcNAc)_4(NeuAc)$, $(Hex)_5(GlcNAc)_4$ $NeuAc)_2$, $(Hex)_6(GlcNAc)_5$ $(NeuAc)_3$ respectively, which were free from the spectral interference of the fragment ions of GlcNAc residues. The corrected reporter ion intensity ratios of the four spectra are summarized in Table S2.

CONCLUSIONS

Combining the isobaric labeling strategy and CE-ESI-MS/MS, the aminoxyTMT-labeled neutral and acidic glycans can be efficiently separated in acidic background electrolytes and quantified with CID MS/MS or pseudo-MS^3 method (MS/MS of Y_1(H) ion). CE coupled with TWIM-CID MS/MS was demonstrated to be efficient in resolving human milk oligosaccharide isomers and improving the relative quantification accuracy. The labeled sialyted isomeric *N*-glycans can also be resolved by CE separation in normal polarity. For the high-mannose type of *N*-glycans, online data-dependent analysis of the precursor ion adducts $[M+H+Na]^{2+}$,$[M+2H]^{2+}$, $[M+K+2H]^{3+}$ could give rise to high-intensity reporter ions for accurate relative quantification between different samples, and CID MS/MS of $[M+H+Na]^{2+}$ ions generate most useful tandem mass spectra in terms of abundant reporter ions and rich the fragment ions for structural inference. For the complex type of *N*-glycans, MS^3 method could be applied to generate reporter ions avoiding interference from the fragment ions of HexNAc residues. Future work will extend the application of these aminoxyTMT tags to more complex biological systems for high throughput glycan biomarker discovery.

ACKNOWLEDGEMENTS

The authors wish to thank Professor David D.Y. Chen at Department of Chemistry at the University of British Columbia for generously providing the CE-ESI-MS interface unit. Our thanks also go to Gary Girdaukas and staff members of the machine shop of Department of Chemistry at the University of Wisconsin-Madison, for their assistance with the CE and ESI source instrumentation. This work was supported by grants from NIH R01DK071801, R56DK071801, and S10RR029531.

REFERENCES

1. *Biological Roles of Glycans, Essentials of Glycobiology*. 2nd Edition ed.; Cold Spring Harbor Laboratories Press: NY, 2009.
2. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. *Nature reviews. Molecular cell biology* **2012,** 13, (7), 448-62.
3. Jankovic, M., GLYCANS AS BIOMARKERS: STATUS AND PERSPECTIVES. *J. Med. Biochem.* **2011,** 30, (3), 213-223.
4. Taniguchi, N., Human Disease Glycomics/Proteome Initiative (HGPI). *Mol Cell Proteomics* **2008,** 7, (3), 626-627.
5. An, H. J.; Kronewitter, S. R.; de Leoz, M. L. A.; Lebrilla, C. B., Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **2009,** 13, (5-6), 601-607.
6. Alley, W. R.; Mann, B. F.; Novotny, M. V., High-sensitivity Analytical Approaches for the Structural Characterization of Glycoproteins. *Chem Rev* **2013,** 113, (4), 2668-2732.
7. Arnold, J. N.; Saldova, R.; Hamid, U. M. A.; Rudd, P. M., Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **2008,** 8, (16), 3284-3293.
8. Hart, G. W.; Copeland, R. J., Glycomics Hits the Big Time. *Cell* **2010,** 143, (5), 672-676.
9. 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* **2010,** 6, (10), 713-723.
10. Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M., Glycan labeling strategies and their use in identification and quantification. *Anal Bioanal Chem* **2010,** 397, (8), 3457-3481.
11. Harvey, D. J., Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2011,** 879, (17-18), 1196-225.
12. Anumula, K. R., Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Anal Biochem* **2006,** 350, (1), 1-23.
13. El Rassi, Z., Recent developments in capillary electrophoresis and capillary electrochromatography of carbohydrate species. *Electrophoresis* **1999,** 20, (15-16), 3134-44.
14. El Rassi, Z., Recent developments in capillary electrophoresis of carbohydrate species. *Electrophoresis* **1997,** 18, (12-13), 2400-7.
15. Suzuki, S.; Honda, S., A tabulated review of capillary electrophoresis of carbohydrates. *Electrophoresis* **1998,** 19, (15), 2539-60.
16. Bigge, J. C.; Patel, T. P.; Bruce, J. A.; Goulding, P. N.; Charles, S. M.; Parekh, R. B., Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal Biochem* **1995,** 230, (2), 229-38.
17. Hase, S.; Hara, S.; Matsushima, Y., Tagging of sugars with a fluorescent compound, 2-aminopyridine. *Journal of biochemistry* **1979,** 85, (1), 217-20.
18. Anumula, K. R., Quantitative determination of monosaccharides in glycoproteins by high-performance liquid chromatography with highly sensitive fluorescence detection. *Anal Biochem* **1994,** 220, (2), 275-83.
19. Chen, F. T.; Evangelista, R. A., Analysis of mono- and oligosaccharide isomers derivatized with 9-aminopyrene-1,4,6-trisulfonate by capillary electrophoresis with laser-induced fluorescence. *Anal Biochem* **1995,** 230, (2), 273-80.

20. Mechref, Y.; Novotny, M. V., Glycomic Analysis by Capillary Electrophoresis-Mass Spectrometry. *Mass Spectrom Rev* **2009,** 28, (2), 207-222.
21. Zaia, J., Mass Spectrometry and the Emerging Field of Glycomics. *Chem Biol* **2008,** 15, (9), 881-892.
22. Mechref, Y.; Hu, Y. L.; Desantos-Garcia, J. L.; Hussein, A.; Tang, H. X., Quantitative Glycomics Strategies. *Mol Cell Proteomics* **2013,** 12, (4), 874-884.
23. Kailemia, M. J.; Ruhaak, L. R.; Lebrilla, C. B.; Amster, I. J., Oligosaccharide Analysis by Mass Spectrometry: A Review of Recent Developments. *Anal Chem* **2014,** 86, (1), 196-212.
24. Alley, W. R.; Novotny, M. V., Structural Glycomic Analyses at High Sensitivity: A Decade of Progress. *Annu Rev Anal Chem* **2013,** 6, 237-265.
25. Orlando, R.; Lim, J. M.; Atwood, J. A.; Angel, P. M.; Fang, M.; Aoki, K.; Alvarez-Manilla, G.; Moremen, K. W.; York, W. S.; Tiemeyer, M.; Pierce, M.; Dalton, S.; Wells, L., IDAWG: Metabolic Incorporation of Stable Isotope Labels for Quantitative Glycomics of Cultured Cells. *J Proteome Res* **2009,** 8, (8), 3816-3823.
26. Kang, P.; Mechref, Y.; Kyselova, Z.; Goetz, J. A.; Novotny, M. V., Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal Chem* **2007,** 79, (16), 6064-6073.
27. Aoki, K.; Perlman, M.; Lim, J. M.; Cantu, R.; Wells, L.; Tiemeyer, M., Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* **2007,** 282, (12), 9127-9142.
28. Zhang, W.; Wang, H.; Tang, H. L.; Yang, P. Y., Endoglycosidase-Mediated Incorporation of O-18 into Glycans for Relative Glycan Quantitation. *Anal Chem* **2011,** 83, (12), 4975-4981.
29. Yuan, J.; Hashii, N.; Kawasaki, N.; Itoh, S.; Kawanishi, T.; Hayakawa, T., Isotope tag method for quantitative analysis of carbohydrates by liquid chromatography-mass spectrometry. *J. Chromatogr. A* **2005,** 1067, (1-2), 145-152.
30. Hitchcock, A. M.; Costello, C. E.; Zaia, J., Glycoform quantification of chondroitin/dermatan sulfate using a liquid chromatography-tandem mass spectrometry platform. *Biochemistry* **2006,** 45, (7), 2350-2361.
31. Prien, J. M.; Prater, B. D.; Qin, Q.; Cockrill, S. L., Mass Spectrometric-Based Stable Isotopic 2-Aminobenzoic Acid Glycan Mapping for Rapid Glycan Screening of Biotherapeutics. *Anal Chem* **2010,** 82, (4), 1498-1508.
32. Lawrence, R.; Olson, S. K.; Steele, R. E.; Wang, L. C.; Warrior, R.; Cummings, R. D.; Esko, J. D., Evolutionary Differences in Glycosaminoglycan Fine Structure Detected by Quantitative Glycan Reductive Isotope Labeling. *J Biol Chem* **2008,** 283, (48), 33674-33684.
33. Bowman, M. J.; Zaia, J., Comparative Glycomics Using a Tetraplex Stable-Isotope Coded Tag. *Anal Chem* **2010,** 82, (7), 3023-3031.
34. Walker, S. H.; Budhathoki-Uprety, J.; Novak, B. M.; Muddiman, D. C., Stable-Isotope Labeled Hydrophobic Hydrazide Reagents for the Relative Quantification of N-Linked Glycans by Electrospray Ionization Mass Spectrometry. *Anal Chem* **2011,** 83, (17), 6738-6745.
35. Wang, C.; Wu, Z.; Yuan, J.; Wang, B.; Zhang, P.; Zhang, Y.; Wang, Z.; Huang, L., Simplified Quantitative Glycomics Using the Stable Isotope Label Girard's Reagent P by Electrospray Ionization Mass Spectrometry. *J Proteome Res* **2014,** 13, (2), 372-384.
36. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C., Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003,** 75, (8), 1895-1904.

37. Xiang, F.; Ye, H.; Chen, R. B.; Fu, Q.; Li, L. J., N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Anal Chem* **2010,** 82, (7), 2817-2825.
38. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** 3, (12), 1154-1169.
39. Yang, S.; Yuan, W.; Yang, W. M.; Zhou, J. Y.; Harlan, R.; Edwards, J.; Li, S. W.; Zhang, H., Glycan Analysis by Isobaric Aldehyde Reactive Tags and Mass Spectrometry. *Anal Chem* **2013,** 85, (17), 8188-8195.
40. Hahne, H.; Neubert, P.; Kuhn, K.; Etienne, C.; Bomgarden, R.; Rogers, J. C.; Kuster, B., Carbonyl-Reactive Tandem Mass Tags for the Proteome-Wide Quantification of N-Linked Glycans. *Anal Chem* **2012,** 84, (8), 3716-3724.
41. Maxwell, E. J.; Chen, D. D., Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry. *Analytica chimica acta* **2008,** 627, (1), 25-33.
42. Zhong, X.; Zhang, Z.; Jiang, S.; Li, L., Recent advances in coupling capillary electrophoresis-based separation techniques to ESI and MALDI-MS. *Electrophoresis* **2014,** 35, (9), 1214-25.
43. Mittermayr, S.; Bones, J.; Guttman, A., Unraveling the Glyco-Puzzle: Glycan Structure Identification by Capillary Electrophoresis. *Anal Chem* **2013,** 85, (9), 4228-4238.
44. Zaia, J., Capillary electrophoresis-mass spectrometry of carbohydrates. *Methods in molecular biology (Clifton, N.J.)* **2013,** 984, 13-25.
45. Zhao, S. S.; Zhong, X.; Tie, C.; Chen, D. D., Capillary electrophoresis-mass spectrometry for analysis of complex samples. *Proteomics* **2012,** 12, (19-20), 2991-3012.
46. Mechref, Y., Analysis of glycans derived from glycoconjugates by capillary electrophoresis-mass spectrometry. *Electrophoresis* **2011,** 32, (24), 3467-81.
47. Albrecht, S.; Schols, H. A.; van den Heuvel, E. G. H. M.; Voragen, A. G. J.; Gruppen, H., CE-LIF-MSn profiling of oligosaccharides in human milk and feces of breast-fed babies. *Electrophoresis* **2010,** 31, (7), 1264-1273.
48. Jayo, R. G.; Thaysen-Andersen, M.; Lindenburg, P. W.; Haselberg, R.; Hankemeier, T.; Ramautar, R.; Chen, D. D. Y., Simple Capillary Electrophoresis-Mass Spectrometry Method for Complex Glycan Analysis Using a Flow-Through Microvial Interface. *Anal Chem* **2014,** 86, (13), 6479-6486.
49. Liu, Y.; Salas-Solano, O.; Gennaro, L. A., Investigation of sample preparation artifacts formed during the enzymatic release of N-linked glycans prior to analysis by capillary electrophoresis. *Anal Chem* **2009,** 81, (16), 6823-9.
50. Gennaro, L. A.; Salas-Solano, O., On-line CE-LIF-MS technology for the direct characterization of N-linked glycans from therapeutic antibodies. *Anal Chem* **2008,** 80, (10), 3838-45.
51. Bunz, S. C.; Rapp, E.; Neususs, C., Capillary Electrophoresis/Mass Spectrometry of APTS-Labeled Glycans for the Identification of Unknown Glycan Species in Capillary Electrophoresis/Laser-Induced Fluorescence Systems. *Anal Chem* **2013,** 85, (21), 10218-10224.

52. Liu, X.; Afonso, L.; Altman, E.; Johnson, S.; Brown, L.; Li, J., O-acetylation of sialic acids in N-glycans of Atlantic salmon (Salmo salar) serum is altered by handling stress. *Proteomics* **2008,** 8, (14), 2849-57.
53. Jayo, R. G.; Li, J. J.; Chen, D. D. Y., Capillary Electrophoresis Mass Spectrometry for the Characterization of O-Acetylated N-Glycans from Fish Serum. *Anal Chem* **2012,** 84, (20), 8756-8762.
54. Ito, E.; Nakajima, K.; Waki, H.; Miseki, K.; Shimada, T.; Sato, T. A.; Kakehi, K.; Suzuki, M.; Taniguchi, N.; Suzuki, A., Structural characterization of pyridylaminated oligosaccharides derived from neutral glycosphingolipids by high-sensitivity capillary electrophoresis-mass spectrometry. *Anal Chem* **2013,** 85, (16), 7859-65.
55. Bunz, S. C.; Cutillo, F.; Neususs, C., Analysis of native and APTS-labeled N-glycans by capillary electrophoresis/time-of-flight mass spectrometry. *Anal Bioanal Chem* **2013,** 405, (25), 8277-8284.
56. Maxwell, E. J.; Ratnayake, C.; Jayo, R.; Zhong, X. F.; Chen, D. D. Y., A promising capillary electrophoresis-electrospray ionization-mass spectrometry method for carbohydrate analysis. *Electrophoresis* **2011,** 32, (16), 2161-2166.
57. Hiraoka, K.; Kudaka, I., Negative-Mode Electrospray-Mass Spectrometry Using Nonaqueous Solvents. *Rapid Commun Mass Sp* **1992,** 6, (4), 265-268.
58. Cole, R. B.; Harrata, A. K., Solvent effect on analyte charge state, signal intensity, and stability in negative ion electrospray mass spectrometry; implications for the mechanism of negative ion formation. *J Am Soc Mass Spectrom* **1993,** 4, (7), 546-56.
59. Yamashita, M.; Fenn, J. B., NEGATIVE-ION PRODUCTION WITH THE ELECTROSPRAY ION-SOURCE. *J. Phys. Chem.* **1984,** 88, (20), 4671-4675.
60. Huang, Y.; Dodds, E. D., Ion mobility studies of carbohydrates as group I adducts: isomer specific collisional cross section dependence on metal ion radius. *Anal Chem* **2013,** 85, (20), 9728-35.
61. Pagel, K.; Harvey, D. J., Ion mobility-mass spectrometry of complex carbohydrates: collision cross sections of sodiated N-linked glycans. *Anal Chem* **2013,** 85, (10), 5138-45.
62. Domon, B.; Costello, C. E., A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-Ms Ms Spectra of Glycoconjugates. *Glycoconjugate J* **1988,** 5, (4), 397-409.
63. Maxwell, E. J.; Zhong, X. F.; Chen, D. D. Y., Asymmetrical Emitter Geometries for Increased Range of Stable Electrospray Flow Rates. *Anal Chem* **2010,** 82, (20), 8377-8381.
64. Maxwell, E. J.; Zhong, X. F.; Zhang, H.; van Zeijl, N.; Chen, D. D. Y., Decoupling CE and ESI for a more robust interface with MS. *Electrophoresis* **2010,** 31, (7), 1130-1137.
65. Zhong, X. F.; Maxwell, E. J.; Chen, D. D. Y., Mass Transport in a Micro Flow-Through Vial of a Junction-at-the-Tip Capillary Electrophoresis-Mass Spectrometry Interface. *Anal Chem* **2011,** 83, (12), 4916-4923.
66. Harvey, D. J., Collision-induced fragmentation of underivatized N-linked carbohydrates ionized by electrospray. *J Mass Spectrom* **2000,** 35, (10), 1178-1190.
67. Harvey, D. J.; Bateman, R. H.; Green, M. R., High-energy collision-induced fragmentation of complex oligosaccharides ionized by matrix-assisted laser desorption/ionization mass spectrometry. *J Mass Spectrom* **1997,** 32, (2), 167-187.

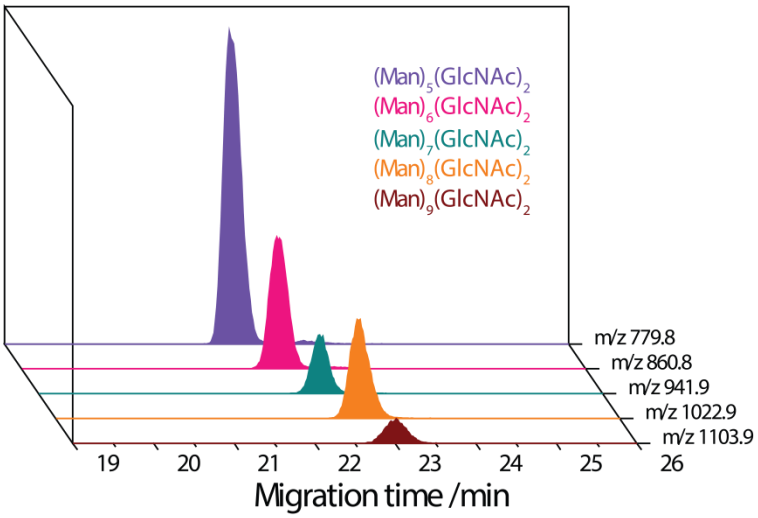


Figure 1. Extracted ion electropherograms of $[M+H+Na]^{2+}$ ions of aminoxyTMT-labeled *N*-glycans released from a glycoprotein RNase B. BGE, 50:49:1 (v/v/v) MeOH/H_2O/formic acid; modifier solution, 50:49:0.2 (v/v/v) MeOH/H_2O/formic acid. Full MS scan was acquired from *m/z* 400 to1500.

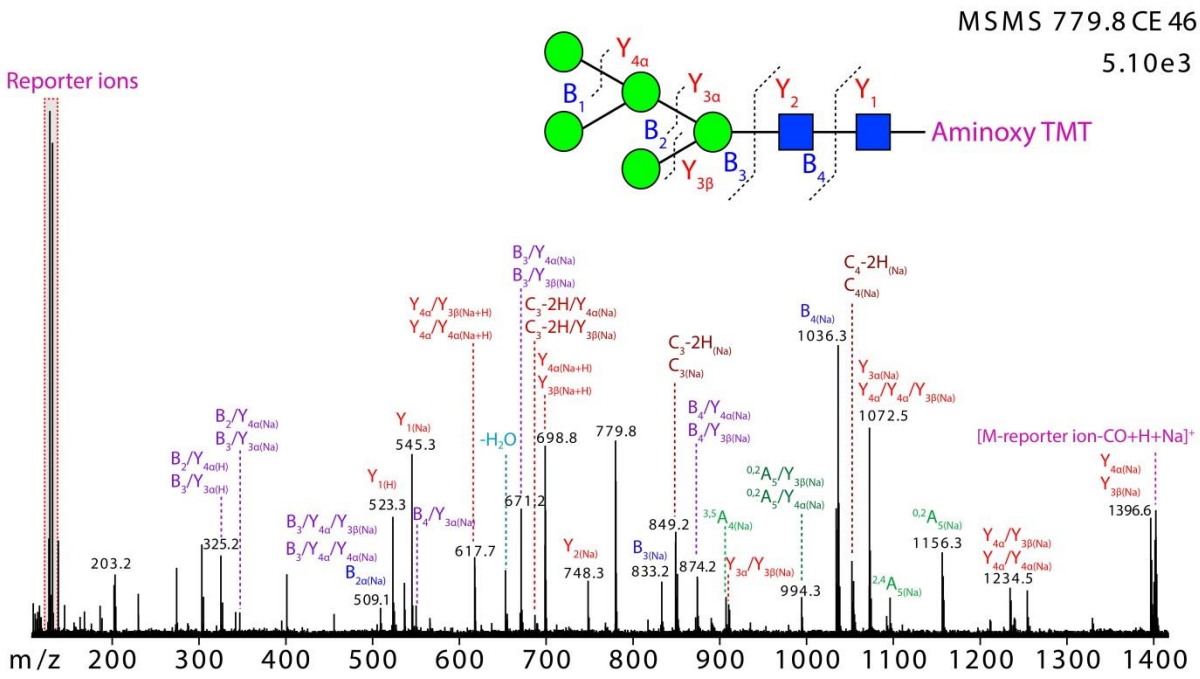


Figure 2. CID MS/MS spectrum of $[(Man)_5(GlcNAc)_2\text{-aminoxyTMT+Na+H}]^{2+}$ *m/z* 779.8 acquired by data-dependent analysis during online CE separation. Domon and Costello nomenclature[62] is used for annotation of fragment ions in this and following figures.

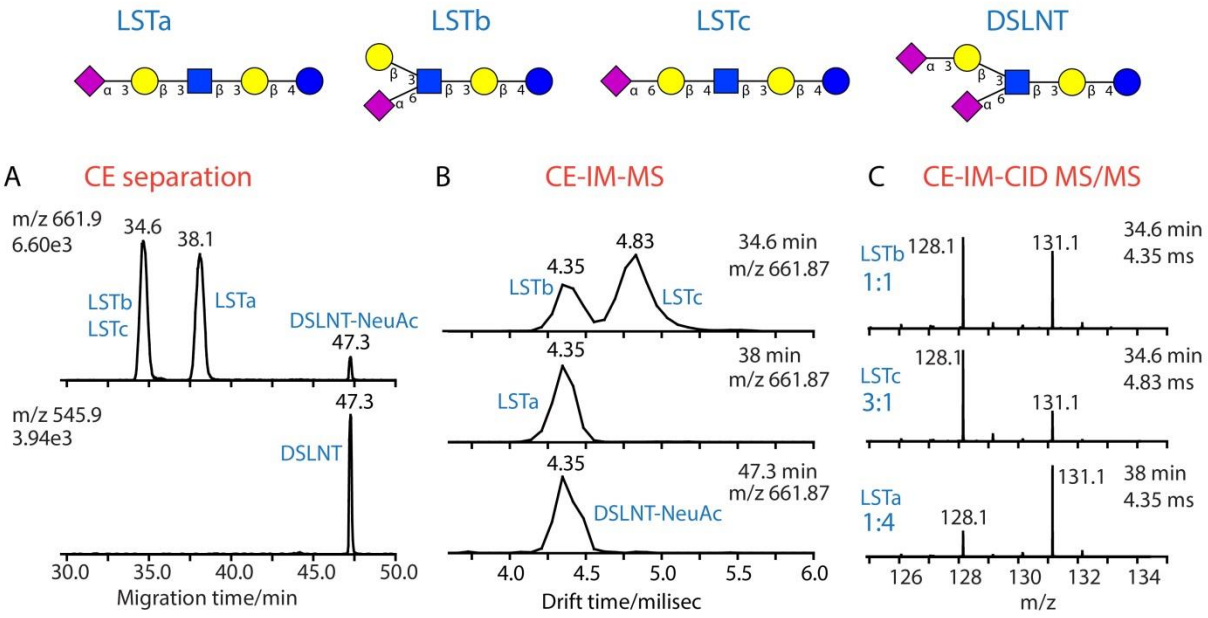


Figure 3. CE-ESI-IM-MS/MS analysis of aminxoyTMT^6-128 and aminoxyTMT^6-131 differentially labeled LSTa (128/131, 1:4), LSTb (128/131, 1:1), LSTc (128/131, 3:1) and DSLNT mixture. (A) Extracted ion electropherograms of [LST-aminoxyTMT+H+Na]$^{2+}$ *m/z* 661.9 (top) and [DLST-aminoxyTMT+H+2Na]$^{2+}$ *m/z* 545.9 (bottom). The peak at 47.3 min in the top pannel corresponds to in-source fragmentation ion $Y_{4\alpha}$(Na+H) or $Y_{3\beta}$(Na+H) of labeled DSLNT. (B) Arrival time distributions of [LST-aminoxyTMT+H+Na]$^{2+}$ *m/z* 661.9 at CE migration time of 34.6 min (top), 38 min (middle) and 47.3 min (bottom). (C) Reporter ion ratios of aminoxyTMT labeled LSTb (top), LSTc (middle), LSTa (bottom) obtained by CE-IM-targeted CID MS/MS. The precursor ion [LST-aminoxyTMT+H+Na]$^{2+}$ *m/z* 661.9 was isolated by quadrupole, separated in the ion mobility drift cell and then subjected to CID MS/MS in the transfer collision cell. CE conditions were as described in Figure 1. Alternative IM-full MS scan and IM-targeted MS/MS scan of *m/z* 661.9 were acquired during CE separation. IMS conditions: trap bias voltage was set at 48 V, helium cell DC 35 V, wave velocity 700 m/s, wave height 40 V. Transfer cell collision energy was set at 50V for targeted MS/MS.

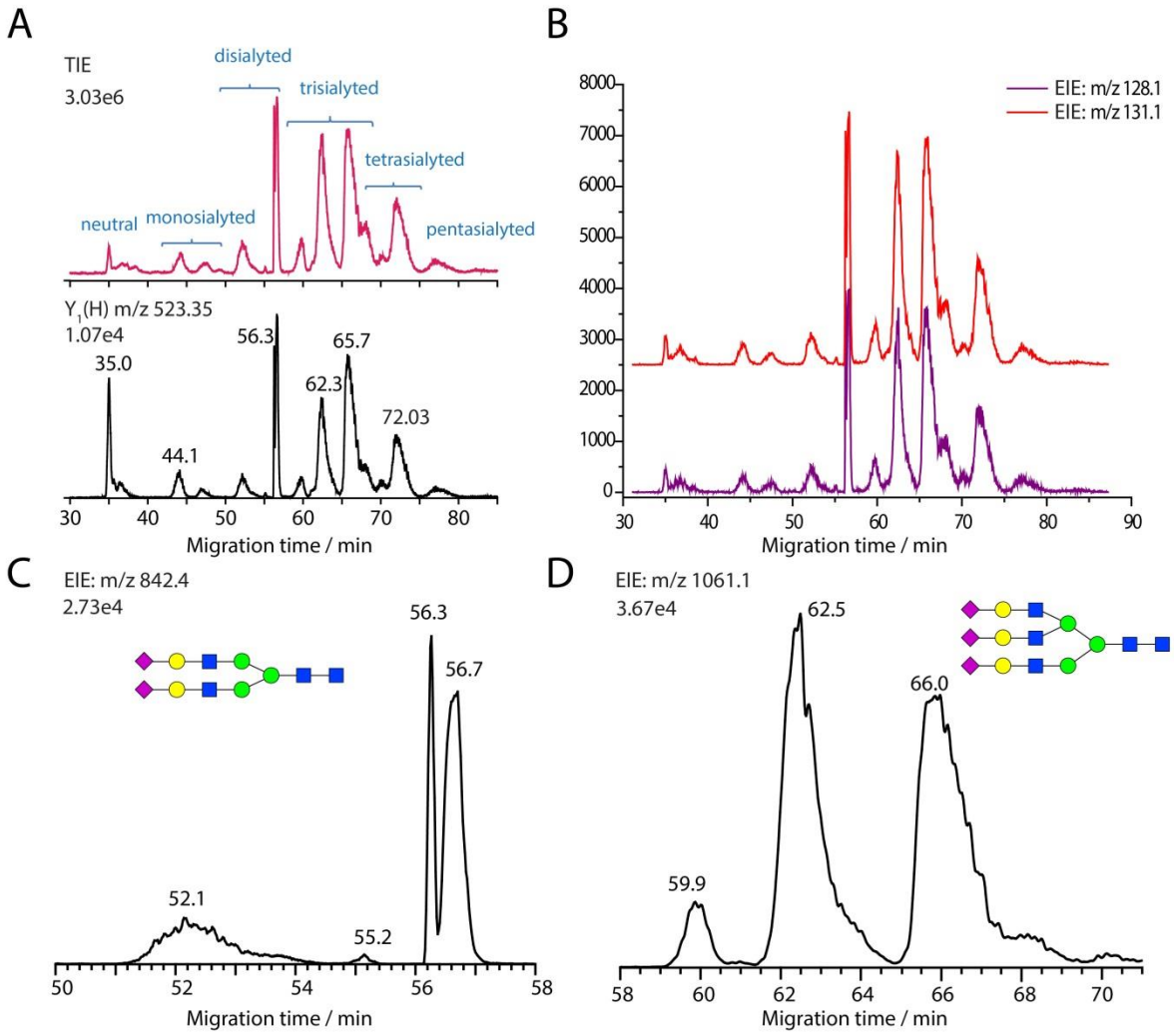


Figure 4. CE-ESI-MS of aminoxyTMT^6-128 and aminoxyTMT^6-130 (1:1) labeled *N*-glycans released from bovine fetuin. (A) Total ion electropherogram (top) and extracted ion electropherogram of Y_1(H) ion *m/z* 523.3 generated by in-source fragmentation. Cone voltage was set at 30 V for full MS scan and scan range was *m/z* 400 ~1700. (B) Extracted ion electropherograms of reporter ions *m/z* 128.1 and *m/z* 131.1 generated by pseudo-MS^3 scans. Cone voltage was set at 100 V, Y_1(H) *m/z* 523.3 was isolated and fragmented in the trap collision cell with collision energy of 40 V, scan range was m/z 100~150. (C) Extracted ion electropherogram of $[Hex_5GlcNAc_4NeuAc_2+3H]^{3+}$ *m/z* 842.4. (D) Extracted ion electropherogram of $[Hex_6GlcNAc_5NeuAc_3+3H]^{3+}$ *m/z* 1061.1.

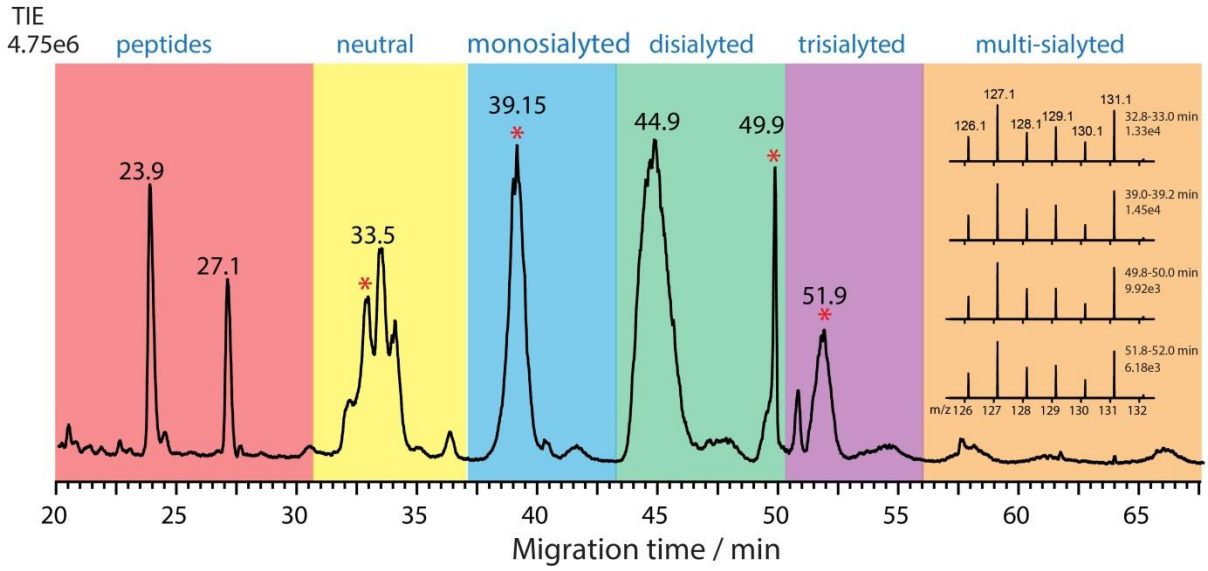


Figure 5. Total ion electropherogram of six-plex aminoxyTMT labeled *N*-glycans released from human serum protein digests. Inset: four representative peudo-MS^3 spectra of $Y_1(H)$ ion *m/z* 523.3 integrated over 0.2 min periods marked with asterisks corresponding to the migration time zones of aminoxyTMT labeled $(Hex)_3(GlcNAc)_4Fuc$, $(Hex)_5(GlcNAc)_4(NeuAc)$, $(Hex)_5(GlcNAc)_4NeuAc)_2$, $(Hex)_6(GlcNAc)_5(NeuAc)_3$ respectively.

Table 1 Reporter ion intensity ratios of the three-plex aminoxyTMT-labeled *N*-glycans released from RNase B. *N*-glycans released from RNase B was labeled with aminoxyTMT^6-128, aminoxyTMT^6-130 and aminoxyTMT^6 -131 in 1:1:1 ratio. The ratios were calculated from CID MS/MS spectra acquired from CE separation coupled with online data-dependent analysis.

Glycan composition	**Precursor *m/z***	**[a] I_{128}:I_{130}:I_{131}**	**[b] CV(%,I_{130}/I_{128})**	**[b]CV(%,I_{131}/I_{128})**
$(Man)_5(GlcNAc)_2$	$779.83_{(Na+H)}$,$768.85_{(2H)}$	1:1.12:0.89	11.1	1.8
$(Man)_6(GlcNAc)_2$	$860.87_{(Na+H)}$, $849.88_{(2H)}$, $579.56_{(2H+K)}$	1:1.11:0.88	3.0	7.8
$(Man)_7(GlcNAc)_2$	$941.88_{(Na+H)}$, $930.91_{(2H)}$, $633.58_{(2H+K)}$	1:1.11:0.91	3.8	10.6
$(Man)_8(GlcNAc)_2$	$1022.92_{(Na+H)}$, $1011.93_{(2H)}$, $687.56_{(2H+K)}$	1:1.12:0.91	2.5	0.3
$(Man)_9(GlcNAc)_2$	$1103.95_{(Na+H)}$, $1092.96_{(2H)}$, $741.62_{(2H+K)}$	1:1.17:0.96	6.9	5.4

[a] Reporter ion intensities from MS/MS scans of precursor ions belonging to the same glycan composition are summed for the calculation of the ratios. [b] CV are calculated from three technical replicates of CE-ESI-MS/MS analysis.

Supporting information:

EXPERIMENTAL SECTION

Chemicals and Materials

Optima LC/MS grade water, methanol (MeOH), acetonitrile (ACN), formic acid, trifluoroacetic acid (TFA) and acetic acid were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Maltopentose, maltohexaose, maltoheptaose, disialyllacto-N-tetraose (DSLNT), ribonuclease B (RNase B) from bovine pancreas, fetuin from fetal bovine serum, dithiothreitol, iodoacetamide, guanidine hydrochloride, triethylammonium acetate were purchased from Sigma-Aldrich (St. Louis, MO, U.S.A.). Maltooctaose (80% purity), sialyllacto-N-tetraose a (LSTa) sodium salt and sialyllacto-N-tetraose c (LSTc) were purchased from Carbosynth Limited (Bershire, UK). Sialyllacto-N-tetraose b (LSTb) was purchased from Prozyme (Hayward, CA). Human serum palette, MS grade trypsin, 10 kDa molecular weight cut-off spin-columns, carbonyl-reactive aminoxyTMT reagents were provided by Thermo Scientific Pierce (Rockford, IL). PNGase F was from Promega (Madison, WI). SepPak C18 solid phase extraction column, 3 cc Oasis MAX and Oasis HLB cartridges were purchased from Waters (Milford, MA). Fused silica capillary (75 and 50 μm ID, 360 μm OD) was purchased from Polymicro Technologies (Phoenix, AZ).

N-glycan release from glycoprotein standards

1 mg RNase B or bovine fetuin was dissolved in 300 μL of 6 M guanidine hydrochloride (pH 8), reduced with dithiothreitol and alkylated by iodoacetamide. Sample was then dialyzed against water using a 10 kDa molecular weight cut-off membrane to remove guanidine hydrochloride and reagents. Dialyzed protein samples were dried in a SpeedVac concentrator (Thermo Scientific, Waltham, MA, USA), resuspended in 50 mM triethylammonium acetate, and digested with

PNGase F for 18 hours to release *N*-glycans. Glycans were purified using SepPak C18 SPE column, eluted with 1% acetic acid in water solution, dried completely using a SpeedVac concentrator, and stored at -20 ℃ before use.

_N_-glycan release from human serum proteins

10 mg of human serum palette was dissolved in 6 M guanidine hydrochloride and reduced by DTT, alkylated with iodoacetamide. Then the sample was dialyzed against water using a 10 kDa MWCO membrane to remove guanidine hydrochloride and reagents. Trypsin digestion of the serum proteins were performed in a solution containing 20% n-propanol and 20 mM solution of triethylammonium bicarbonate with 1:50 enzyme-to-substrate ratio. The sample was digested for 18 hours and then acidified with TFA. After the digestion, large insoluable proteins/peptides were removed by spinning the sample at 5,000g for 20 minutes, and filtering the solution with 10 kDa MWCO spin-columns. The filtered solution was dried in a speedvac, followed by glycopeptide enrichment with OasisMAX SPE cartridges. The MAX cartridges were equilibrated with acetonitrile, 1 M triethylammonium acetate, water and 1% TFA in 95% acetonitrile sequentially. The digested peptides was dissolve in 200 μL of 1:1 water/acetonitrile with 0.5% TFA solution and loaded into the cartridge filled with washing solution. The SPE column was then washed by 1% TFA in 95% acetonitrile, and the bound glycopeptides was eluted with with 1:1 water/acetonitrile 0.5% TFA solution. The dried glycopeptides were resuspended in 50 mM triethylammonium acetate, and digested with PNGase F for 18 hours to release *N*-glycans. The glycans were then purified with SepPak C18 SPE column.

AminoxyTMT labeling of glycans

Oligosaccharide standards (less than100 μg) or purified *N*-glycan released from ~200 μg of glycoprotein standards was dissolved in 50 μL 0.1% acetic acid solution before labeling. 200 μg of aminoxyTMT reagent was dissolved in 100 μL methanol and mixed with the glycan solution. The reaction mixture was incubated at room temperature under constant shaking for 15 minutes, and then evaporated to dryness by a SpeedVac concentrator. The dried contents was re-suspended in 100 μL of 1:3 (v/v) water/methanol solution, vortexed, and evaporated to dryness again. The excess aminoxyTMT reagent was quenched by adding 200 μL of 1:4:95 (v/v/v) benzaldehyde/H_2O/MeOH into the labeled sample.

After drying down, the samples labeled by different aminoxyTMT channels were combined and cleaned up using Oasis HLB solid phase extraction cartridges. The Oasis HLB cartridge was first conditioned with 1:1 (v/v) ACN/H_2O, and washed by 3:97 (v/v) ACN/H_2O. The samples dissolved in 100 μL of 1:1 (v/v) ACN/H_2O were dispensed into the cartridge filled with the washing solution. The excess labeling reagent was removed by flushing the SPE cartridge by 3:97 (v/v) ACN/H_2O, and the labeled glycans were eluted by 1:1 (v/v) MeOH/H_2O. The eluates were evaporated to dryness and re-suspended in 1:1 (v/v) MeOH/H_2O solution for subsequent mass spectrometric analysis.

Direct infusion ESI-MS/MS and ESI-IMS-MS/MS

Waters Synapt G2 mass spectrometer was employed to acquire both direct infusion ESI-MS/MS and IMS-MS/MS data. Direct infusion experiments were performed with the nanoSpray ion source. The aminoxyTMT labeled glucose oligomers and human milk oligosaccharides were dissolved in 1:1 (v/v) water/methanol and infused at a flow rate of 0.4 μL/min for ESI-CID MS/MS or ESI-IMS-MS analysis. The mass spectrometer was operated in positive mode. The source temperature

was set at 100 °C, the electrospray voltage 3.1 kV, the sampling cone 30 V, extraction cone voltage 4.0 V, and the cone gas flow 20 L/hr. The collision gas in the trap cell and transfer cell was argon, and the low mass resolution of resolving quadrupole was set to 8 during MS/MS data acquisition. All the data were acquired in resolution mode.

For the traveling wave ion mobility spectrometry study, the trap bias voltage was set at 48 V, helium cell DC voltage 35 V, and different combination of traveling wave velocity (m/s)/ traveling wave height (V) (500/30, 500/35, 600/35, 600/40, 700/35, 700/40, and 800/40) were compared to optimize the separation efficiency of aminoxyTMT labeled sialyllacto-N-tetraose isomers.

CE-ESI-MS setup

An Agilent HP 3D capillary electrophoresis system was coupled with the Waters Synapt G2 MS using a junction-at-the tip type interface originally developed by Maxwell et al[63-65]. Briefly, the terminal end of the CE column was inserted into a concentric stainless steel (s.s.) needle, which was internally tapered and externally bevelled at the tip, and secured using PEEK fittings through the opposing ports of a tee union. A second capillary, connected to the orthogonal port of the tee union, delivered the modifier solution to the flow-through microvial formed by the interior of the sprayer needle tip and the column end. The nano-source of the mass spectrometer was adapted to host the s.s. tee union with the s.s needle positioned in an orthogonal configuration. The bevelled face of the sprayer tip was positioned towards the sampling cone.

Capillary electrophoresis was performed with an 80 cm long, 360 μm OD×50 μm ID bare fused silica capillary column. The background electrolyte (BGE) solution was made up of a solution of 50:49:1 (v/v/v) MeOH/H_2O/formic acid. Prior to each run, the column was flushed with water and BGE for 5 minutes each. The sample dissolved in 1:1 (v/v) MeOH/H_2O was loaded

into the column by applying 50 mbar for 20 seconds at the capillary inlet. 30 kV was applied at the capillary inlet throughout the separation and the current was around 4 μA. To facilitate ESI and electrical continuity at the column end, the modifier solution containing MeOH/H_2O/formic acid (v/v/v) 50:49:0.2 was infused at 0.6 μL/min during the CE-MS run. The MS source parameters in CE-MS experiments were set the same as those for direct infusion experiments unless specified.

RESULTS AND DISCUSSION

CID MS/MS fragmentation of aminoxyTMT-labeled glycan standards

Native glycans are easily ionized as alkali metal adducts in positive ESI mode[66]. Labeling the glycans with aminoxyTMT increases the molecules' overall hydrophobicity, and the tertiary amine on the reporter ion group provides a basic site for proton attachment. Both factors make the labeled compounds more easily ionized in ESI. Two glycan standards, maltooctaose and disialyllacto-N-tetraose (DSLNT), were labeled with aminoxyTMT and direct infusion CID MS/MS experiments were performed to examine the fragmentation pattern of different types of ion adduct of the labeled compounds. Precursor ions such as $[M+2H]^{2+}$, $[M+Na+H]^{2+}$, $[M+K+2H]^{3+}$, $[M+K+H]^{2+}$ were commonly observed in MS^1 scan, and decent amount of reporter ions could be generated from these multiply charged protonated precursors at optimized collision energy. For both aminoxyTMT labeled maltooctaose and disialyllacto-N-tetraose (DSLNT), CID MS/MS of the precursor ion $[M+Na+H]^{2+}$ at a trap cell collision energy of 45 V provides the highest yield of reporter ions for accurate MS^2 based quantification as well as the most informative fragment ions (Figure S1). Figure S1A shows the tandem mass spectrum of threeplex aminoxyTMT (128, 130, 131) labeled maltooctaose. The reporter ions *m/z* 128.1, 130.1, 131.1 were the base peaks in the MS^2 spectrum, and the major fragment ions generated were B_n(Na), C_n(Na), C_n-2H(Na),

Y_n(H+Na), Y_n(H) and Y_n(Na) formed by glycosidic bond cleavage. Small amount of $^{0,2}A_n$(Na), $^{3,5}A_n$(Na), $^{2,4}A_n$(Na), $^{1,5}A_n$(Na) ions generated by cross-ring fragmentation of the glucose unit were also observed, and some of these masses could be attributed to internal fragment ions (A/Y ions) formed by cross-ring fragmentation at the reducing terminus and glycosidic cleavage at the non-reducing end. The [M-reporter ion-CO+H+Na]$^+$ ion, generated from the cleavage of the amide bond near the reporter group, exhibited an isotopic mass ratio distribution related to the multiplexed isobaric labeling of the glycan (Figure S1A inset). We have also noticed that the formation the C_n-2H (Na) ion was not commonly observed with ESI-CID fragmentation or MALDI-post source decay of native glycans. For this type of fragment ions, a hydroxy lactone structure and an epoxy-hydroxycyclopentanone structure of the residue at reducing end were proposed by Harvey et al in an earlier study [67]. Figure S1B shows the tandem mass spectrum of duplex aminoxyTMT (128, 131) labeled DSLNT. The reporter ions were intensive in the low mass region. As the NeuAc residues are very labile during activation, most of the fragments observed were internal fragment ions (C/Y, B/Y, A/Y) with the two NeuAc monosaccharide residues cleaved off.

Relative quantification of isobarically labeled glucose oligomers

Previous studies showed that low yield of reporter ions generated from post-source decay of the sodiated aminoxyTMT-labeled glycans using MALDI ToF/ToF compromised the quantification accuracy and dynamic range[40]. Here, we assessed the relative quantification accuracy and dynamic range by CID fragmentation of isobarically labeled glycans on an ESI-Q-ToF mass spectrometer. Glucose oligomer mixtures (DP 5~8) with known ratios of concentration (1:1:1, 1:4:8, 4:8:1), were first labeled differentially with aminoxyTMT^6-128, TMT^6-130, TMT^6 -131 and then pooled together for direct infusion ESI-MS analysis. Each [$(Glc)_n$-aminoxyTMT+Na+H]$^{2+}$ precursor (n =

5-8) was isolated by quadrupole and fragmented in the trap collision cell. As shown in Figure S2A, the pronounced signals of reporter ions recovered from $[M+Na+H]^{2+}$ precursor made the accurate quantification possible. Since the mass resolution of the ToF analyzer is not enough to resolve the isotopic peak of reporter ion *m/z* 130.1 from the reporter ion *m/z* 131.1, the measured ion intensities of reporter ion *m/z* 131.1 was corrected using the following equation $I_{131,corrected} = I_{131,measured} - 0.0494 \times I_{130,measured}$. In Figure S2B, the corrected reporter ion intensities ratios showed good agreement with the expected ratios (error <20%) and acceptable variation (CV<13%) within at least one order of magnitude.

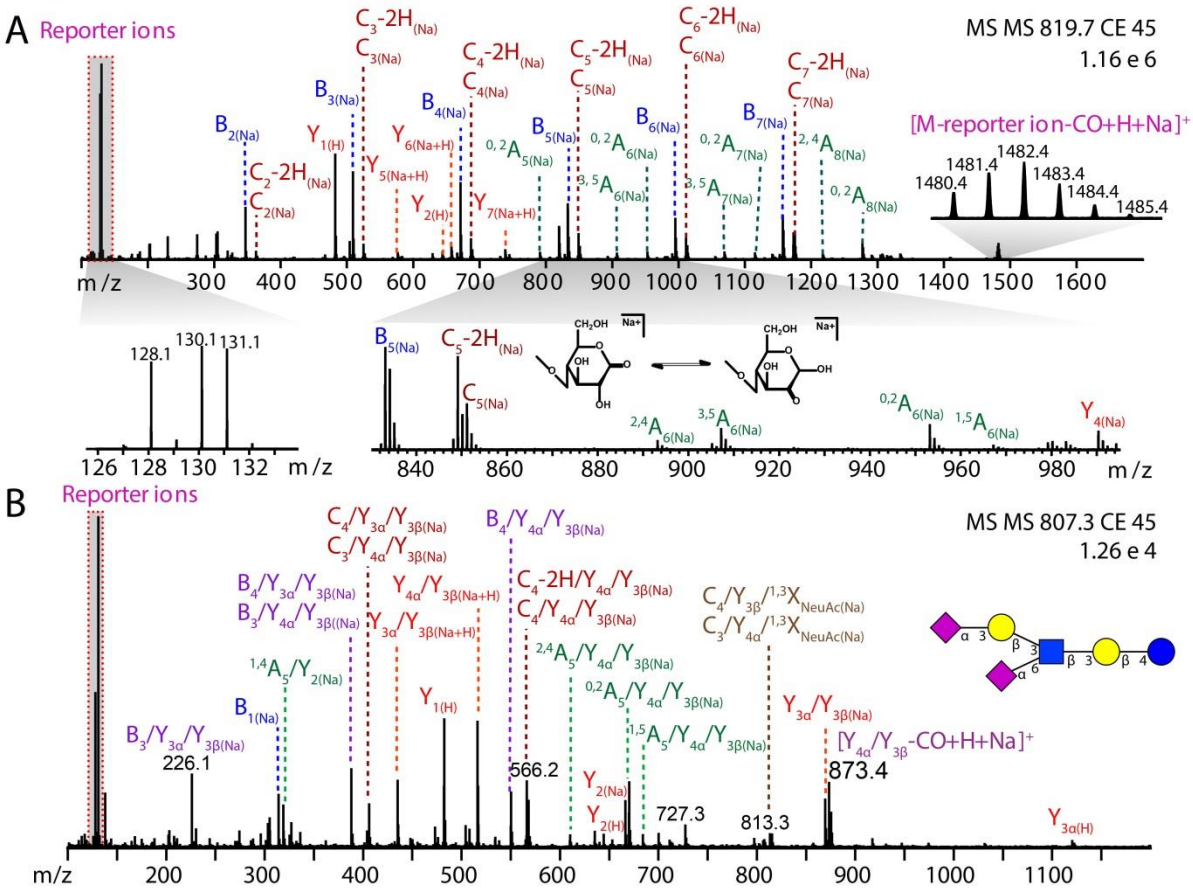


Figure S1. CID MS/MS spectra of (A) $[(Glc)_8$-aminoxyTMT +Na+H$]^{2+}$ at *m/z* 819.7, (B) $[(DSLNT)_8$-aminoxyTMT +H+Na$]^{2+}$ at *m/z* 807.3. Domon and Costello nomenclature[62] is used for annotation of fragment ions in this and following figures. B,Y and C ions generated by the cleaveage of glycosidic bonds and A ions generated from cross-ring fragmentation were observed as different types of ion adducts, which are indicated in the bracketed subspript of the fragment ion anotation. Top pannel inset, proposed chemical structures of the glucose residue at the reducing end of the C-$2H_{(Na)}$ ions[67].

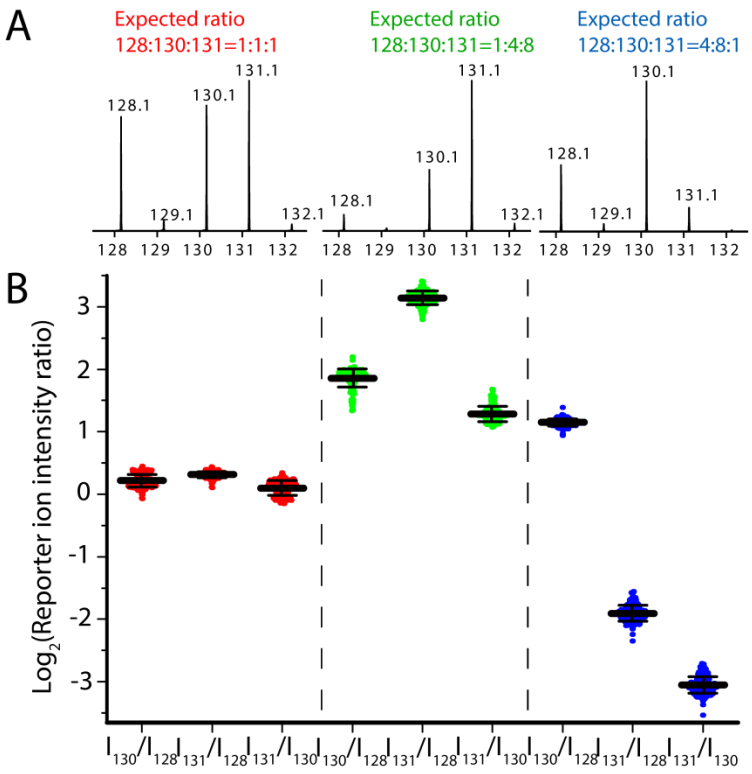


Figure S2. (A) Representative MS/MS spectra and (B) reporter ion intensity ratios measured by CID MS/MS of three groups of glucose oligomers (DP 5~8) mixture differentially labeled using three-plex aminoxyTMT reagents. Mixture of glucose oligomers (DP 5~8) with known ratios (left, 1:1:1; middle, 1:4:8; right, 4:8:1) were labeled with aminoxyTMT^6-128, aminoxyTMT^6-130 and aminoxyTMT^6-131 respectively, and mixed for direct infusion MS/MS analysis. The sample concentration was approximately 30 ng/μL for each glucose oligomer. The trap collision cell voltage was set as 30 V for *m/z* 576.7, 40 V for *m/z* 657.7 and *m/z* 738.8, 45V for *m/z* 819.8. 58 spectra at 0.5 Hz scan rate were recorded for each $[(Glc)_n\text{-aminoxyTMT+Na+H}]^{2+}$ precursor (n = 5-8). Ion intensities of reporter *m/z* 131.1 used in Figure 2B were corrected for isotopic peak interfering from reporter ion *m/z* 130.1 using the equation $I_{131,corrected} = I_{131,measured} - 0.0494 \times I_{130,measured}$. The measured average reporter ion ratios (I_{128}:I_{130}:I_{131}) of the three groups of differentially labeled samples were 1:1.04:1.20 (left), 1:3.83:8.67 (middle), 3.68:8.44:1(right).

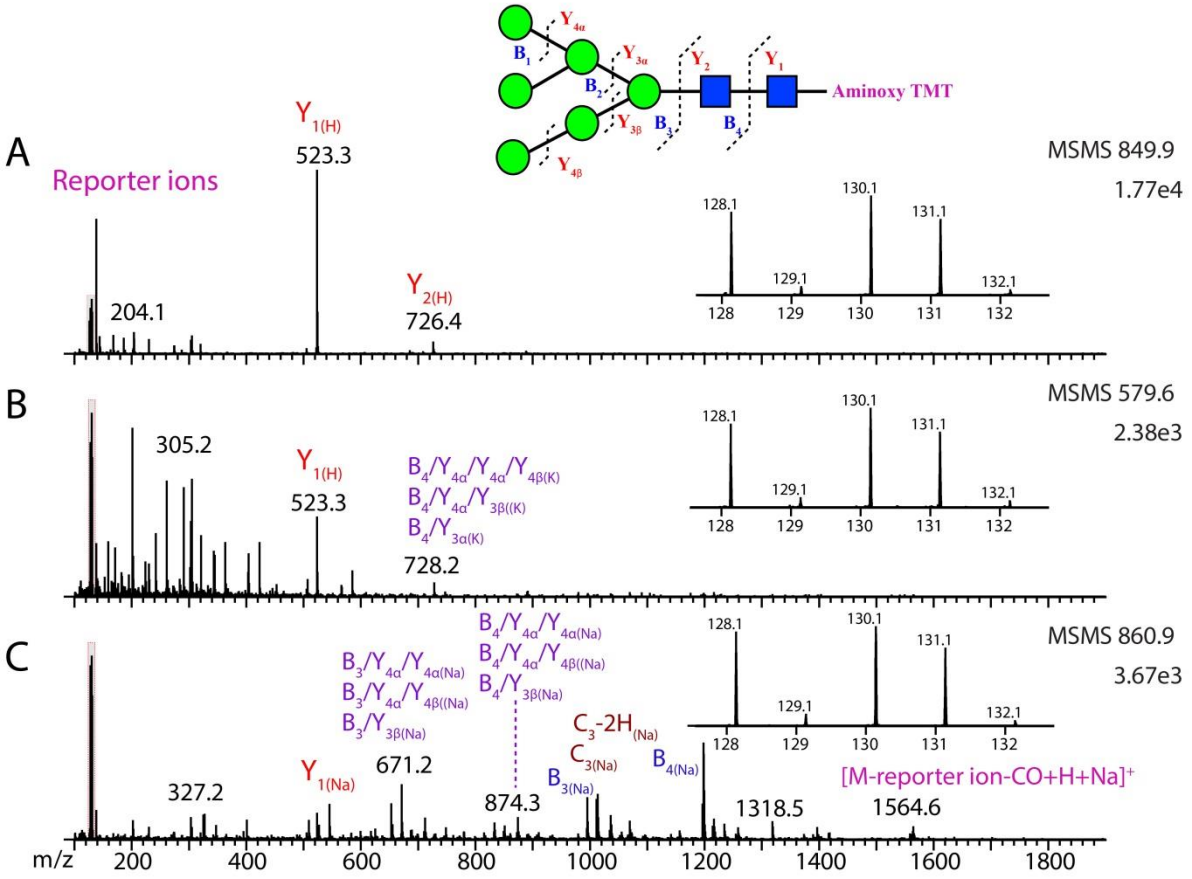


Figure S3. CID MS/MS spectra of (A) $[(Man)_6(GlcNAc)_2\text{-aminoxyTMT+2H}]^{2+}$ *m/z* 849.9, (B) $[(Man)_6(GlcNAc)_2\text{-aminoxyTMT+2H+K}]^{3+}$ *m/z* 579.6, (C) $[(Man)_6(GlcNAc)_2\text{-aminoxyTMT+Na+H}]^{2+}$ *m/z* 860.9 acquired by data-dependant analysis during online CE separation.

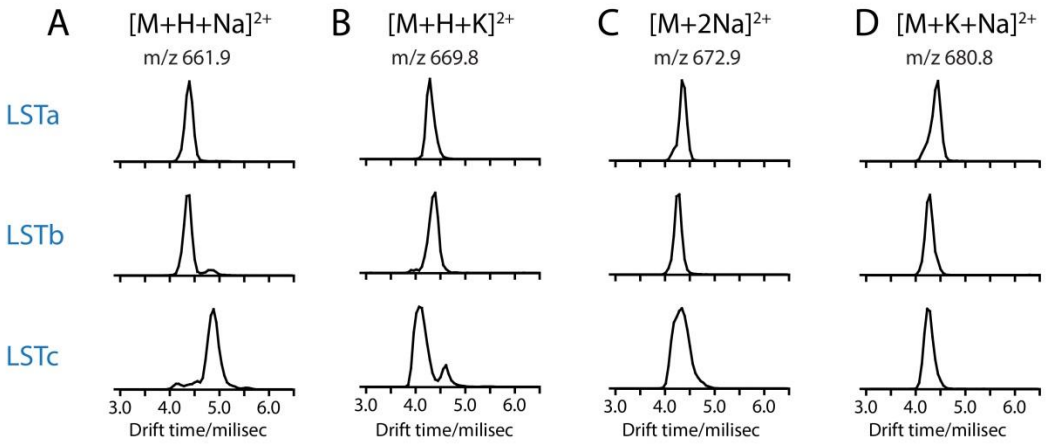


Figure S4. Arrival time distributions of (A) $[M+H+Na]^{2+}$ *m/z* 661.9, (B) $[M+H+K]^{2+}$ *m/z* 669.8, (C) $[M+2Na]^{2+}$ *m/z* 672.9, (D) $[M+K+Na]^{2+}$ *m/z* 680.8 ion adducts of aminoxyTMT labeled LSTa (top panel), LSTb (middle panel), LSTc (bottom panel). IMS conditions: trap bias voltage was set at 48 V, helium cell CD 35 V, wave velocity 700 m/s, wave height 40 V.

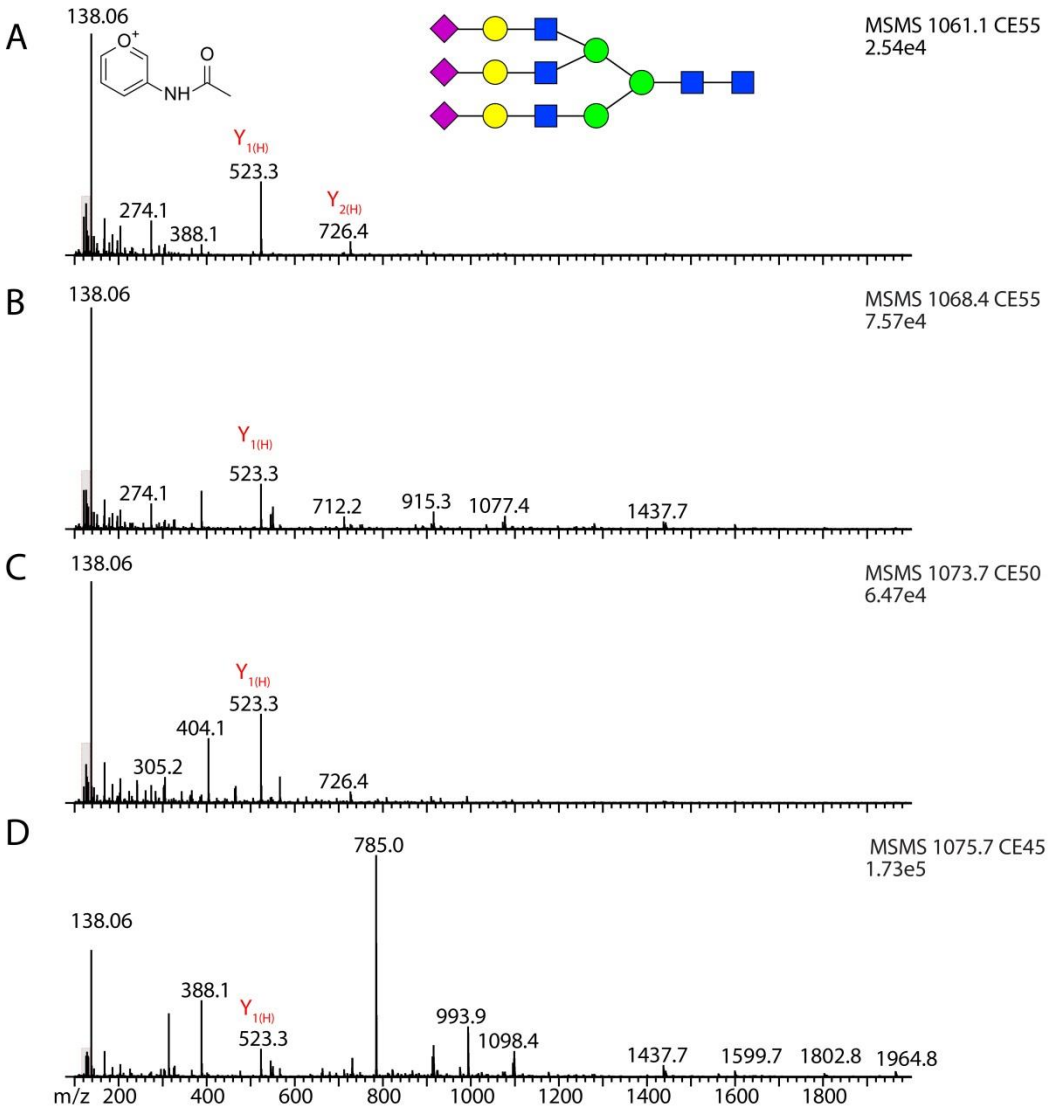


Figure S5. CID MS/MS spectra of (A) $[(Hex)_6(GlcNAc)_5(NeuAc)_3\text{-aminoxyTMT+3H}]^{3+}$ *m/z* 1061.1, (B) $[(Hex)_6(GlcNAc)_5(NeuAc)_3\text{+2H+Na}]^{3+}$ *m/z* 1068.4, (C) $[(Hex)_6(GlcNAc)_5(NeuAc)_3\text{+2H+K}]^{3+}$ *m/z* 1073.7, (D) $[(Hex)_6(GlcNAc)_5(NeuAc)_3\text{+2Na+H}]^{3+}$ *m/z* 1075.7 of tri-antennary complex *N*-glycan from bovine fetuin.

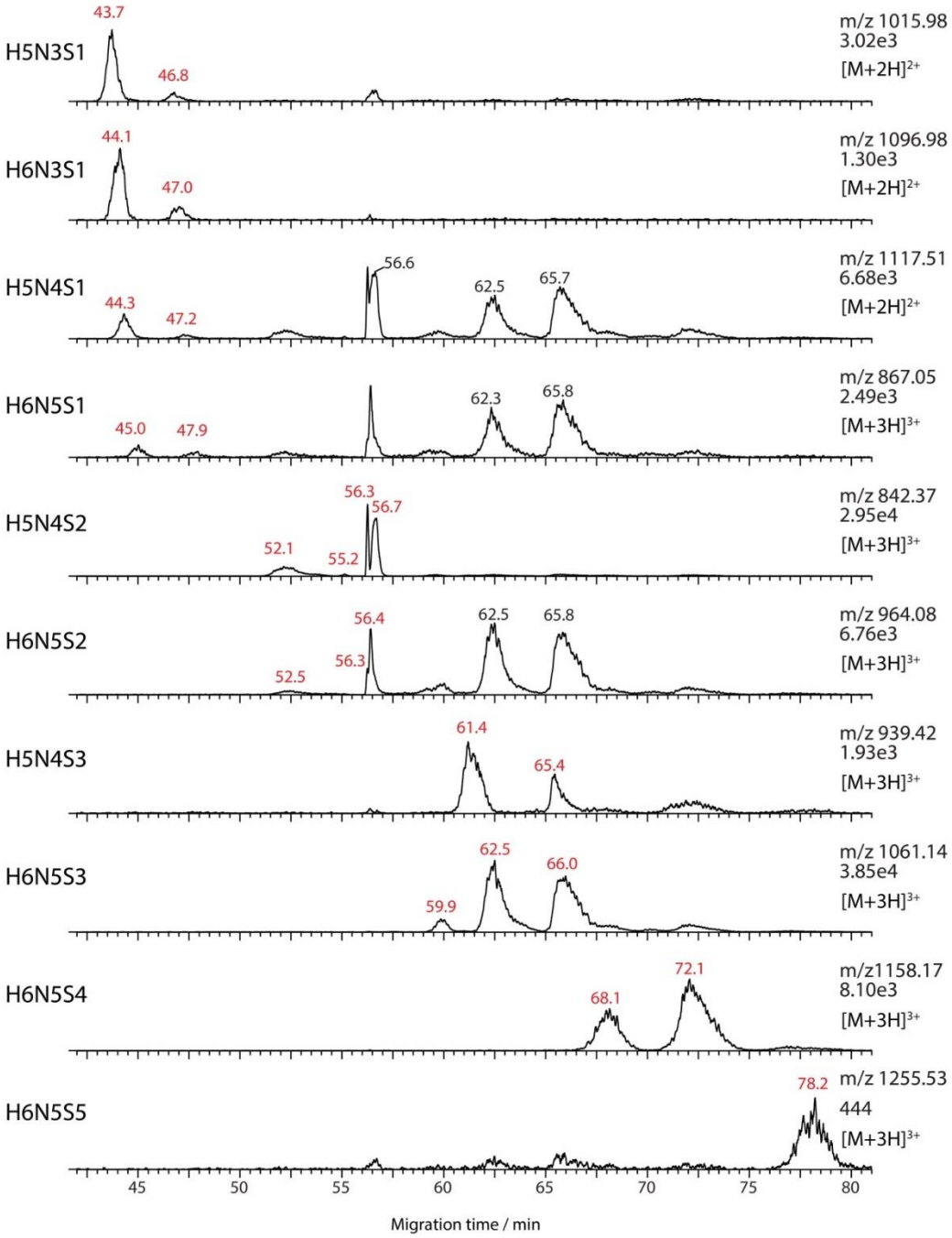


Figure S6. Extracted ion electropherograms of sialyted *N*-glycan compositions from bovine fetuin. Peaks labeled in red correspond to isomers associated with the same composition. Nomenclature is as follows: H = mannose or galactose (Hex),N = N-acetylglucosamine (HexNAc), F = fucose, and S = sialic acid(Neu5Ac).

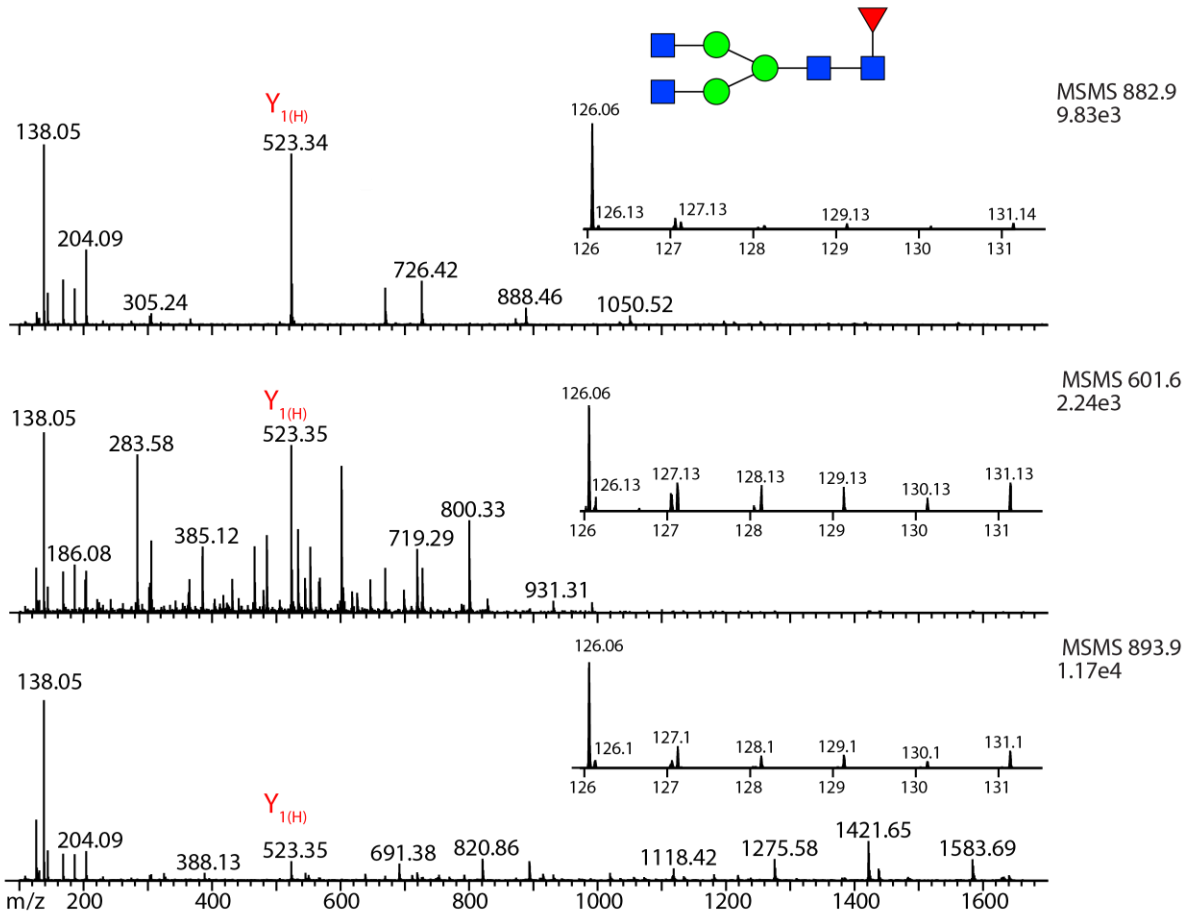


Figure S7. CID MS/MS spectra of (A) $[(Hex)_3(GlcNAc)_4(Fuc)\text{-aminoxyTMT}+2H]^{2+}$ *m/z* 882.9, (B) $[(Hex)_3(GlcNAc)_4(Fuc)+2H+K]^{2+}$ *m/z* 601.6, (C) $[(Hex)_3(GlcNAc)_4(Fuc)+Na+H]^{2+}$ *m/z* 893.9 acquired by data-dependent analysis during online CE separation of N-glycans released from human serum proteins.

Table S1 Compositions of *N*-glycans from human serum proteins identified from CE-ESI-MS.

No.	Hex	HexNAc	Fucose	NeuAc	Native glycan mass	Labeled glycan mass	Isomers observed	Migration time (min)	Most intense adduct ions observed
1	3	3	0	0	1113.407	1414.639	3	31.7, 32.4, 33.2	H+Na
2	4	2	0	0	1072.381	1373.612	2	32.0, 32.6	2H
3	3	3	1	0	1259.465	1560.697	1	32.7	H
4	3	4	0	0	1316.487	1617.718	2	32.5, 33.1	H+K
5	4	3	0	0	1275.460	1576.692	3	32.4, 33.6, 34.1	2H
6	5	2	0	0	1234.433	1535.665	1	32.0	2H
7	3	4	1	0	1462.544	1763.776	2	33.0, 33.5	2H
8	3	5	0	0	1519.566	1820.798	1	32.9	2H
9	4	3	1	0	1421.518	1722.750	4	32.8, 33.4, 34.0, 34.1	H+Na
10	4	4	0	0	1478.539	1779.771	1	32.9	2H+Na
11	5	3	0	0	1437.513	1738.744	3	32.7, 33.4, 34.1	H+K
12	6	2	0	0	1396.486	1697.718	1	32.6	2H+K
13	3	5	1	0	1665.624	1966.856	1	33.1	2H
14	4	4	1	0	1624.597	1925.829	1	33.4	2H
15	4	5	0	0	1681.619	1982.850	1	33.1	2H+Na
16	5	4	0	0	1640.592	1941.824	1	33.5	2H+Na
17	6	3	0	0	1599.566	1900.797	1	33.3	H+Na
18	7	2	0	0	1558.539	1859.771	1	33.7	3H
19	4	5	1	0	1827.677	2128.908	1	33.6	H+Na
20	5	4	1	0	1786.650	2087.882	1	34.1	2H
21	5	5	0	0	1843.672	2144.903	1	33.6	2H+Na
22	6	4	0	0	1802.645	2103.877	1	34.1	2H+Na
23	8	2	0	0	1720.592	2021.823	2	33.2, 33.6	2H
24	5	4	2	0	1932.708	2233.940	1	34.2	3H
25	5	5	1	0	1989.729	2290.961	2	38.6, 39.2	3H
26	5	4	3	0	2078.766	2379.998	1	39.4	3H
27	5	5	3	0	2281.845	2583.077	1	39.2	3H
28	6	5	2	0	2297.840	2599.072	2	39.0, 39.8	3H
29	4	3	0	1	1566.555	1867.787	2	38.3, 39.0, 39.1	2H
30	4	3	1	1	1712.613	2013.845	2	38.7, 39.3	2H
31	4	4	0	1	1769.635	2070.866	1	38.7	2H
32	5	3	0	1	1728.608	2029.840	1	38.5	2H
33	4	4	1	1	1915.693	2216.924	2	39.2, 41.6	H+K
34	5	4	0	1	1931.688	2232.919	2	39.2, 41.6	2H
35	5	4	1	1	2077.745	2378.977	1	39.5	3H
36	5	5	0	1	2134.767	2435.999	1	38.9	3H
37	5	5	1	1	2280.825	2582.057	1	39.2	3H
38	6	5	0	1	2296.820	2598.051	2	39.2, 41.8	2H+Na
39	6	5	1	1	2442.878	2744.109	3	39.3, 39.7, 41.9	2K+Na
40	5	4	0	2	2222.783	2524.015	3	44.9, 47.9, 49.9	3H

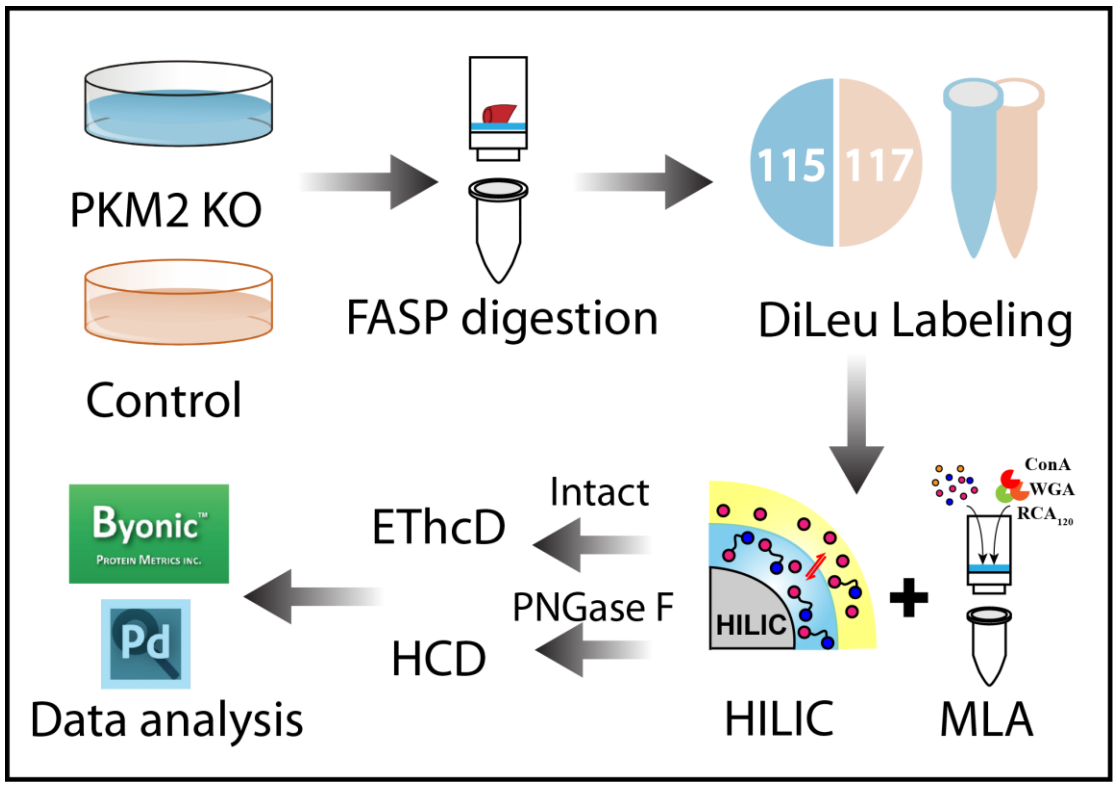
41	5	4	1	2	2368.841	2670.073	3	44.9, 47.9, 49.9	3H
42	5	5	1	2	2571.920	2873.152	3	44.9, 47.9, 49.9	2H+K
43	6	5	0	2	2587.915	2889.147	3	44.9, 47.9, 49.9	3H
44	6	5	1	2	2733.973	3035.205	3	44.9, 47.9, 49.9	H+2Na
45	6	5	0	3	2879.011	3180.242	4	50.8,51.9,54.9,58.2	3H
46	6	5	1	3	3025.068	3326.300	4	50.8,51.7,54.2,57.8	3H

Table S2 Corrected reporter ion ratios of 6-plex aminoxyTMT labeled *N*-glcyans released from human serum proteins. Data were generated by pseudo MS^3 method. The following equations are used to correct the isotopic contribution from adjacent reporter ion where I_c represents corrected intensity. $I_{126} = I_{c,126}$ ， $I_{127} = I_{c,127} + 0.0927 \times I_{c,126}$， $I_{128} = I_{c,128} + 0.0891 \times I_{c,127}$， $I_{129} = I_{c,129} + 0.0711 \times I_{c,128}$ ， $I_{130} = I_{c,130} + 0.0675 \times I_{c,129}$ ， $I_{131} = I_{c,131} + 0.0494 \times I_{c,130}$

Migration time (min)	$\mathbf{I_{c,127}/I_{c,126}}$	$\mathbf{I_{c,128}/I_{c,126}}$	$\mathbf{I_{c,129}/I_{c,126}}$	$\mathbf{I_{c,130}/I_{c,126}}$	$\mathbf{I_{c,131}/I_{c,126}}$
32.8-33.0	2.21	0.97	1.32	0.69	2.04
39.0-39.2	2.10	1.05	1.30	0.53	1.95
49.8-50.0	2.77	1.30	1.49	0.67	2.58
51.8-52.0	2.18	1.08	1.20	0.65	1.89

Chapter 6

Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD)



- Adapted from **Chen, Z.**,[#] Yu, Q.,[#] Hao, L., Liu, F., Johnson, J., Tian, Z., Kao. J., Xu, W. & Li, L. (2018). Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD). Analyst, 143(11), 2508-2519. (# These authors contributed equally)

- **Author contribution:** study was designed by **Chen, Z.**, Yu, Q, Hao, L., Kao, J., Xu, W. and Li, L.; experiment was performed by Chen, Z., Yu, Q., Liu, F. and Tian, Z.; manuscript was written by Chen, Z., and edited by Yu, Q. and Li, L.

Abstract:

The system-wide site-specific analysis of intact glycopeptides is crucial for understanding the exact functional relevance of protein glycosylation. Dedicated workflow with the capability to simultaneously characterize and quantify intact glycopeptides in a site-specific and high-throughput manner is essential to reveal specific glycosylation alteration patterns in complex biological systems. In this study, an enhanced, dedicated, large-scale site-specific quantitative N-glycoproteomics workflow has been established, which includes improved specific extraction of membrane-bound glycoproteins using filter aided sample preparation (FASP) method, enhanced enrichment of N-glycopeptides using sequential hydrophilic interaction liquid chromatography (HILIC) and multi-lectin affinity (MLA) enrichment, site-specific N-glycopeptide characterization enabled by EThcD, relative quantitation utilizing isobaric N, N-dimethyl leucine (DiLeu) tags and automated FDR-based large-scale data analysis by Byonic. For the first time, our study shows that HILIC complements to a very large extent to MLA enrichment with only 20% overlapping in enriching intact N-glycopeptides. When applying the developed workflow to site-specific N-glycoproteome study in PANC1 cells, we were able to identify 1067 intact N-glycopeptides, representing 311 glycosylation sites and 88 glycan compositions from 205 glycoproteins. We further applied this approach to study the glycosylation alterations in PKM2 knockout cells vs. parental breast cancer cells and revealed altered N-glycoprotein/N-glycopeptide patterns and very different glycosylation microheterogeneity for different types of glycans. To provide a more comprehensive map of glycoprotein alterations, N-glycopeptides after treatment with PNGase F were also analyzed. A total of 484 deglycosylated peptides were quantified, among which 81

deglycosylated peptides from 70 glycoproteins showed significant changes. KEGG pathway analysis revealed that the PI3K/Akt signaling pathway was highly enriched, which provided evidence for the previous finding that PKM2 knockdown cancer cells rely on activation of Akt for their survival. With glycosylation being one of the most important signaling modulators, our results provide additional evidence that signaling pathways is closely regulated by metabolism.

Keywords: glycoproteomics, mass spectrometry, PTM enrichment, PKM2, N-glycosylation, cancer

Introduction:

Glycosylation is one of the most prevalent and complex protein post-translational modifications (PTMs). It has been reported to play essential role in many key biological processes including cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction and endocytosis.[1-4] Altered glycosylation has been associated with different diseases such as cancer,[5-7] neurodegenerative diseases,[8] and rheumatoid arthritis.[9] Thus, the accurate site-specific characterization and quantitation of N-glycopeptides is of vital importance to provide insights into the molecular mechanisms of diseases and reveal potential biomarkers and therapeutic targets.[10]

Enrichment of N-glycopeptides is often needed due to their low abundance and signal suppression by more abundant non-glycosylated peptides. Various glycopeptide enrichment strategies have been developed, including hydrophilic interaction chromatography (HILIC),[11-14]

lectin affinity,[15, 16] TiO_2 affinity,[17] immobilized metal affinity chromatography (IMAC),[18, 19] hydrazide [20] and boronic acid chemistry.[21, 22] Among them, MLA and HILIC are the most widely used strategies, which not only enables the analysis of deglycosylated peptides after PNGase F treatment but also the analysis of intact N-glycopeptides.[23] The former strategy relies on N-glycopeptides' glycan structure motif recognition, while the latter takes advantage of the differences in physicochemical properties between non-glycosylated peptides and glycopeptides due to glycan attachment. Previous studies [24, 25] have indicated these two methods complement each other in mapping glycosylation sites, and recommended combined strategies should be used for better glycosylation site coverage. Very recently Zhang *et al* [26] showed that HILIC alone is almost as effective as the combined strategies in mapping glycosylation sites. However, these prior studies targeted at deglycosylated peptides after treatment with PNGase F and none of these studies had investigated performance of HILIC and MLA in enriching intact N-glycopeptides. We propose that an evaluation of the performance of these two most commonly used strategies in enriching intact N-glycopeptides will provide valuable guidance for the site-specific glycoproteomics.

Previous glycoproteomics studies were somewhat biased towards glycosylation sites mapping or released glycan analysis mainly due to a multitude of structural complexity encompassing attached glycans and a lack of enabling analytical technology.[23, 27, 28] However, with the rapid development of various glycoproteomics approaches, especially more advanced instrumentation equipped with new features, site-specific characterization of glycoproteins has become increasingly feasible.[29] Among various emerging techniques, the electron transfer and higher-energy collision dissociation (EThcD) technique has significantly contributed to the enhanced site-

specific glycoproteomics analyses. Commonly used for PTMs analysis, electron-transfer dissociation (ETD) was well known to preserve labile PTM on peptide backbones. It provides peptide backbone fragmentation c/z-ion series with glycan unit attached for confident mapping of glycosylation site and peptide identification; however, it only reveals the glycan as a "blind" modification with an accurate mass. On the other hand, glycan diagnostic oxonium ions and glycan backbone B- and Y-ion series produced by high-energy collision dissociation (HCD) provides confident glycan identification,[30, 31] but without glycosylation site information. Since its initial introduction by Heck and coworkers,[32] the "hybrid" dissociation method (EThcD) has shown great potential for PTM analysis with improved site localization, as well as generating richer and more informative backbone fragmentation spectra.[33, 34] When applied to intact glycopeptide analysis, EThcD can produce rich fragment ion information for glycan, peptide and glycosylation site identification in one spectrum, providing the opportunity for site-specific glycoproteomics.[35, 36]

Cancer cells display altered patterns of metabolism in order to support their growth and survival.[37] These changes in metabolism are often described as the Warburg effect,[38, 39] which diverts intermediates away from oxidation respiration in mitochondria into the biosynthesis of carbohydrates, proteins, lipids and nucleic acids, supporting the biosynthetic requirements of uncontrolled proliferation. One of the metabolism pathway alteration is the preferential expression of pyruvate kinase isozymes M2 (PKM2) in the glycolysis pathway. Unlike other pyruvate kinases such as PKL, PKR, and PKM1 that form stable tetramers (the active form of PK), PKM2 exists as both dimers and tetramers.[40] The PKM2 dimer has a reduced enzyme activity, which results in a buildup of glycolytic intermediates that enhances anabolic metabolism.[41, 42] Pentose phosphate

pathway (PPP) and hexosamine biosynthetic pathway (HBP) are two of those affected pathways that will utilize these glycolytic intermediates as substrates to synthesize nucleotide sugars such as uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), which are utilized as donor substrates for protein glycosylation.[43] In fact, studies have shown glycosylation alterations in various cancers, including breast cancer, pancreatic cancer and lung cancer.[44] With an increased expression of PKM2 in cancer cells at the same time, this leads to the speculation that PKM2 plays a vital role in the glycosylation alterations during cancer progression.[37] However, no study has been conducted to explore how PKM2 affects the overall glycosylation patterns in cancer cells.

In this work, we first evaluated the intact N-glycopeptide enrichment efficiency of HILIC and MLA respectively. The parallel experiments showed that sequential enrichment utilizing both HILIC and MLA largely increased N-glycopeptide coverage by 80% compared to utilizing either one of them alone. With hybrid fragmentation technique EThcD incorporated in the workflow, we confidently achieved glycan, glycosylation site and peptide backbone identification in one spectrum. Besides highly confident intact glycopeptide identification, reporter ions from DiLeu labeled glycopeptides generated in the low mass region enabled accurate quantitation. Utilizing the enhanced workflow, we were able to conduct comparative, quantitative, site-specific glycoproteomics study between PKM2 knockout cells and parental breast cancer cells (**Figure 1**), which revealed altered N-glycoprotein/N-glycopeptide expression in PKM2 knockout cells with increased fucosylation in several of the examined glycosylation sites. Further deglycoproteomics study confirmed the significantly altered glycosylation pattern after PKM2 knockout and revealed the altered PI3K/Akt signaling pathway, which supports the previous finding that PKM2

knockdown cancer cells rely on activation of Akt for their survival.

Experimental section

Chemicals and materials

Dithiothreitol (DTT), PNGase F and sequencing grade trypsin were from Promega (Madison, WI). Optimal LC/MS grade acetonitrile (ACN), methanol (MeOH) and water were from Fisher Scientific (Pittsburgh, PA). Concanavalin A (ConA), wheat germ agglutinin (WGA), *Ricinuscommunis* agglutinin (RCA120), iodoacetamide (IAA), acetyl-D18 glucosamine, D-lactose, methyl α-D-mannopyranoside and manganese dichloride were obtained from Sigma-Aldrich (St. Louis, MO). Tris base, urea (UA), sodium chloride, ammonium bicarbonate (ABC) and calcium chloride ($CaCl_2$) were obtained from Fisher Scientific (Pittsburgh, PA). Trifluoroacetic acid (TFA), triethylammonium bicarbonate (TEAB), N, N-dimethylformamide (DMF), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4- methylmorpholinium tetrafluoroborate (DMTMM), N-methylmorpholine (NMM) and dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St. Louis, MO). Hydroxylamine solution was purchased from Alfa Aesar (Ward Hill, MA). C18 OMIX tips were obtained from Agilent (Santa Clara, CA). Hydrophilic interaction chromatography material (PolyHYDROXYETHYL A) was obtained from PolyLC (Columbia, MD). Microcon filters YM-30 (30 kDa) was purchased from Merck Millipore (Billerica, MA). PANC-1 pancreatic ductal adenocarcinoma cells were from ATCC (Manassas, VA). Duplex DiLeu tags were custom synthesized in our own lab.[45]

PANC-1 cells and breast cancer cells

Commercially available PANC1 pancreatic ductal adenocarcinoma cells were routinely maintained in complete media of DMEM/Ham's F-12 (1:1) (ATCC) supplemented with 10% fetal bovine serum (Hyclone) and 1% antibiotic-antimycotic solution (Cellgro). Cell culture flasks were placed in an incubator containing 5% CO_2 and 98% humidity. Cells were used for a maximum of 15 passages and trypsinized using 0.25% trypsin EDTA solution (Gibco) once 80% confluence was achieved. Cell pellets were rapidly washed twice with phosphate-buffered saline, flash frozen in dry ice, and stored at -80 ℃. PKM2 knockout cancer cells and parental breast cancer cells were provided by our collaborator Xu's group.[46]

Protein extraction and digestion

PANC1 cell pellets were lysed by sonication in a solution containing digest buffer (4% SDS, 100 mM Tris/Base pH 8.0). The bicinchoninic acid assay (BCA assay) was applied to determine the protein concentration. To extract proteins from PKM2 knockout breast cancer cells and parental cells, after removing the cell media, cells were rapidly rinsed with saline, quenched with methanol, and scraped with addition of water. Chloroform was added at certain methanol/water/chloroform ratio (v/v/v). Protein precipitate in the middle layer was collected and dried. The upper and lower layers of supernatant were collected as polar and nonpolar metabolites for a separate study. The proteins were stored at -80℃. The experiments were performed with three biological replicates.

Trypsin digestion was performed based on previously reported filter-aided sample preparation (FASP) protocol[47] with some modifications. Briefly, the protein extracts were thawed and centrifuged at 16, 000 ×g for 5 min. Then 200 μg protein was taken out to the vial and 1 M DTT in digest buffer was added to make DTT final concentration 0.1 M. Incubate the sample at 95℃

for 3 min to reduce disulfide bonds. 200 μL of urea (UA) buffer (8 M UA in 100 mM Tris/Base) was added into the vial and transferred onto the 30 kDa filter. The filter was centrifuged at 14, 000 ×g for 15 min. Another 200 μL of UA buffer was added to the sample and centrifuge at 14, 000 ×g for 15 min. Add 100 μL of IAA buffer (0.05 M IAA in UA buffer) onto the filter and gently swirl to mix, then incubate in darkness for 20 min, followed by centrifugation at 14, 000 ×g for 10 min. Add 100 μL of UA buffer onto the filter, and centrifuge at 14, 000 ×g for 15 min. Repeat another 2 times. Add 100 μL of ABC buffer (50 mM) onto the filter, and centrifuge at 14, 000 ×g for 15 min. Repeat another 2 times. All the centrifugation was at 20℃. 10 μL of trypsin and 40 μL of ammonium bicarbonate (ABC) buffer was added onto the filter. The filter was incubated at 37 ℃ water bath for 18h. After incubation, the filter was transferred to a fresh collection vial and centrifuged at 14, 000 ×g for 10min. 50 μL 0.5 M NaCl solution was added onto the filter and centrifuged at 14, 000 ×g for 10min. Repeat for one more time. TFA was added into the vial to make TFA final concentration 0.25%. Samples were desalted using a SepPak C18 SPE cartridge (Waters, Milford, MA).

Protein digest labeling with DiLeu tags for quantitation

Digested peptides from PKM2 knockout cells and parental cells were dissolved in 0.5 M TEAB prior to labeling. DiLeu tags were suspended in anhydrous DMF and combined with DMTMM and NMM at 0.6× molar ratios to DiLeu. The mixture was then vortexed at room temperature for 1h. Following centrifugation, the supernatant was immediately mixed with protein digest. Labeling was performed by addition of labeling solution at a 10:1 label to peptide digest ratio by weight and vortexing at room temperature for 2 h. The labeling reaction was quenched by

addition of hydroxylamine to a concentration of 0.25%, and the labeled peptide samples were dried *in vacuo*. The samples labeled with different DiLeu channels were combined and were subject to strong cation exchange (SCX) chromatography to remove excess DiLeu tags.

HILIC enrichment

HILIC enrichment was conducted following a previously reported protocol with minor modification.[26] Namely, 5 mg of HILIC beads (PolyLC) were first activated in 100 μL elution buffer (0.1% TFA) by vortexing for 30 min. Then the activated beads were washed with 100 μL binding buffer (0.1% TFA, 19.9% H_2O, 80% ACN) for two times. 100 μg tryptic peptides were dissolved in 250 μL of binding buffer and mixed with beads at a 1:50 peptide-to-beads mass ratio. Vortex for 1 h to allow N-glycopeptides to bind to beads. The beads were washed with 50 μL binding buffer for 6 times. N-glycopeptides were eluted by washing the beads with elution buffer for 5 times. The eluted N-glycopeptides were dried down *in vacuo*. Supernatant during the process were collected for sequential MLA enrichment. The separation between beads and supernatant was achieved by centrifugation.

MLA enrichment

MLA enrichment was performed based on a previously reported protocol with some modifications.[48-51] Tryptic peptides were dissolved in 80 μL 1×binding solution (1 mM $CaCl_2$, 1 mM $MnCl_2$, 0.5 M NaCl in 20 mM Tris/Base, pH 7.3) and transferred to the 30 kDa filter. Adding 36 μL lectin mixtures (90 μg ConA, 90 μg WGA and 90 μg RCA120 in 2×biniding buffer) onto the filter. Incubate at room temperature for 1 h and unbound peptides were eluted by centrifuging at 14, 000 ×g for 10 min at 18℃. Wash with 200 μL binding solution for four times. Transfer the

filter to a new collection vial. Add 100 μL sugar mixtures (300 mM N-acetyl-D-glucosamine, D-lactose, methyl α-D-mannopyranoside in 1xbinding buffer) and incubate at room temperature for 30 min. This step was repeated once. The N-glycopeptides were eluted by centrifugation. Then the samples were acidified to 0.25% TFA and desalted by C18 OMIX tip. To prepare deglycosylated N-glycopeptides, after adding lectin mixtures and incubation for 1 h, wash with 200 μL binding solution for four times and 50 μl digest buffer (50 mM NH4HCO3) for two times. 8 μl PNGase F in 72 μl digest buffer was added onto the filter and incubated in 37℃ water-bath for 3 h. The deglycosylated N-glycopeptides were eluted with 2×50 μl digest buffer by centrifugation at 14, 000 ×g for 10 min at 18℃.

LC-MS/MS analysis

Samples were dissolved in 0.1% FA and analyzed on the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Dionex UPLC system. A binary solvent system composed of H_2O containing 0.1% formic acid (A) and MeCN containing 0.1% formic acid (B) was used for all analyses. Peptides were loaded and separated on a 75 μm x 15 cm homemade column packed with 1.7 μm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661). The LC gradient for intact N-glycopeptides was set as follows, 3%-30% B (18-98min), 30%-75% B (100-108 min) and 75%-95% B (108-118min). The mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS acquisition**.** For intact N-glycopeptide analysis, an MS1 scan was acquired from 400–1800 (120,000 resolution, $4e^5$ AGC, 100 ms injection time) followed by EThcD MS/MS acquisition of the precursors with the highest charge states in an order of intensity and detection

in the Orbitrap (60,000 resolution, $3e^5$ AGC, 100 ms injection time). EThcD was performed with optimized user defined charge dependent reaction time (2+ 50 ms; 3+ 20 ms; 4+ 20 ms; 5+ 20ms; 6 + 9 ms; 7+; 9 ms; 8+ 9ms) supplemented by 33% HCD activation. The LC gradient for deglycosylated N-glycopeptides was set as follows: 3%-30% B (18-100min), 30%-75% B (100-110 min) and 75%-95% B (110-111min). For deglycosylated N-glycopeptide analysis, an MS1 scan was acquired from 300–1500 (60,000 resolution, $2e^5$ AGC, 100 ms injection time) followed by MS/MS data-dependent acquisition of the 15 most intense ions with HCD and detection in the Orbitrap (15,000 resolution, $1e^4$ AGC, 30 NCE, 100 ms injection time).

Data analysis

All raw data files were searched against UniProt *Homo sapien*s reviewed database (08.10.2016, 20, 152 sequences). Deglycosylated peptides data were searched using Sequest HT algorithm incorporated in Proteome Discoverer (version 2.1.1.21, Thermo Scientific). Trypsin was selected as the enzyme and two maximum missed cleavages were allowed. Searches were performed with a precursor mass tolerance of 20 ppm and a fragment mass tolerance of 0.03 Da. Static modifications consisted of DiLeu labels on peptide N-termini and lysine residue (+145.1240 Da), carbamidomethylation of cysteine residues (+57.02146 Da). Dynamic modifications consisted of oxidation of methionine residues (+15.99492 Da), deamidation of asparagine (+0.98402 Da) and oxidation of methionine (+15.9949 Da). Three maximal dynamic modifications were allowed. Peptide spectral matches (PSMs) were validated based on q-values to 1% FDR (false discovery rate) using percolator. Quantitation was performed in Proteome Discoverer with a reporter ion integration tolerance of 20 ppm for the most confident centroid.

For intact N-glycopeptide analysis, data were searched using PTM-centric search engine Byonic (version 2.9.38, Protein Metrics, San Carlos, CA) incorporated in Proteome Discoverer (**Figure 1**). All the parameters were set the same as deglycosylated N-glycopeptide analysis except the following settings. N-glycosylation on asparagine and deamidation on asparagine and glutamine were set as common modifications. Oxidation of methionine (+15.9949 Da) was set as rare modification. Two common modification and one rare modification were allowed. Human N-glycan database embedded in Byonic, which included 182 glycan entities, was used. As for glycopeptide FDR control, Byonic default settings was applied that cut the protein list after the 20th decoy protein or at the point in the list at which the protein FDR first reached 1%, whichever cut resulted in more proteins. After that, Byonic estimated the spectrum-level FDR of the remaining PSMs to the reported proteins which typically were in the range 0-5%. Only those N-glycopeptides with PSMs FDR ≤ 1% were reported.

Results and Discussion

Comparison between HILIC and MLA

In a previous study, ion paring HILIC has been reported to increase the difference in hydrophilicity between glycosylated and non-glycosylated peptides, which largely decreased the non-specific binding and increased the coverage of N-glycosylation sites.[52] Thus, ion pairing HILIC (IP-HILIC) method was used in our study by adding TFA as the ion pairing reagent. Compared with single lectin method, multi-lectin affinity (MLA) method can increase N-glycopeptide coverage by binding to different glycan moiety.[50] In our study, lectin mixtures

containing ConA, WGA and RCA 120 were selected to simultaneously capture mannose, sialic acid, N-acetylglucosamine and galactose to improve N-glycopeptide enrichment ability. Starting from 200 μg proteins from PANC1, a total of 433 intact N-glycopeptides were identified using HILIC method and 582 intact N-glycopeptides were identified using MLA method. The Venn diagram analysis in **Figure 2** shows that only 103 N-glycopeptides were overlapped. HILIC and MLA enrichment were also carried out in a separate cell aliquot to demonstrate the reproducibility of the results (**Figure S2**). A closer look at the composition of the 433 N-glycopeptides enriched by HILIC shows that sialylated N-glycopeptides comprise 71% of the identified N-glycopeptides. This is probably because sialic acid increases the hydrophilic interactions between the N-glycopeptides and HILIC beads, resulting in preferential enrichment of sialylated N-glycopeptides. It has been reported that sialylated N-glycopeptides bind HILIC sorbent stronger than non-sialylated N-glycopeptides under commonly used acidic conditions.[53] While for MLA method, the sialylated N-glycopeptides enriched only comprise 12%, even though WGA in the lectin mixtures binds to sialic acid. One possible reason is that WGA also binds to N-acetylglucosamine, which may lead other high-mannose and complex glycans to compete the binding. Our results showed that HILIC preferentially enriches sialylated N-glycopeptides, while MLA preferentially enriches non-sialylated N-glycopeptides. For example, N-glycopeptide [R].EAGNHTSGAGLVQINK.[S] from cation-dependent mannose-6-phosphate receptor, which was N-glycosylated at the 4th asparagine residue, was identified using both of these two methods. With HILIC method, eight out of nine N-glycans identified at this site were sialylated glycans. In sharp contrast, three out of eight were sialylated glycans using MLA method. Another interesting observation between these two

methods is the ability to enrich paucimannosidic N-glycopeptides, which consist of two *N*-acetylglucosamine, 1-3 mannose residues and an optional fucose residue. Out of 433 glycopeptides identified by HILIC, only 3 of them were paucimannosidic N-glycopepitdes. For MLA enrichment method, 80 out of 582 were paucimannosidic N-glycopeptides. This is because, in order for glycopeptides to be captured by HILIC beads, a minimum degree of local hydrophilicity is needed, which does not appear to be provided by the paucimannosidic *N*-glycopepitdes.[54] This indicates that the nature of the individual glycan significantly influences the HILIC retention behavior of these lowly hydrophilic truncated N-glycans attached to peptides. In comparison, lectin captures N-glycans through recognizing a specific glycan moiety, which is not influenced by the hydrophilicity of glycans. Therefore, in order to achieve more comprehensive enrichment of different N-glycopeptides, our results indicate that there is a need for sequential enrichment. With no salt used in HILIC method, MLA enrichment could be directly performed with the flow-through after HILIC enrichment. After combining these two enrichment methods, a total of 1067 intact N-glycopeptides were identified, representing 311 glycosylation sites and 88 N-glycan compositions from 205 glycoproteins as shown in **Figure 3A**. As a consequence of sequential enrichment, there are about 80% increase in the number of intact N-glycopeptides, 40% increase in glycosylation sites and 35% increase in glycoproteins compared to individual enrichment method. Among the 88 N-glycans identified along with the 1067 N-glycopeptides, we found that almost all categories of N-glycans were identified, including high-mannose, hybrid and complex types along with the N-glycan synthesis pathways (**Figure 3B)**. Besides 79 of N-glycans identified in the classical synthesis pathway, 9 paucimannosidic N-glycans from 84 N-glycopeptides and 38 glycoproteins

were also identified in the unconventional truncation pathway. Gene ontology analysis of these glycoproteins revealed that they were most enriched in lysosome, which was in accordance with the previous finding.[55]

Intact N-glycopeptide quantitation enabled by EThcD

Intact N-glycopeptide analysis was conducted by utilizing EThcD as fragmentation method. As shown in **Figure 4A**, a series of fragment ions including c/z, b/y ions, glycan fragment ions and glycopeptide with one or more loss of monosaccharides were detected. In the low mass region, the glycan signature oxonium ions including *m/z* 138.06 (HexNAc-$2H_2O$-CH_2O), 168.06 (HexNAc-$2H_2O$), 186.08 (HexNAc-H_2O) and 204.09 (HexNAc), were detected, confirming that the spectrum belonged to a glycopeptide. For relative quantitation, intact N-glycopeptides from different samples were labeled by isobaric duplex DiLeu tags, which were custom synthesized by incorporating stable isotopes (13C, 2H, 18O and 15N) into the reporter group and mass-balanced group, creating reporter ions *m/z* at 115.12476 and 117.13731. When designing the DiLeu tags, we have intentionally grouped the deuterium atoms surrounding the tertiary amine group in the DiLeu tag and also only used a small number of deuterium atoms in DiLeu tag to minimize the isotope effect. As a result, the increased polarity of the amine group has successfully offset the small deuterium number difference in the tags with a negligible differences in retention time on RP chromatography.[45] The reliable and robust quantitation ability of DiLeu tags up to 12-plex for quantitative proteomics studies have been demonstrated.[56-61] Recently, our group has successfully applied DiLeu tags for quantitation of labile PTM enabled by EThcD.[62] Even though ETD has been well-known to preserve labile PTMs, the intensity of reporter ions produced is often quite

low so that the quantitation accuracy is compromised. While HCD provides high reporter ion intensity, labile PTMs are poorly preserved. Utilizing the novel EThcD hybrid fragmentation method, where a supplemental energy is applied to all fragment ions formed by ETD, our group was able to successfully apply DiLeu tags to reliable quantitative phosphoproteomics study with increased phosphorylation site confidence.[62] In this regard, we propose quantitation of intact glycopeptides could also benefit from EThcD for site-specific characterization as well as accurate quantitation. As shown in **Figure 4A**, reporter ions with a decent intensity next to the base peak (HexNAc oxonium ion) were produced in the low mass region, which ensured the accurate and reliable quantitation of intact glycopeptides. As oxonium ions are easily generated due to the labile glycosidic bonds, the oxonium ion would always be the base peak, which was also observed for isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT) labeled intact N-glycopeptides.[63-65]

To delineate to what extent the labile glycans affect the overall reporter ion yield, we compared the reporter ion yield efficiency between the labeled N-glycopeptides and labeled nonglycosylated peptides. After HILIC and MLA enrichment, there were still a few nonglycosylated peptides residuals, because of overlapping hydrophobicity between N-glycopeptides and non-glycosylated peptides and non-specific binding of nonglycosylated peptides to lectin. Thus, the reporter ions generated from the labeled N-glycopeptides and labeled nonglycosylated peptides in the same run would serve as perfect examples for their reporter ion yield comparison with a minimized instrument and sample variations. As shown in **Figure S1**, compared to nonglycosylated peptides, the average reporter ion intensity generated by labeled N-glycopeptides was 6× times lower and

the average precursor intensity was 3× times lower, which is largely due to the lower abundance and poorer ionization efficiency of N-glycopeptides. In addition, the fragile glycosidic bonds impair the yield of reporter ions, which leads to a 2-fold lower reporter ion yield for N-glycopeptides. To this end, compared to EThcD, HCD mode may generate a higher reporter ion yield of intact N-glycopeptides given sufficient collision energy used. However, it should be noted that if a higher HCD collision energy was used to increase the reporter ion yield, glycan fragments may dominate and compromise the peptide backbone fragmentation. Unlike the longer cycle time needed for ETD reaction, HCD also affords a faster scan speed that allows more spectra to be collected in one cycle. But the EThcD could produce much more fragments including c/z ions and peptide ions with partial or intact glycan preserved, resulting in better spectra quality. To conclude, the reporter ion yield, the number of total scans, and spectra quality should be considered when choosing the most effective fragmentation technique for quantitative analysis of intact glycopeptides.

Intact N-glycopeptide quantitation between PKM2 knockout cells vs. parental breast cancer cells.

PKM2 knockout breast cancer cell lines were recently generated by CRISPR/cas9 method[46] and used in glycopeptide studies. In total, 45 intact N-glycopeptides (39 up-regulated and 6 down-regulated) on 22 glycoproteins were significantly regulated ($p \leq 0.05$ and ±1.5-fold) in PKM2 knockout cells vs. parental breast cancer cells (**Figure 5B**). Using the developed approach, we were able to quantify different glycoforms on a single glycosylation site and reveal different glycosylation pattern. One example is lysosome-associated membrane glycoprotein 1 (LAMP-1),

which has been shown heavily N-glycosylated previously.[66, 67] A total of 20 N-glycopeptides from 6 glycosylation sites were quantified and N-glycopeptides from four representing glycosylation sites were shown in **Figure 7**. Five glycoforms were quantified on the 7th asparagine residue on glycopeptide [R].LLNINPNK.[T], including one high-mannose N-glycan and four fucosylated N-glycans. The results show that the high-mannose glycoform was up-regulated, while the four fucosylated glycoforms remained unaltered after PKM2 knockout, which indicates that different kinds of glycoforms respond differently even on the same glycosylation site. On the third asparagine residue of N-glycopeptide [R].KDNTTVTR.[L], 4 out of 5 fucosylated glycoforms detected showed an up-regulation after PKM2 knockout. Overall, there was an up-regulation of fucosylation on glycoprotein LAMP-1, which was also observed for LAMP-2. LAMP1 and LAMP-2 were mainly found in the lysosomes and late endosomes and were known to maintain the lysosomal acidification and lysosomal membrane integrity.[68] In addition to their role in lysosomal biogenesis, it was also reported that LAMP-1 and LAMP-2 is associated with autophagosome accumulation and biogenesis.[69, 70] A lack of fucosylation may alter the integrity of LAMP-1 and LAMP-2, which in turn may affect the dynamic distribution of lysosomes. In addition, the fucosylated LAMP-1 and LAMP-2 forms are responsible for regulating autophagy biogenesis, which may influence tumor development and progression.[71] Besides LAMP-1 and LAMP-2, an increased fucosylation was also observed for other glycoproteins including prosaposin, palmitoyl-protein thioesterase 1, galectin-3-binding protein and dipeptidyl peptidase 2. In fact, increased fucosylation has been reported to have unique biological significance in cancer.[72] For example, in breast cancer, increased core fucosylation of epidermal growth factor receptor (EGFR) has been

found to result in increased EGFR-mediated signaling associated with tumor cell growth and malignancy.[73, 74]

On the other hand, PKM2 was found to be preferentially expressed in cancer cells[75] and it significantly affected the levels of substrates used for protein glycosylation by regulating the last step of glycolysis.[76] Thus, it has been speculated that PKM2 attributes to glycosylation alteration in cancer by affecting the availability and abundance of the sugar nucleotide donors.[75] However, whether PKM2 alters the glycosylation pattern remains unknown. Our comparative glycoproteomics study using paired parental and PKM2 knockout breast cancer cells revealed alteration in glycosylation patterns after PKM2 knockout, providing evidence that PKM2 contributed to the altered glycosylation in cancer. In particular, fucosylation, one of the most widely reported glycosylation often altered in cancer cells, was notably affected by knocking out PKM2.

Deglycosylated peptide quantitation enabled by HCD

Due to glycosylation micro-heterogeneity, various glycoforms can exist on the same glycosylation site. By enzymatically removing these glycoforms with PNGase F, deglycosylated peptides could be obtained with the marker of asparagine (Asn) being deamidated indicating the exact glycosite.[77] After this step, the abundance of glycopeptides with different glycoforms will be merged into one glycosylation site-containing peptide, increasing the chance to be detected. Also the ionization efficiency will be largely increased by removing the hydrophilic glycan and signal suppression was greatly reduced by other more abundant nonglycosylated peptide, increasing the coverage of glycosylation site. Despite of losing the site-specific glycoform information,

deglycoproteomics approach could provide complementary information about alterations of glycoproteins. Thus, deglycosylated peptides were analyzed using HCD after PNGase F treatment. **Figure 4B** shows the spectrum of a duplex DiLeu labeled deglycosylated peptide, with a wealth of peptide backbone fragments and reporter ions fragments being detected. The N-glycosylation site was identified by the signature asparagine deamidation after PNGase F treatment. To avoid any false-positive assignment of N-glycosites brought by the spontaneous deamidation,[77] aliquots of intact N-glycopeptides were used as negative control samples and analyzed directly by LC-MS/MS without PNGase F treatment. Furthermore, all the deglycosylated glycopeptides identified in the study contain NXS/T (where X≠proline) consensus motifs to ensure confident identification. A total of 484 deglycosylated peptides were quantified, among which 81 glycosylations sites from 70 glycoproteins, showed significant changes with more than 1.5-fold ($p \leq 0.05$) alterations (**Figure 5A**). KEGG pathway analysis revealed that these altered glycoproteins are highly enriched in ECM-receptor interaction, arrhythmogenic right ventricular cardiomyopathy (ARVC), PI3K/Akt signaling pathway, cell adhesion molecules (CAMs) and focal adhesion pathways (**Figure 6A**).

One interesting observation was the alteration of the PI3K/Akt signaling pathway that was highly enriched. In our study, 10 glycoproteins along the PI3K/Akt pathway were found to be significantly altered in PKM2 knockout cells, including insulin like growth factor 1 receptor (IGF1R), insulin receptor (INSR), integrin subunit alpha V (ITGAV), integrin subunit beta 1 (ITGB1), integrin subunit beta 6 (ITGB6), ephrin A4 (EFNA4), laminin subunit alpha 5 (LAMA5), laminin subunit beta 1(LAMB1), laminin subunit beta 2 (LAMB2) and thrombospondin 1 (THBS1) (**Table 1**). PKM 2 plays an important role for anabolic glycolysis to support rapid proliferation of

cancer cells.[76] Previous studies have shown that PKM2 knockdown only leads to a modest impairment of cancer cell survival but the mechanism remains elusive. One study showed that knockdown of PKM2 leads to activation of Akt and inhibition of PI3K/Akt signaling pathway, resulting in significant growth inhibition and apoptosis in PKM2 knockdown cells, which indicated Akt activation was necessary for the survival of PKM2 knockdown cells.[78] In fact, activated Akt is a well-established survival factor and activated Akt modulates the function of many substrates involved in the regulation of cell survival, cell cycle progression and cellular growth.[79] Thus, our results offer additional evidence to support that PKM2 knockout cells may rely on regulation of the PI3K/Akt pathway to survive. Gene ontology biological process analysis revealed that cell adhesion, positive regulation of cell migration, cell adhesion mediated by integrin, viral entry into host cell, extracellular matrix organization, leukocyte migration and integrin-mediate signaling pathways are the main biological processes that have been affected (**Figure 6C**). According to our observation, there was a decrease in cell migration after PKM2 knockout, which agreed well with the alteration of positive regulation of cell migration pathway. Gene ontology cellular component analysis showed that both intracellular and extracellular glycoproteins were affected, including those localized in extracellular exosome, cell surface, and endoplasmic reticulum, etc. (**Figure 6B**).

Both altered glycosylation and metabolic reprogramming have been extensively reported in cancer cells.[80, 81] Glycosylation is generated from metabolites and is also a nutrient-sensitive post-translational modification of proteins. At the same time, as one of the key cue for cellular signals, glycosylation also regulates metabolism through various receptors and signal transduction pathways that regulate metabolism.[37] This implies that these two most common features,

glycosylation signaling and metabolism, of cancers cells might be intervened.[82] PKM2, a limiting glycolytic enzyme that catalyzes the final step in glycolysis, plays essential role in metabolic changes in cancer cells. Many studies[83-85] have shown PKM2 was preferentially expressed in cancer cells and it contributes to the accumulation of metabolites intermediates, which are the sources of substrates for protein glycosylation. Thus, PKM2 has been implicated to be involved in the glycosylation alteration in cancer.[37] To directly address the question whether and how PKM2 knockout alters the overall glycosylation pattern in cancer cells, we employed PKM2 knockout cell lines for quantitative glycosylation studies. Indeed, altered glycosylation was detected by knocking out PKM2, which confirms the previous speculation as well as provides direct evidence that metabolism and signaling pathway were tightly related. On the other hand, previous study showed that activation of Akt is necessary for the survival of PKM2 knockdown cells, suggesting that combination of PKM2 knockdown with PI3K or Akt inhibitors may be therapeutically effective. Our study identified 10 N-glycoproteins with altered glycosylation in the PI3K/Akt pathway upon loss of PKM2, which could provide the explicit targets for follow-up studies.

Conclusions

In the present study, a dedicated large-scale site-specific quantitative N-glycoproteomics workflow has been established. By employing FASP method for trypsin digestion, SDS could be used to enhance the extraction efficacy of membrane glycoproteins. Intact N-glycopeptides enrichment efficiency with HILIC and MLA was compared for the first time. The results show that HILIC preferentially enriched sialylated N-glycopepitdes while MLA preferentially enriched

nonsialylated glycopeptides, implying that a sequential enrichment procedure would improve intact N-glycopeptide enrichment efficiency. Benefiting from EThcD, site-specific intact glycopeptide structure characterization and quantitation can be achieved in one spectrum, which also provides an opportunity for FDR control at the intact glycopeptide level. Using PTM-centric search engine Byonic, FDR-based large-scale glycoproteomics was achieved. Furthermore, we successfully applied this developed workflow to study the glycosylation alteration in PKM2 knockout breast cancer cells vs. parental cells. Upon loss of PKM2, the abundance ratios of different glycoforms on the same glycosylation vary differently and an increased fucosylation was observed in several of the examined glycosylation sites. Further deglycoproteomics study revealed that the 10 glycoproteins in PI3K/Akt signaling pathway were altered, which supports the previous finding that PKM2 knockdown cancer cells rely on the activation of Akt for their survival.

Acknowledgement

This research was supported in part by the National Institutes of Health (NIH) grants R21AG055377, R01 DK071801, R01 CA213293, and R01AG052324. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges a Vilas Distinguished Achievement Professorship and Janis Apinis Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. We thank Dr. Marshall Bern from Protein Metrics for providing access to Byonic software package.

References:

1. Gewinner, C.; Hart, G.; Zachara, N.; Cole, R.; Beisenherz-Huss, C.; Groner, B., The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *J. Biol. Chem.* **2004,** 279, 3563-3572.

2. Alikhani, Z.; Alikhani, M.; Boyd, C. M.; Nagao, K.; Trackman, P. C.; Graves, D. T., Advanced glycation end products enhance expression of pro-apoptotic genes and stimulate fibroblast apoptosis through cytoplasmic and mitochondrial pathways. *J. Biol. Chem.* **2005,** 280, 12087-12095.

3. Phan, U. T.; Waldron, T. T.; Springer, T. A., Remodeling of the lectin–EGF-like domain interface in P-and L-selectin increases adhesiveness and shear resistance under hydrodynamic force. *Nat. Immunol.* **2006,** 7, 883-889.

4. Partridge, E. A.; Le Roy, C.; Di Guglielmo, G. M.; Pawling, J.; Cheung, P.; Granovsky, M.; Nabi, I. R.; Wrana, J. L.; Dennis, J. W., Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis. *Science* **2004,** 306, 120-124.

5. Hakomori, S.-i., Tumor malignancy defined by aberrant glycosylation and sphingo (glyco) lipid metabolism. *Cancer Res.* **1996,** 56, 5309-5318.

6. Couldrey, C.; Green, J. E., Metastases: the glycan connection. *Breast Cancer Res.* **2000,** 2, (5), 321.

7. Drake, P. M.; Cho, W.; Li, B.; Prakobphol, A.; Johansen, E.; Anderson, N. L.; Regnier, F. E.; Gibson, B. W.; Fisher, S. J., Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin. Chem.* **2010,** 56, 223-236.

8. Hwang, H.; Zhang, J.; Chung, K. A.; Leverenz, J. B.; Zabetian, C. P.; Peskind, E. R.; Jankovic, J.; Su, Z.; Hancock, A. M.; Pan, C., Glycoproteomics in neurodegenerative diseases. *Mass spectrometry reviews* **2010,** 29, (1), 79-125.

9. Rademacher, T. W.; Parekh, R. B.; Dwek, R. A.; Isenberg, D.; Rook, G.; Axford, J. S.; Roitt, I. In *The role of IgG glycoforms in the pathogenesis of rheumatoid arthritis*, 1988, 1988; Springer: 1988; pp 231-249.

10. Thaysen-Andersen, M.; Packer, N. H., Advances in LC–MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N-and O-glycoproteome. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2014,** 1844, (9), 1437-1452.

11. Hägglund, 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. *Journal of Proteome Research* **2004,** 3, (3), 556-566.

12. Wuhrer, M.; Koeleman, C. A.; Hokke, C. H.; Deelder, A. M., Protein glycosylation analyzed by normal-phase nano-liquid chromatography− mass spectrometry of glycopeptides. *Anal. Chem.* **2005,** 77, (3), 886-894.

13. Thaysen-Andersen, M.; Thøgersen, I. B.; Lademann, U.; Offenberg, H.; Giessing, A. M.; Enghild, J. J.; Nielsen, H. J.; Brünner, N.; Højrup, P., Investigating the biomarker potential of glycoproteins using comparative glycoprofiling—application to tissue inhibitor of metalloproteinases-1. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2008,** 1784,

(3), 455-463.
14. Chen, Z.; Zhong, X.; Tie, C.; Chen, B.; Zhang, X.; Li, L., Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents. *Anal. Bioanal. Chem.* **2017,** 409, (18), 4437-4447.
15. Hirabayashi, J., Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconjugate J.* **2004,** 21, (1), 35-40.
16. Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wiśniewski, J. R., Screening for N‐glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **2004,** 4, (2), 454-465.
17. Larsen, M. R.; Jensen, S. S.; Jakobsen, L. A.; Heegaard, N. H., Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Mol. Cell. Proteomics* **2007,** 6, (10), 1778-1787.
18. Glover, M. S.; Yu, Q.; Chen, Z.; Shi, X.; Kent, K. C.; Li, L., Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass spectrom.* **2017**.
19. Zhu, J.; Wang, F.; Cheng, K.; Dong, J.; Sun, D.; Chen, R.; Wang, L.; Ye, M.; Zou, H., A simple integrated system for rapid analysis of sialic‐acid‐containing N‐glycopeptides from human serum. *Proteomics* **2013,** 13, (8), 1306-1313.
20. 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. *Nature biotechnology* **2003,** 21, (6), 660.
21. Chen, W.; Smeekens, J. M.; Wu, R., A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol. Cell. Proteomics* **2014,** 13, (6), 1563-1572.
22. Li, X.; Liu, H.; Qing, G.; Wang, S.; Liang, X., Efficient enrichment of glycopeptides using phenylboronic acid polymer brush modified silica microspheres. *Journal of Materials Chemistry B* **2014,** 2, (16), 2276-2281.
23. 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* **2016,** 15, 1773-1790.
24. 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* **2008,** 71, 304-317.
25. Ma, C.; Zhao, X.; Han, H.; Tong, W.; Zhang, Q.; Qin, P.; Chang, C.; Peng, B.; Ying, W.; Qian, X., N‐linked glycoproteome profiling of human serum using tandem enrichment and multiple fraction concatenation. *Electrophoresis* **2013,** 34, 2440-2450.
26. Zhang, C.; Ye, Z.; Xue, P.; Shu, Q.; Zhou, Y.; Ji, Y.; Fu, Y.; Wang, J.; Yang, F., Evaluation of different N-glycopeptide enrichment methods for N-glycosylation sites mapping in mouse brain. *J. Proteome Res.* **2016,** 15, 2960-2968.

27. Chen, Z.; Glover, M. S.; Li, L., Recent advances in ion mobility–mass spectrometry for improved structural characterization of glycans and glycoconjugates. *Curr. Opin. Chem. Biol.* **2018,** 42, 1-8.
28. Zhong, X.; Chen, Z.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L., Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags. *Anal. Chem.* **2015,** 87, (13), 6527-6534.
29. 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. *Molecular & Cellular Proteomics* **2016,** 15, (6), 1773-1790.
30. Domon, B.; Costello, C. E., A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **1988,** 5, 397-409.
31. Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* **1990,** 193, 886.
32. Frese, C. K.; Altelaar, A. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012,** 84, 9668-9673.
33. Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. M.; Mechtler, K.; Heck, A. J.; Mohammed, S., Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **2013,** 12, 1520-1525.
34. Liao, R.; Zheng, D.; Nie, A.; Zhou, S.; Deng, H.; Gao, Y.; Yang, P.; Yu, Y.; Tan, L.; Qi, W., Sensitive and precise characterization of combinatorial histone modifications by selective derivatization coupled with RPLC-EThcD-MS/MS. *J. Proteome Res.* **2017,** 16, (2), 780-787.
35. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.-W.; Kent, K. C.; Li, L., Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J. Am. Soc. Mass Spectrom.* **2017,** 28, (9), 1751-1764.
36. Marino, F.; Bern, M.; Mommen, G. P.; Leney, A. C.; van Gaans-van den Brink, J. A.; Bonvin, A. M.; Becker, C.; van Els, C. A.; Heck, A. J., Extended O-GlcNAc on HLA class-I-bound peptides. *J. Am. Chem. Soc.* **2015,** 137, 10922-10925.
37. Cairns, R. A.; Harris, I. S.; Mak, T. W., Regulation of cancer cell metabolism. *Nature Reviews Cancer* **2011,** 11, 85-95.
38. Warburg, O.; Wind, F.; Negelein, E., The metabolism of tumors in the body. *The Journal of general physiology* **1927,** 8, 519.
39. Warburg, O., On the origin of cancer cells. *Science* **1956,** 123, 309-314.
40. Mazurek, S., Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells. *The international journal of biochemistry & cell biology* **2011,** 43, 969-980.
41. Vander Heiden, M. G.; Locasale, J. W.; Swanson, K. D.; Sharfi, H.; Heffron, G. J.; Amador-Noguez, D.; Christofk, H. R.; Wagner, G.; Rabinowitz, J. D.; Asara, J. M., Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science* **2010,** 329, 1492-1499.
42. Ye, J.; Mancuso, A.; Tong, X.; Ward, P. S.; Fan, J.; Rabinowitz, J. D.; Thompson, C. B.,

Pyruvate kinase M2 promotes de novo serine synthesis to sustain mTORC1 activity and cell proliferation. *Proceedings of the National Academy of Sciences* **2012,** 109, 6904-6909.
43. Shirato, K.; Nakajima, K.; Korekane, H.; Takamatsu, S.; Gao, C.; Angata, T.; Ohtsubo, K.; Taniguchi, N., Hypoxic regulation of glycosylation via the N-acetylglucosamine cycle. *J. Clin. Biochem. Nutr.* **2010,** 48, 20-25.
44. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015,** 15, 540-555.
45. Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L., N, N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. *Analytical chemistry* **2010,** 82, (7), 2817-2825.
46. Liu, F.; Ma, F.; Wang, Y.; Hao, L.; Zeng, H.; Jia, C.; Wang, Y.; Liu, P.; Ong, I. M.; Li, B., PKM2 methylation by CARM1 activates aerobic glycolysis to promote tumorigenesis. *Nat. Cell Biol.* **2017,** 19, (11), 1358.
47. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009,** 6, 359.
48. Zielinska, D. F.; Gnad, F.; Schropp, K.; Wiśniewski, 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* **2012,** 46, 542-548.
49. Deeb, S. J.; Cox, J.; Schmidt-Supprian, M.; Mann, M., N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol. Cell. Proteomics* **2014,** 13, 240-251.
50. Zielinska, D. F.; Gnad, F.; Wiśniewski, J. R.; Mann, M., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010,** 141, 897-907.
51. Pan, Y.; Bai, H.; Ma, C.; Deng, Y.; Qin, W.; Qian, X., Brush polymer modified and lectin immobilized core–shell microparticle for highly efficient glycoprotein/glycopeptide enrichment. *Talanta* **2013,** 115, 842-848.
52. Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M., Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Analytical Chemistry* **2010,** 82, (13), 5598-5609.
53. Thaysen-Andersen, M.; Larsen, M. R.; Packer, N. H.; Palmisano, G., Structural analysis of glycoprotein sialylation–part I: pre-LC-MS analytical strategies. *Rsc Advances* **2013,** 3, 22683-22705.
54. Loke, I.; Packer, N. H.; Thaysen-Andersen, M., Complementary LC-MS/MS-based N-glycan, N-glycopeptide, and intact N-glycoprotein profiling reveals unconventional Asn71-glycosylation of human neutrophil cathepsin G. *Biomolecules* **2015,** 5, 1832-1854.
55. Medzihradszky, K. F.; Kaasik, K.; Chalkley, R. J., Tissue-specific glycosylation at the glycopeptide level. *Mol. Cell. Proteomics* **2015,** 14, 2103-2110.
56. Greer, T.; Hao, L.; Nechyporenko, A.; Lee, S.; Vezina, C. M.; Ricke, W. A.; Marker, P. C.; Bjorling, D. E.; Bushman, W.; Li, L., Custom 4-plex DiLeu isobaric labels enable relative quantification of urinary proteins in men with lower urinary tract symptoms (LUTS). *PLoS One* **2015,** 10, (8), e0135415.

57. Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L., Mass Defect-Based N, N-Dimethyl Leucine Labels for Quantitative Proteomics and Amine Metabolomics of Pancreatic Cancer Cells. *Anal. Chem.* **2017,** 89, (2), 1138-1146.
58. Frost, D. C.; Greer, T.; Li, L., High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Anal. Chem.* **2014,** 87, 1646-1654.
59. Zeng, W.-F.; Liu, M.-Q.; Zhang, Y.; Wu, J.-Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; Chi, H.; Sun, R.-X.; Wong, C. C. L.; He, S.-M.; Yang, P., pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Scientific Reports* **2016,** 6, 25102.
60. Frost, D. C.; Greer, T.; Xiang, F.; Liang, Z.; Li, L., Development and characterization of novel 8‐plex DiLeu isobaric labels for quantitative proteomics and peptidomics. *Rapid Commun. Mass Spectrom.* **2015,** 29, 1115-1124.
61. Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L., N, N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. *Anal. Chem.* **2010,** 82, 2817-2825.
62. Yu, Q.; Shi, X.; Feng, Y.; Kent, K. C.; Li, L., Improving data quality and preserving HCD-generated reporter ions with EThcD for isobaric tag-based quantitative proteomics and proteome-wide PTM studies. *Anal. Chim. Acta* **2017,** 968, 40-49.
63. Shah, P.; Wang, X.; Yang, W.; Eshghi, S. T.; Sun, S.; Hoti, N.; Chen, L.; Yang, S.; Pasay, J.; Rubin, A., Integrated proteomic and glycoproteomic analyses of prostate cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. *Mol. Cell. Proteomics* **2015,** 14, 2753-2763.
64. Ye, H.; Boyne, M. T.; Buhse, L. F.; Hill, J., Direct approach for qualitative and quantitative characterization of glycoproteins using tandem mass tags and an LTQ Orbitrap XL electron transfer dissociation hybrid mass spectrometer. *Anal. Chem.* **2013,** 85, 1531-1539.
65. Lee, H.-J.; Cha, H.-J.; Lim, J.-S.; Lee, S. H.; Song, S. Y.; Kim, H.; Hancock, W. S.; Yoo, J. S.; Paik, Y.-K., Abundance-ratio-based semiquantitative analysis of site-specific N-linked glycopeptides present in the plasma of hepatocellular carcinoma patients. *J. Proteome Res.* **2014,** 13, 2328-2338.
66. Carlsson, S.; Fukuda, M., The polylactosaminoglycans of human lysosomal membrane glycoproteins lamp-1 and lamp-2. Localization on the peptide backbones. *J. Biol. Chem.* **1990,** 265, (33), 20488-20495.
67. Carlsson, S. R.; Lycksell, P.-O.; Fukuda, M., Assignment of O-glycan attachment sites to the hinge-like regions of human lysosomal membrane glycoproteins lamp-1 and lamp-2. *Arch. Biochem. Biophys.* **1993,** 304, 65-73.
68. Huynh, K. K.; Eskelinen, E. L.; Scott, C. C.; Malevanets, A.; Saftig, P.; Grinstein, S., LAMP proteins are required for fusion of lysosomes with phagosomes. *The EMBO journal* **2007,** 26, (2), 313-324.
69. Eskelinen, E.-L.; Illert, A. L.; Tanaka, Y.; Schwarzmann, G.; Blanz, J.; von Figura, K.; Saftig, P., Role of LAMP-2 in lysosome biogenesis and autophagy. *Mol. Biol. Cell* **2002,** 13, (9), 3355-3368.
70. Tanaka, Y.; Guhde, G.; Suter, A.; Eskelinen, E.-L.; Hartmann, D.; Lüllmann-Rauch, R.;

Janssen, P. M.; Blanz, J.; von Figura, K.; Saftig, P., Accumulation of autophagic vacuoles and cardiomyopathy in LAMP-2-deficient mice. *Nature* **2000,** 406, (6798), 902-906.
71. Tan, K.-P.; Ho, M.-Y.; Cho, H.-C.; Yu, J.; Hung, J.-T.; Yu, A. L.-T., Fucosylation of LAMP-1 and LAMP-2 by FUT1 correlates with lysosomal positioning and autophagic flux of breast cancer cells. *Cell Death Dis.* **2016,** 7, (8), e2347.
72. Miyoshi, E.; Moriwaki, K.; Nakagawa, T., Biological function of fucosylation in cancer biology. *J. Biochem.* **2008,** 143, (6), 725-729.
73. Takahashi, M.; Kuroki, Y.; Ohtsubo, K.; Taniguchi, N., Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydr. Res.* **2009,** 344, (12), 1387-1390.
74. Liu, Y.-C.; Yen, H.-Y.; Chen, C.-Y.; Chen, C.-H.; Cheng, P.-F.; Juan, Y.-H.; Chen, C.-H.; Khoo, K.-H.; Yu, C.-J.; Yang, P.-C., Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proceedings of the National Academy of Sciences* **2011,** 108, (28), 11332-11337.
75. Cairns, R. A.; Harris, I. S.; Mak, T. W., Regulation of cancer cell metabolism. *Nature Reviews Cancer* **2011,** 11, (2), 85-95.
76. Wong, N.; De Melo, J.; Tang, D., PKM2, a central point of regulation in cancer metabolism. *Int. J. Cell Biol.* **2013,** 2013.
77. Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R., Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J. Proteome Res.* **2012,** 11, 1949-1957.
78. Qin, X.; Du, Y.; Chen, X.; Li, W.; Zhang, J.; Yang, J., Activation of Akt protects cancer cells from growth inhibition induced by PKM2 knockdown. *Cell & bioscience* **2014,** 4, (1), 20.
79. Di Cristofano, A.; Pandolfi, P. P., The multiple roles of PTEN in tumor suppression. *Cell* **2000,** 100, (4), 387-390.
80. Hsu, P. P.; Sabatini, D. M., Cancer cell metabolism: Warburg and beyond. *Cell* **2008,** 134, (5), 703-707.
81. Peracaula, R.; Barrabés, S.; Sarrats, A.; Rudd, P. M.; de Llorens, R., Altered glycosylation in tumours focused to cancer diagnosis. *Dis. Markers* **2008,** 25, (4-5), 207-218.
82. Wellen, K. E.; Thompson, C. B., A two-way street: reciprocal regulation of metabolism and signalling. *Nature reviews Molecular cell biology* **2012,** 13, (4), 270-276.
83. Schneider, J.; Neu, K.; Grimm, H.; Velcovsky, H.-G.; Weisse, G.; Eigenbrodt, E., Tumor M2-pyruvate kinase in lung cancer patients: immunohistochemical detection and disease monitoring. *Anticancer Res.* **2001,** 22, (1A), 311-318.
84. Brinck, U.; Fischer, G.; Eigenbrodt, E.; Oehmke, M.; Mazurek, S., L-and M2-pyruvate kinase expression in renal cell carcinomas and their metastases. *Virchows Archiv* **1994,** 424, (2), 177-185.
85. Christofk, H. R.; Vander Heiden, M. G.; Harris, M. H.; Ramanathan, A.; Gerszten, R. E.; Wei, R.; Fleming, M. D.; Schreiber, S. L.; Cantley, L. C., The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **2008,** 452, (7184), 230-233.

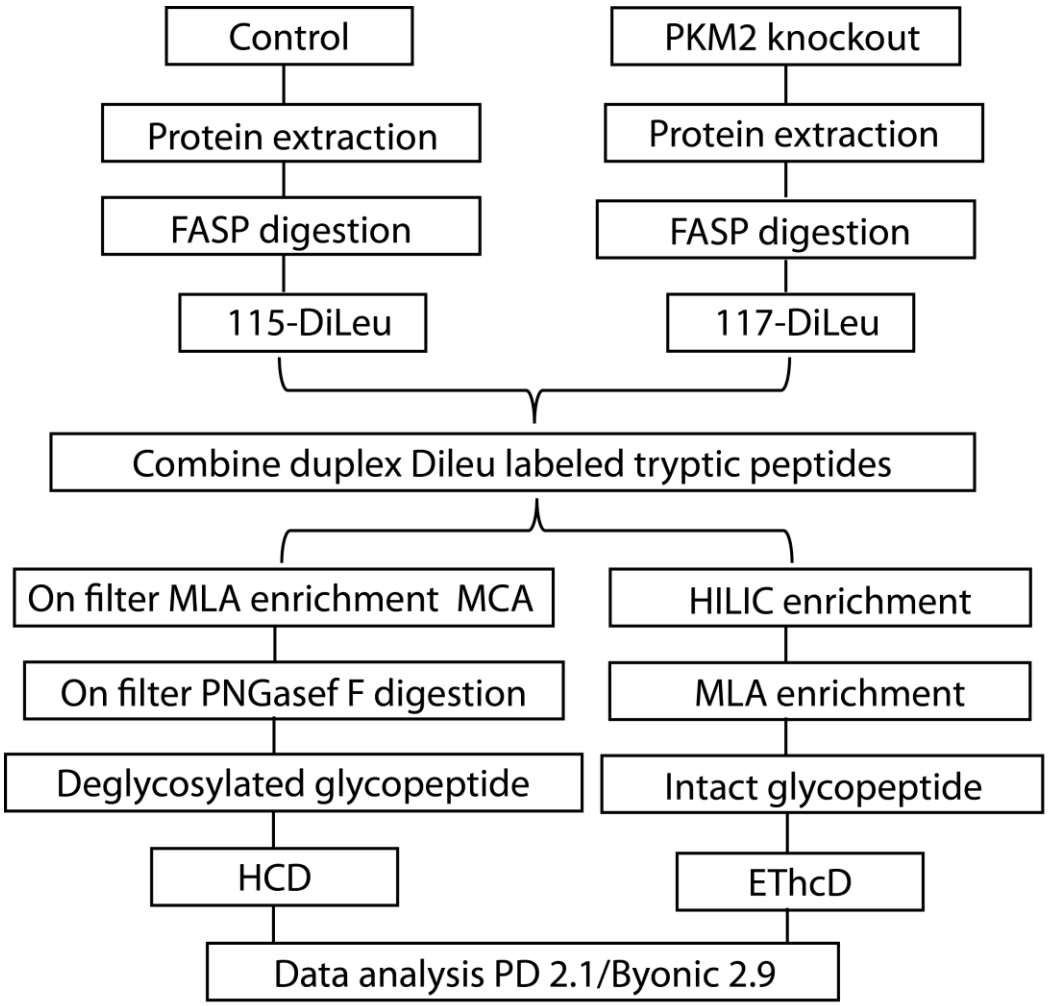


Figure 1. Schematic illustration of the workflow for intact N-glycopeptides and deglycosylated peptide analysis.

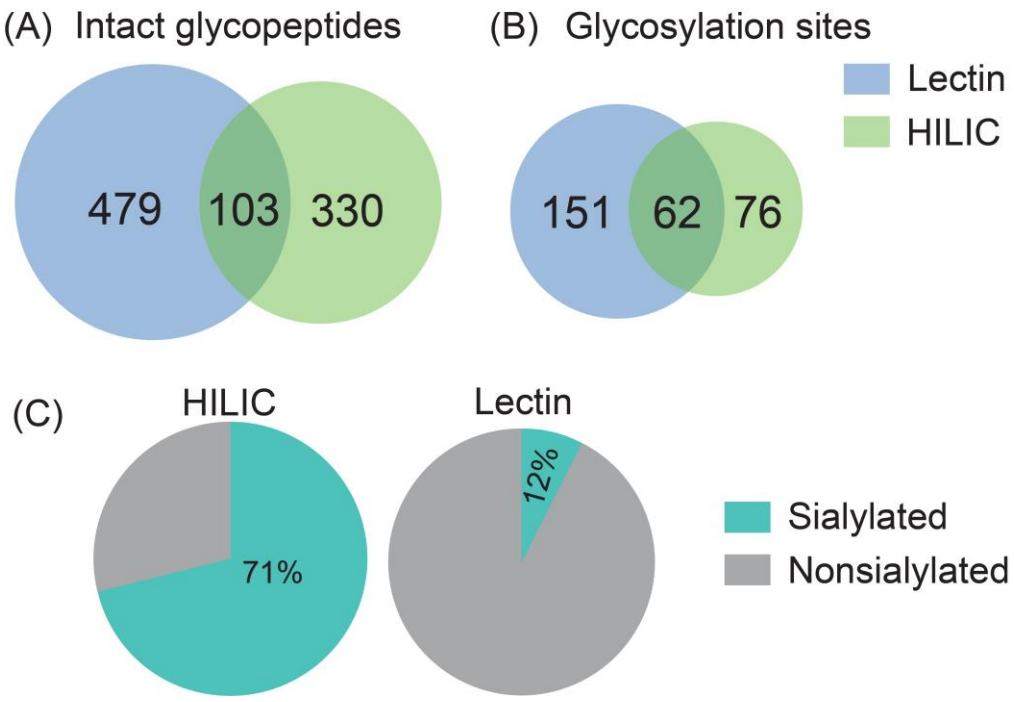


Figure 2. The Venn diagram analysis of intact N-glycopeptides (A) and N-glycosylation sites (B) enriched by HILIC and MLA methods. (C) The comparison of the percentage of sialylated N-glycopeptides enriched by HILIC and MLA respectively.

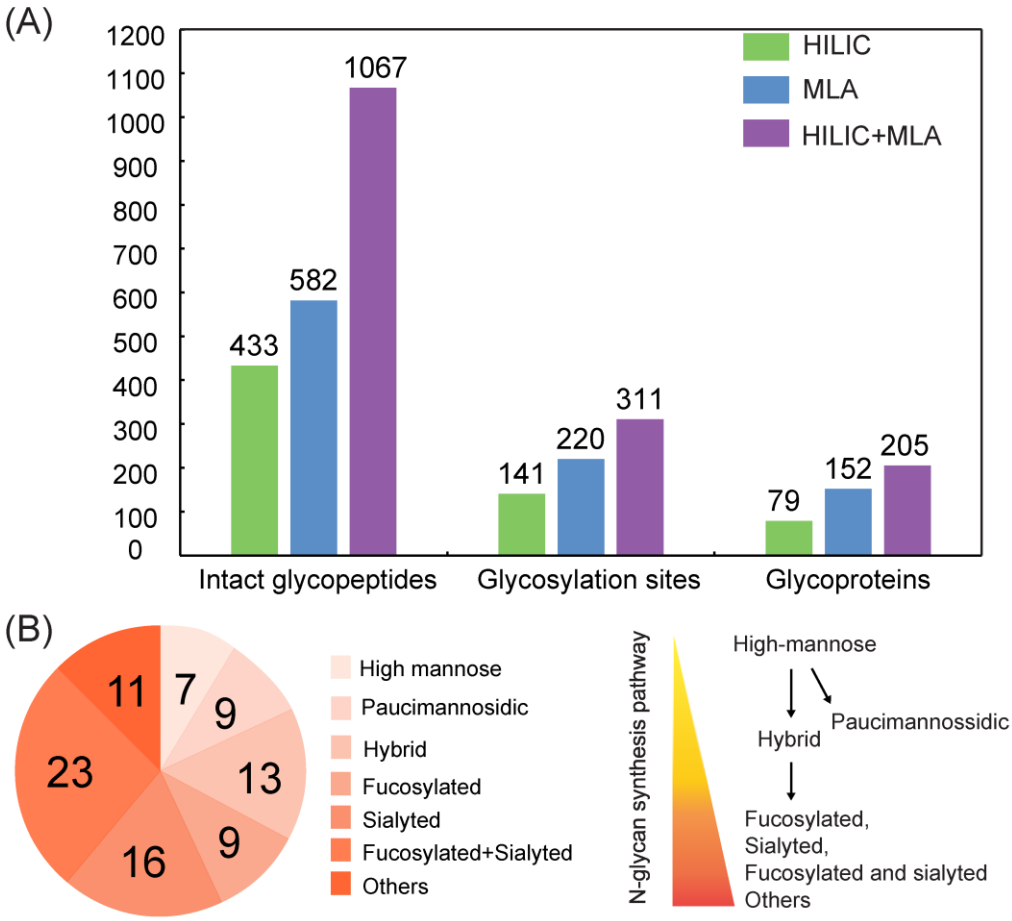


Figure 3. (A) The comparison of the number of intact N-glycopeptides, N-glycosylation sites and N-glycoproteins enriched by HILIC, MLA and sequential HILIC and MLA enrichment. (B) Different types of N-glycans along the N-glycan synthesis enriched by sequential HILIC and MLA enrichment. EThcD was used for N-glycopeptide identification.

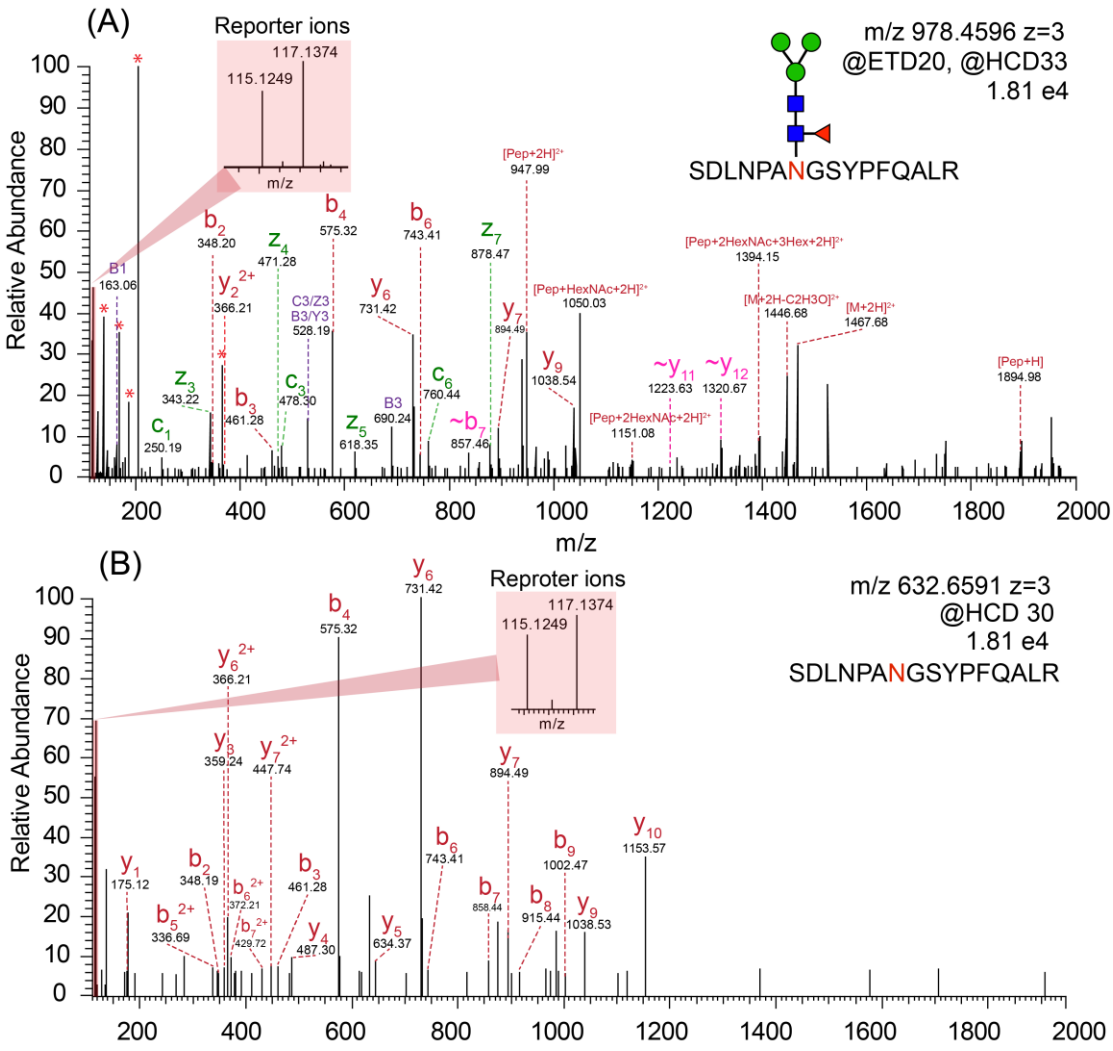


Figure 4. (A) The EThcD mass spectrum of duplex DiLeu labeled intact N-glycopeptide. (B) The HCD mass spectrum of duplex DiLeu labeled deglycosylated peptide.

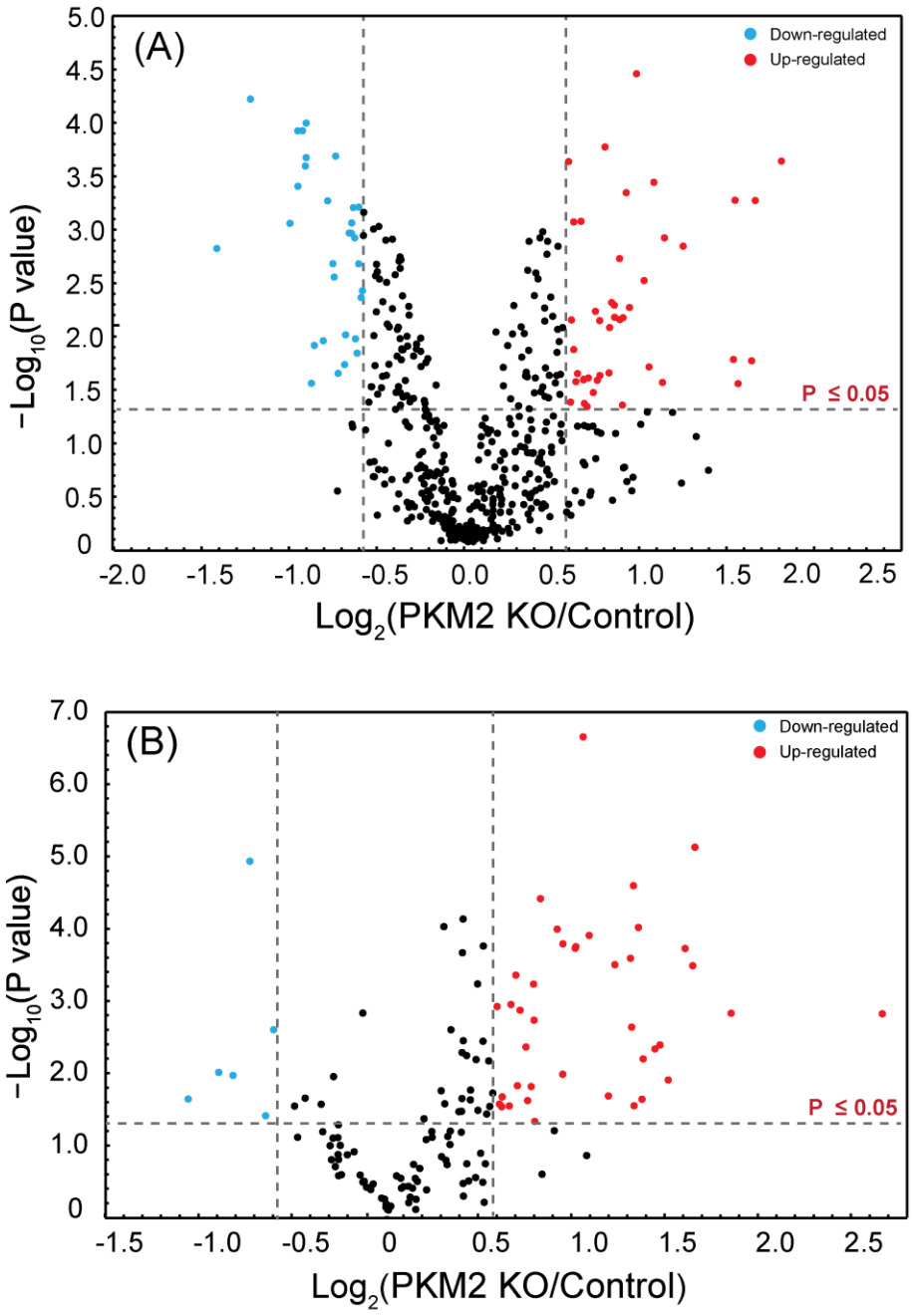


Figure 5. (A) Volcano plot of the quantitation results of deglycosylated peptides in PKM2 knockout (PKM2 KO) and parental breast cancer cells. (B) Volcano plot of the quantitation results of intact N-glycopeptides in PPKM2 KO and parental breast cancer cells.

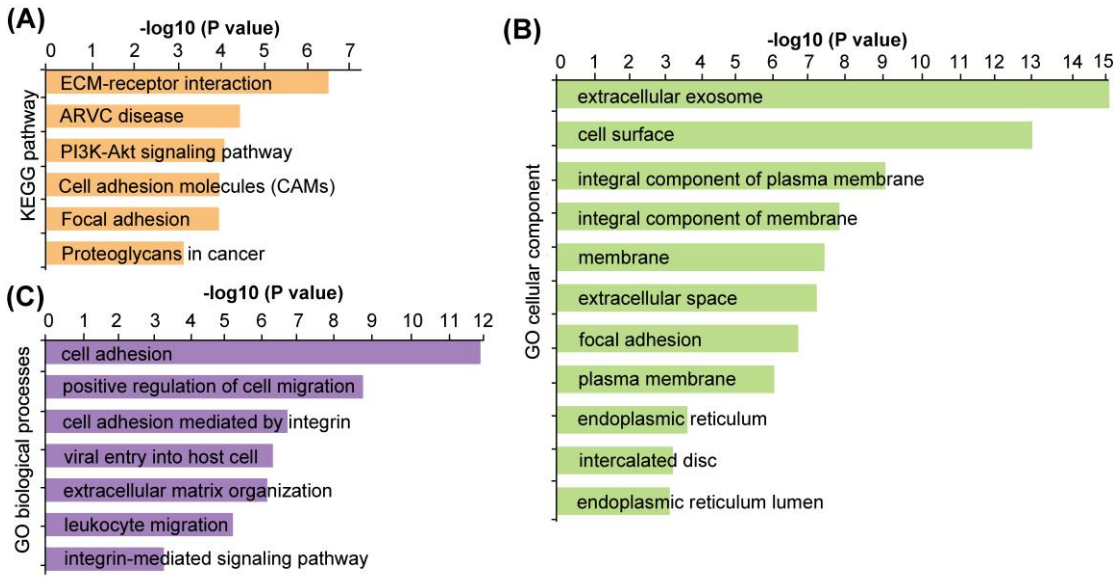


Figure 6. KEGG pathway analysis (A), gene ontology cellular component (B) and gene ontology cellular component (C) analysis of the significantly altered glycoproteins.

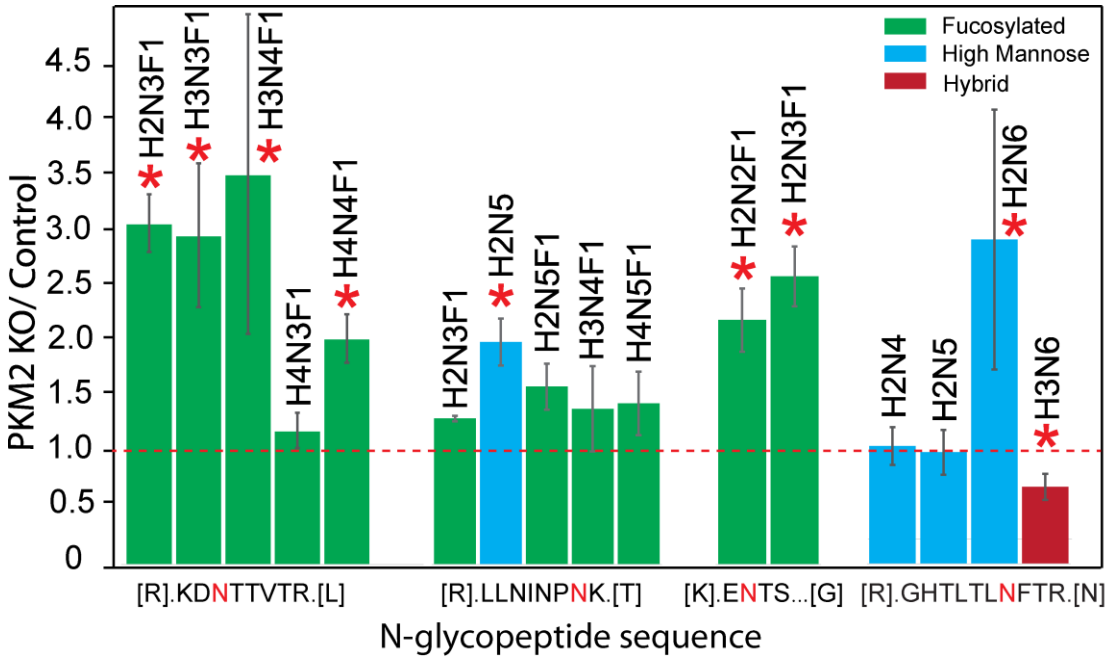


Figure 7. Quantitative analysis of intact N-glycopeptides from Lysosomal-associated membrane protein 1 (LAMP-1). The letter "N" in red denotes the N-glycosylation site.

Table 1. Altered glycoproteins involved in PI3K-Akt signaling pathway

Uniprot ID	Gene Name	Glycosylation position	Fold change	Up or down
P52798	EFNA4	P52798 [28-38]	0.63	↓
P08069	IGF1R	P08069 [223-252]	0.56	↓
		P08069 [741-753]	0.47	↓
P06213	INSR	P06213 [697-714]	1.67	↑
P06756	ITGAV	P06756 [73-88]	1.84	↑
		P06756 [865-883]	2.07	↑
P05556	ITGB1	P05556 [403-414]	1.89	↑
P18564	ITGB6	P18564 [459-466]	3.17	↑
		P18564 [567-594]	3.21	↑
O15230	LAMA5	O15230 [3106-3123]	0.61	↓
P07942	LAMB1	P07942 [1271-1283]	3.58	↑
P55268	LAMB2	P55268 [1347-1365]	0.57	↓
P07996	THBS1	P07996 [1065-1077]	0.50	↓

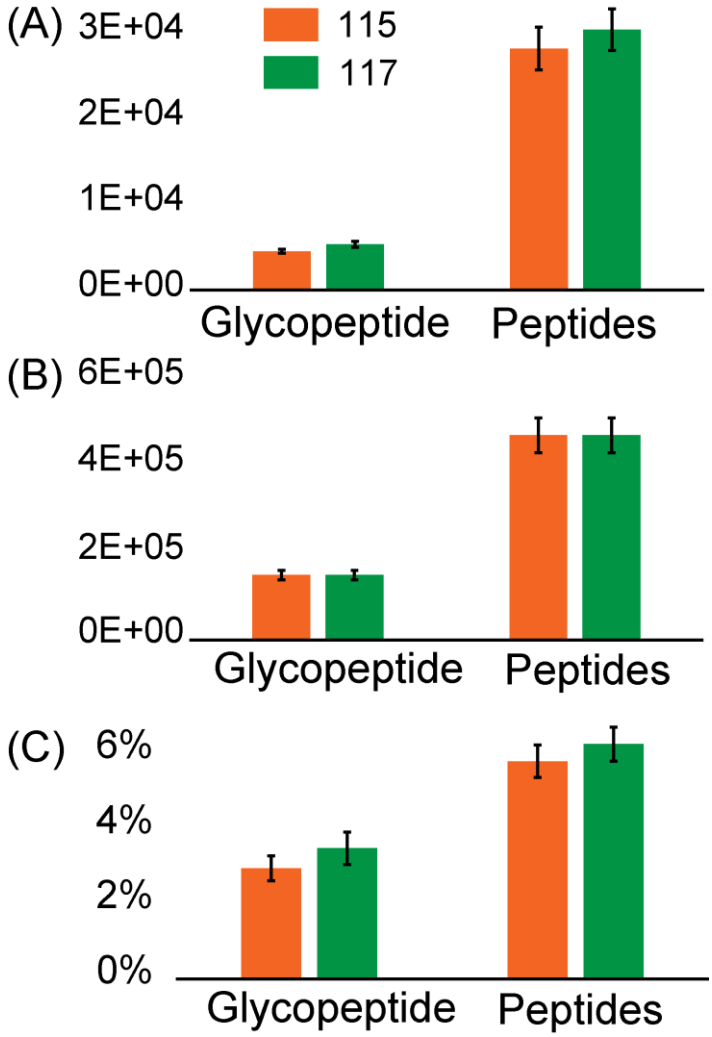


Figure S1. The reporter ions intensity (A), precursor ions intensity (B) and reporter ion yield (C) of glycopeptides and nonglycosylated peptides in the same run.

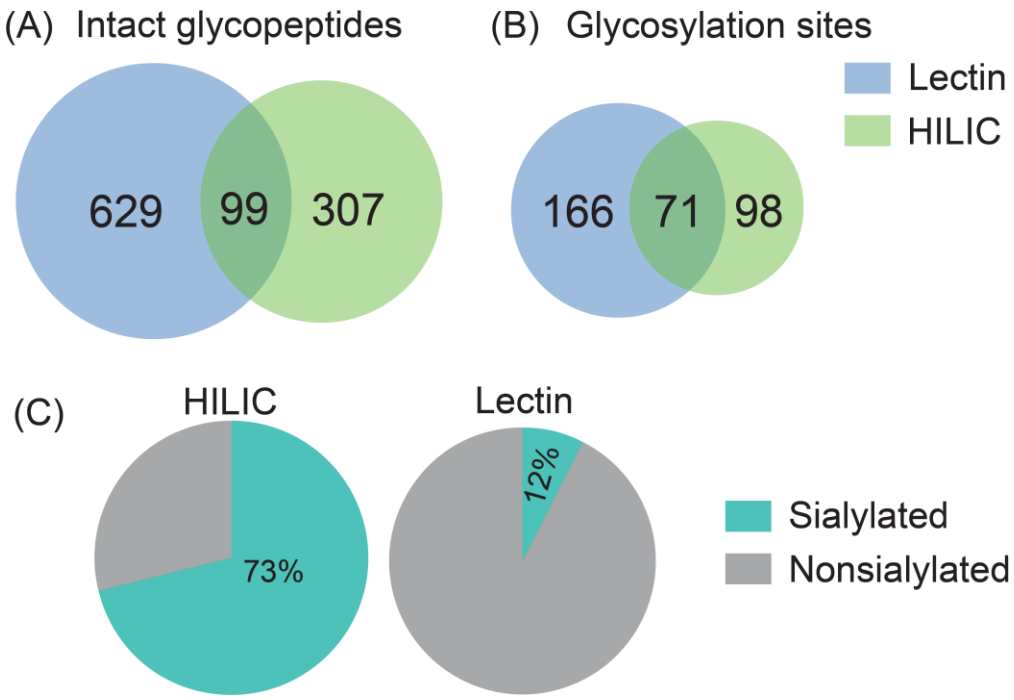


Figure S2. The Venn diagram analysis of (A) intact N-glycopeptides (A) and (B) N-glycosylation sites identified with HILIC and MLA enrichment methods. (C) The comparison of the percentage of sialylated N-glycopeptides enriched between two methods**.**

Chapter 7

In-depth site-specific analysis of N-glycoproteome in human cerebrospinal fluid (CSF) and glycosylation landscape changes in Alzheimer's disease (AD)

Adapted from **Chen, Z.**, Yu, Q., Johnson, J., Shipman, R., Zhong, X., Huang, J., Zetterberg, H., Asthana, S., Carlsson, C., Okonkwo, O., Li, L. *To be submitted to Mol. Cell. Proteomics. (2018)*

Author contribution: study was designed by Chen, Z., Zetterberg, H., Asthana, S., Carlsson, C., Okonkwo, O., Li, L; experiment was performed by Chen, Z., Yu, Q., Johnson, J., Shipman, R.,; data was analyzed by Chen, Z., Johnson, J., Shipman, R., Zhong, X., Huang, J.,; manuscript was written by Chen, Z. and edited by Yu Q., Zetterberg, H., Okonkwo, O. and L. Li.

Abstract

As the only body fluid that directly interchanges with the extracellular fluid of the central nervous system (CNS), CSF serves as a rich source for CNS-related disease biomarker discovery. Extensive proteome profiling has been conducted for CSF, but studies aimed at unraveling components of the site-specific CSF N-glycoproteome are lacking. Initial efforts into site-specific N-glycoproteomics study in CSF yielded quite low coverage, hindering further experiment design into glycosylation-based disease biomarker discovery in CSF. In the present study, we developed an N-glycoproteomics approach that combines enhanced N-glycopeptide sequential enrichment by hydrophilic interaction chromatography (HILIC) and boronic acid enrichment with electron transfer and higher-energy collision dissociation (EThcD) for large-scale intact N-glycopeptides analysis. After applying the developed approach on CSF samples, a total of 3596 intact N-glycopeptides from 676 N-glycosylation sites and 358 N-glycoproteins were identified. To our knowledge, this is the largest site-specific N-glycoproteome dataset reported for CSF to date. As accumulating evidence suggests that defects in glycosylation may be involved in Alzheimer's disease (AD) pathogenesis, we employed the method to examine the N-glycoproteome in CSF from AD patients and controls. A similar overall N-glycopeptide coverage in AD patients and controls was obtained but striking differences in certain glycoforms were evident, particularly decreased fucosylation in AD CSF. Altered glycosylation patterns were detected for a number of N-glycoproteins including alpha-1-antichymotrypsin, Ephrin-A3, carnosinase CN1, and voltage-dependent T-type calcium channel subunit alpha-1H etc., which serve as interesting targets for further glycosylation-based AD study and may eventually contribute to improved understanding

of the role of glycosylation in AD progression.

Keywords: Glycoproteomics, Mass spectrometry, Alzheimer's disease, N-glycosylation, Cerebrospinal fluid, Glycoproteins

Introduction

Originating from the brain ventricles and subarachnoid spaces around the brain and spinal cord, cerebrospinal fluid (CSF) surrounds and supports the central nervous system (CNS) [1]. It is predominantly secreted by the choroid plexuses, with a minor portion produced from cerebral interstitial fluid and cerebral capillaries [2]. For healthy individuals, around 80% of the CSF proteins are derived from plasma filtrate, while the remaining 20% originates directly from CNS [3]. Apart from acting as a protection against mechanical trauma, CSF plays an important role in brain development, regulation of brain interstitial fluid homeostasis, and neuronal functioning. It is the only body fluid that directly interchanges with the extracellular fluid of CNS, and reflects the ongoing pathological changes in the CNS most directly [4, 5]. Thus, biochemical analysis of CSF has great potential for CNS-related diseases diagnostics, such as neurological disorders [6].

The global CSF proteome has been extensively mapped, which has largely contributed to our understanding of CNS functioning under both physiological and pathological conditions [7]. Various protein post-translational modifications (PTMs) such as phosphorylation, glycosylation, methylation *etc.* greatly increase the proteome complexity and functional diversity, and analysis of these "subproteomes" would further improve our understanding of protein function. However,

the characterization of CSF "subproteomes" is still lagging behind proteome study due to underdeveloped and less sophisticated analytical technologies. Among PTMs, glycosylation represents one of the most common and complex PTMs, and acts as a key regulatory mechanism controlling protein folding, cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction, and endocytosis [8].

Defects in glycosylation in humans and their links to disease suggest that glycosylation contains a remarkable amount of biological information that can potentially help elucidate various disease mechanisms and provide potential targets for disease diagnosis and therapeutic strategies [9]. In fact, glycosylation-based biomarker discovery has achieved great success for cancer research, contributing to improved cancer diagnosis, and monitoring of malignant progression and prognosis [9]. On the other hand, alterations in protein glycosylation have also been related to human neurodegenerative disease (ND) states, such as Alzheimer's disease (AD), Parkinson disease and Creutzfeldt-Jakob disease [10, 11]. Abnormal glycosylation patterns of amyloid precursor protein (APP), tau and numerous other proteins have been reported in AD [11]. It has also been shown that O-glycosylation protects tau against aberrant phosphorylation and subsequent aggregation [12-14].

Nevertheless, there are few reports on the CSF N-glycoproteome. One study has reported 846 N-glycosylation sites from 520 N-glycoproteins in CSF after removing the glycan part with PNGase F [7]. Another study utilizing glycomics approach identified 90 N-glycan structures in human CSF [15]. These studies are either protein-targeted (deglycoproteomics) or glycan-targeted (glycomics), losing the site-specific information of each individual glycan. Initial efforts into site-specific analysis of intact glycopeptides in CSF have been made, but the depth reached is not

satisfactory, with one study showing the identification of 36 N-glycosylation sites from 23 N-glycoproteins [16], and another study showing the identification of 55 N-glycosylation sites from 36 N-glycoproteins [17]. With such limited site-specific N-glycoproteome information, the process of uncovering potential roles of different glycoproteins in the CNS is hampered, and it will also hinder the design of studies to explore disease-related glycosylation alterations.

In the present study, we developed an enhanced dedicated large-scale site-specific glycoproteomics approach for in-depth CSF N-glycoproteome analysis, including sequential HILIC and boronic acid enrichment for improved N-glycopeptide coverage, intact N-glycopeptide characterization enabled by EThcD and automated FDR-based large-scale data analysis by Byonic. This approach allows us to analyze thousands of intact N-glycopeptides from CSF proteins in a high-throughput manner, generating information about glycopeptide sequences, glycosylation site and glycan composition. In total, 3596 intact N-glycopeptides from 676 N-glycosylation sites and 358 N-glycoproteins were identified in CSF, representing the largest reported site-specific CSF N-glycoproteome dataset so far. This developed strategy was also applied to N-glycoproteome analysis of CSF samples from AD patients, allowing us to conduct a glycosylation pattern comparison between healthy control and AD. A comparable N-glycoproteome coverage was detected in AD, but very different glycoform patterns were detected for glycoproteins such as alpha-1-antichymotrypsin, Ephrin-A3, Carnosinase CN1, voltage-dependent T-type calcium channel subunit alpha-1H *etc.*, which serve as promising glycosylation-based biomarker candidates for AD.

Experimental Procedures

Chemicals and materials

Dithiothreitol (DTT) and sequencing-grade trypsin were from Promega (Madison, WI). Optimal LC/MS grade acetonitrile (ACN), methanol (MeOH) and water were from Fisher Scientific (Pittsburgh, PA). Concanavalin A (ConA), wheat germ agglutinin (WGA), *Ricinuscommunis* agglutinin (RCA120), iodoacetamide (IAA), acetyl-D18 glucosamine, D-lactose, methyl α-D-mannopyranoside and manganese dichloride were obtained from Sigma-Aldrich (St. Louis, MO). Tris base, urea (UA), sodium chloride, ammonium bicarbonate (ABC) and calcium chloride ($CaCl_2$) were obtained from Fisher Scientific (Pittsburgh, PA). Formic acid (FA), 10% Sodium dodecyl sulfate solution (SDS), trifluoroacetic acid (TFA), dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St. Louis, MO). C18 OMIX tips and phenylboronic acid (PBA) solid phase extraction cartridges were obtained from Agilent (Santa Clara, CA). Hydrophilic interaction chromatography material (PolyHYDROXYETHYL A) was obtained from PolyLC (Columbia, MD). Microcon filters YM-30 (30 kDa) and amicon Ultra-0.5 mL centrifugal filters (10 kDa) were purchased from Merck Millipore (Billerica, MA). PANC-1 pancreatic ductal adenocarcinoma cells were from ATCC (Manassas, VA).

CSF samples

32 enrollees in the Wisconsin Alzheimer's Disease Research Center (ADRC) participated in this study. The subjects comprised of 16 cognitively normal individuals who enrolled in the Wisconsin ADRC at late middle age and 16 individuals with AD dementia. Detailed subjects information can

be found in **Supplemental Table S1**. All AD participants were diagnosed via applicable clinical criteria in standardized and multidisciplinary consensus conferences. (25,26) Cognitive normalcy was determined based on intact cognitive performance by a comprehensive battery of neuropsychological tests, lack of functional impairment, and absence of neurological or psychiatric conditions that might impair cognition (27, 28). CSF was collected by lumbar puncture of individuals under written informed consent. The University of Wisconsin Institutional Review Board approved all study procedures. Each enrollee provided a signed informed consent form before participation. CSF aliquots from each of the 16 individuals at each stage were combined into a pool of 1 mL for control and AD subjects.

PANC1 cells

The proteins extracted from PANC1 cells were intended for the optimization of enrichment method. Commercially available PANC1 pancreatic ductal adenocarcinoma cells were routinely maintained in complete media of DMEM/Ham's F-12 (1:1) (ATCC) supplemented with 10% fetal bovine serum (Hyclone) and 1% antibiotic-antimycotic solution (Cellgro). Cell culture flasks were placed in an incubator containing 5% CO_2 and 98% humidity. Cells were used for a maximum of 15 passages and trypsinized using 0.25% trypsin EDTA solution (Gibco) once 80% confluence was achieved. Cell pellets were rapidly washed twice with phosphate-buffered saline, flash frozen in dry ice, and stored at -80 ℃.

Protein extraction and digestion from PANC1 cells

PANC1 cell pellets were lysed by sonication in a solution containing digest buffer (4% SDS, 100 mM Tris/Base pH 8.0). The bicinchoninic acid assay (BCA assay) was applied to determine the protein concentration. Trypsin digestion was performed based on previously reported filter-aided sample preparation (FASP) protocol [18] with some modifications. Briefly, the proteins was thawed and centrifuged at 16×000g for 5 min. Then 200 μg protein were taken out to the vial and 1 M DTT in digest buffer was added to make DTT final concentration 0.1 M. The sample was incubated at 95℃ for 3 min to reduce disulfide bonds. 200 μL of UA buffer (8 M UA in 100 mM Tris/Base) was added into the vial and transferred onto the 30 kDa filter. The filter was centrifuged at 14×000g for 15 min. Another 200 μL of UA buffer was added to the sample and centrifuge at 14×000g for 15 min. 100 μL of IAA buffer (0.05 M IAA in UA buffer) was added onto the filter; the buffer was gently swirled to mix and then incubated in darkness for 20 min followed by centrifugation at 14×000g for 10 min. One hundred μL of UA buffer was added onto the filter followed by centrifugation at 14×000g for 15 min. The procedure was repeated twice. 100 μL of ABC buffer (50 mM) was added onto the filter and centrifuged at 14×000g for 15 min. The procedure was repeated twice. All the centrifugation steps took place at 20℃. 10 μL of trypsin and 40 μL of ABC buffer was added onto the filter. The filter was incubated at 37 ℃ water bath for 18h. After incubation, the filter was transferred to a fresh collection vial and centrifuged at 14×000g for 10min. 50 μL 0.5 M NaCl solution was added onto the filter and centrifuge the filter at 14×000g for 10min. Repeat for one more time. TFA was added into the vial to make TFA final concentration 0.25%. Samples were desalted using a SepPak C18 SPE cartridge (Waters, Milford, MA).

CSF sample processing

1 mL of pooled CSF samples were separated into peptides fraction and protein fraction using 10 kDa MWCO following the previous protocol [19, 20]. The peptide fractions were analyzed in a separate study. Protein fraction were dissolved in 8 M urea, reduced (5 mM DTT, 1 h at room temperature) and alkylated (15 mM IAA, 30 min at room temperature in the dark). Alkylation was quenched by incubation in 9 mM DTT by adding a second aliquot of DTT at room temperature. The samples were diluted with 50 mM tris buffer to make urea below 1 M. Trypsin was added in a 1:50 (w/w) ratio and incubated for 18 h at 37 ℃. The digestion was quenched by the addition of TFA to a final concentration of 0.3%. Finally, the samples were desalted on a C18 SepPak cartridge (Waters, Milford, MA) and dried under vacuum.

HILIC enrichment

HILIC enrichment was conducted following a previously reported protocol with minor modifications [21]. In brief, 5 mg of HILIC beads (PolyLC) were first activated in 100 μL elution buffer (0.1% TFA) by vortexing for 30 min. Then the activated beads were washed with 100 μL binding buffer (0.1% TFA, 19.9% H_2O, 80% ACN) for two times. 100 μg tryptic peptides were dissolved in 250 μL of binding buffer and mixed with beads at a 1:50 peptide-to-beads mass ratio. Vortex for 1 h to allow N-glycopeptides bind to beads. The beads were washed with 50 μL binding buffer for 6 times. N-glycopeptides were eluted by washing the beads with elution buffer for 5 times. The volume of solvent used was scaled up accordingly when different starting material of peptides were used. The eluted N-glycopeptides were dried down under vacuum. Supernatant

during the process were collected for sequential lectin affinity enrichment or boronic acid enrichment. The separation between beads and supernatant was achieved by centrifugation.

Lectin affinity enrichment

Lectin affinity enrichment was performed based on a previously reported protocol with some modifications [22-25]. Briefly, 200 μg tryptic peptides were dissolved in 80 μL 1×binding solution (1 mM $CaCl_2$, 1 mM $MnCl_2$, 0.5 M NaCl in 20 mM Tris/Base, pH 7.3) and transferred to the 30 kDa filter. Adding 36 μL lectin mixtures (90 μg ConA, 90 μg WGA and 90 μg RCA120 in 2×biniding buffer) onto the filter. Incubate at room temperature for 1 h and unbound peptides were eluted by centrifuging at 14, 000 g for 10 min at 18℃. Wash with 200 μL binding solution for 4 times. Transfer the filter to a new collection vial. Add 100 μL sugar mixtures (300 mM N-acetyl-D-glucosamine, D-lactose, methyl α-D-mannopyranoside in 1xbinding buffer) and incubate at room temperature for 30 min. This step was repeated once. The N-glycopeptides were eluted by centrifugation. Then the samples were acidified to 0.25% TFA and desalted by C18 OMIX tip. The enriched glycopeptides were dried down under vacuum.

Boronic acid enrichment

Boronic acid enrichment was performed based on previous reported protocol with slight modifications [26]. PBA cartridges were first washed with 1 mL of anhydrous DMSO for 3 times. Tryptic peptides were dissolved in 35μL DMSO, loaded onto the cartridge, and incubated in 37℃ for 2 h with both ends of the cartridge sealed. The non-bound peptides were washed away with 1

mL anhydrous ACN for 3 times. Then the glycopeptides were eluted after incubation in 600μL of 0.1% TFA in 37℃ for 1 h. In order to elute all the bound glycopeptides, this step was repeated once. The enriched glycopeptides were dried down under vacuum.

High-pH fractionation

Enriched glycopeptides were fractionated using a C18 reverse-phase column (2.1 ×150 mm, 5 μm, 100 Å) operating at 0.3 mL/min. Samples were reconstituted in 100 μL of 10 mM ammonium formate at pH 10 (HPLC mobile phase A). Mobile phase B consisted of 90% ACN, 10 mM ammonium formate, pH=10. Glycopeptides were eluted with a gradient as follows: 1 % A (0–3 min), 1-35% (3-50 min), 35-60% (50-54 min), 60-70% (54-58 min), 70-100 % (58-59 min). Seven fractions were collected from 4 min to 62 min. The column effluent was monitored at 280 nm with a Waters 2489 UV/Visible detector. Each fraction was dried down under vacuum.

LC-MS/MS analysis

Samples were dissolved in 0.1% FA and analyzed on the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Dionex UPLC system. A binary solvent system composed of H_2O containing 0.1% formic acid (A) and MeCN containing 0.1% formic acid (B) was used for all analyses. Peptides were loaded and separated on a 75 μm x 15 cm homemade column packed with 1.7 μm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661). The LC gradient for intact N-glycopeptides was set as follows, 3%-30% A (18-98min), 30%-75% A (100-108 min) and 75%-95% A (108-118min). The

mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS acquisition. For intact N-glycopeptides analysis, an MS1 scan was acquired from 400–1800 (120,000 resolution, $4e^5$ AGC, 100 ms injection time) followed by EThcD MS/MS acquisition of the precursors with the highest charge states in an order of intensity and detection in the Orbitrap (60,000 resolution, $3e^5$ AGC, 100 ms injection time). EThcD was performed with optimized user defined charge dependent reaction time (2+ 50 ms; 3+ 20 ms; 4+ 20 ms; 5+ 20ms; 6 + 9 ms; 7+; 9 ms; 8+ 9ms) supplemented by 33% HCD activation.

Data analysis

All raw data files were searched against UniProt *homo sapien*s reviewed database (08.10.2016, 20, 152 sequences), using PTM-centric search engine Byonic (version 2.9.38, Protein Metrics, San Carlos, CA) incorporated in Proteome Discoverer (PD 2.1). Trypsin was selected as the enzyme and two maximum missed cleavages were allowed. Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.01 Da. Static modifications consisted of carbamidomethylation of cysteine residues (+57.02146 Da). Dynamic modifications consisted of oxidation of methionine residues (+15.99492 Da), deamidation of asparagine and glutamine (+0.98402 Da), and N-glycosylation on asparagine. Oxidation and deamidation were set as “rare” modification, and N-glycosylation was set as “common” modification through Byonic node. Two rare modification and one common modification were allowed. Human N-glycan database embedded in Byonic, which contains 182 glycan entities, was used. Only these N-glycopeptides with PSMs with an FDR ≤ 1% and Byonic score over 50 were reported, and such criteria were set

based on the expertise experience of manual spectra check and previous evaluation of the software [27]. Each glycopeptide identified should have the consensus motif NX/T/S, X≠P.

Experimental design and statistical rationale

The experiment design is summarized in **Fig. 1** and **Fig. S1**. Firstly, different N-glycopeptide enrichment methods and their combinations were evaluated in order to get an optimized enrichment strategy to maximize the glycoproteome coverage. The most commonly used enrichment methods including HILIC, lectin and boronic acid were evaluated respectively, and the benefits of high pH fractionation were also investigated. Secondly, in order to obtain an overall picture of human CSF N-glycoproteome, CSF samples from 16 healthy individuals were pooled together to obtain 1 mL sample. Trypsin digestion and following optimized enrichment strategy were applied to the CSF samples. A total of 5 fractions of enriched glycopeptides were collected, including 1 from HILIC enrichment and 4 fractions from boronic acid enrichment. For each fraction, two technical replicates were performed for LC-MS/MS analysis. Meanwhile, the same amount of CSF samples from 16 AD patients were processed in parallel to get a landscape N-glycoproteome picture in AD state. As a descriptive and qualitative comparison of site-specific N-glycoproteome between healthy control and AD state were conducted, no statistical tool was used. Further discussions could be found in the *Discussion* section.

Results

Optimization of enrichment strategies

Although advances in various analytical technologies have made large-scale analysis of glycopeptides feasible, the depth of glycoproteome study in complex sample is still lagging behind

other PTMs studies, such as phosphoproteomics. This is because glycopeptides only constitute a minor portion (2% to 5%) of the total peptides mixtures, and the signal is often suppressed in the presence of other more abundant non-glycosylated peptides [28, 29]. The glycoform heterogeneity (microheterogeneity) at each glycosylation site will further reduce the relative abundance of each unique glycopeptide. Thus, glycopeptide enrichment is a key step in the success of glycopeptide analysis in complex biological samples. Many enrichment strategies have been developed in recent years, including hydrophilic interaction chromatography (HILIC), titanium dioxide (TiO_2) affinity, lectin affinity, hydrazide and boronic acid chemistry [30]. Among them, HILIC and lectin affinity enrichment are two of the most widely used strategies, with enrichment occurring at the intact glycopeptide level fully preserving the native glycan information. Due to the inherent glycopeptide structure complexity, single enrichment strategy is usually not enough to enrich all kinds of glycopeptides, suggesting multiple enrichment strategies should be used to maximize the glycoproteome coverage of a certain sample. To further reduce sample complexity and improve glycoproteom coverage, off-line fractionation such as high-pH fractionation (HpH) is often utilized due to its ease of implementation and its capacity for large amounts of starting material, which has also been shown highly orthogonal to the following LC-MS/MS analysis using low-pH reversed-phase chromatography [31].

Proteins extracted from PANC1 cells were used for method optimization. As a starting point, a sequential enrichment with HILIC and lectin followed by HpH fractionation was employed for N-glycoproteome analysis of 200 μg proteins from PANC1 cells. In total, 732 intact N-glycopeptides were identified using 7 fractions and only a slight improvement was found after

increased to 10 fractions with 777 N-glycopeptides identifications (**Fig. 1**). Lectin affinity enrichment relies on certain glycan motif recognition, while HILIC takes advantage of increased hydrophilicity due to glycan attachment. Both of these two methods are somewhat biased towards certain categories of glycopeptides, and these glycopeptides with lower lectin binding affinity or more hydrophobic peptide sequences may be lost. Besides, hydrophilicity overlapping region exists between glycopeptides and non-glycosylated peptides, which results in non-specific enrichment of non-glycosylated peptides hampering the glycopeptide detection [32]. Compared to HILIC and lectin enrichment, boronic acid enrichment is a less biased and more universal enrichment strategy. Boronic acids enrichment works by forming a strong reversible covalent bond with 1,2 or 1,3 cis-diols present in any glycan moiety under alkaline conditions [33]. The reversible property of the covalent bond allows the intact glycopeptides to be easily released under acidic conditions without any side effect.

With the same amount of starting material (200 μg protein), boronic acid enrichment with 7 fractions resulted in 1236 intact N-glycopeptide identifications, a 59% increase compared to HILIC and lectin enrichment (**Fig. 1**). A closer look at the number of N-glycopeptides identified in the 7 HpH fractions showed that N-glycopeptides were mainly in fractions 2, 3 and 4 as shown in **Supplemental Fig. S2-a**, whereas the number of peptides identified are almost evenly distributed in 7 fractions in a typical proteomics study as shown in **Supplemental Fig. S2-b**. This is not surprising considering the increased hydrophilicity of glycopeptides, which results in an earlier elution on a C18 column. Then, another round of enrichment combining fractions 1, 5, 6, and 7 with four technical replicates was conducted, resulting in a further increase of N-

glycopeptide identifications (1289 N-glycopeptides). The results showed that boronic acid enrichment alone yielded higher N-glycopeptide coverage than combined HILIC and lectin enrichment.

Further investigation was needed to decide which enrichment method, HILIC or lectin, provides the most complementary enrichment capability to boronic acid enrichment? A comparison between these three enrichment strategies (**Supplemental Fig. S3**) showed that although HILIC gave the lowest glycopeptide identifications, the number of sialyated glycopeptides identified was the highest among these three methods. This is probably because sialic acid increases the hydrophilic interactions between the glycopeptides and HILIC beads, resulting in a preferential enrichment of sialylated N-glycopeptides [34]. On the other hand, although boronic acid enrichment gives the highest total glycopeptide coverage, the percentage of sialylated glycopeptides is only 16% (**Supplemental Fig. S4**). Thus, we propose that combining HILIC with boronic acid enrichment will further increase the N-glycopeptide coverage, especially sialylated glycopeptides. As shown in **Fig. 1**, after combining HILIC enrichment with boronic acid enrichment, a total of 1422 N-glycopeptides were identified, with an increase of 133 N-glycopeptides compared to boronic acid enrichment alone. Benefiting from using HILIC, the number of sialylated glycopeptides identified had a 1.8-fold increase (**Supplemental Fig. S4**). Such improvements are significant when a larger number of sialylated glycopeptides exist in the sample, such as in the case of CSF.

Site-specific intact N-glycopeptide characterization

Since its introduction by Heck and co-workers [35], electron transfer and higher-energy collision dissociation (EThcD) has shown great potential for labile PTMs analysis (i.e. phosphorylation) with improved site localization, as well as generating richer backbone spectra [36]. When applied to intact glycopeptide characterization, EThcD can produce rich fragment ions information for glycan (B/Y ions), peptide (b/y, c/z ions) and glycosylation site (c/z ions) identification, providing the opportunity for site-specific intact glycopeptide analysis [37, 38]. The representative EThcD spectra, deriving from N-glycopeptides with three main N-glycan classes including high-mannose, hybrid and complex types attached, were shown in **Fig. 2.** A series of fragment ions including c/z, b/y ions, glycan fragment ions and glycopeptide with one or more loss of monosaccharides were detected. The highly abundant oxonium ions were detected in the lower mass region, including 138.06 (HexNAc-$2H_2O$-CH_2O), 168.06 (HexNAc-$2H_2O$), 186.08 (HexNAc-H_2O) and 204.09 (HexNAc), confirming that the spectra belonged to a glycopeptide. For the fucoyslated and sialylated glycopeptide, the signature fragment ions supporting the presence of fucose (i.e. HexHexNAcFuc) and sialic acid (i.e. NeuAc, NeuAc-18) were used to further confirm their identity apart from the accurate precursor mass match. Notably, besides the glycopeptides with N-glycan classes synthesized in the classical pathway, glycopeptide with truncated N-glycan (GlcNAc0-2Man0-3Fuc0-1) in an unconventional pathway were also detected and confidently identified in a site-specific manner. As shown in **Fig. 2a**, the glycan oxonium ions 138.06, 168.06, 186.08 and 204.09 from HexNAc and fucose containing glycan fragment ion 553.23 (HexNAc2Fuc1) along with the precursor accurate mass match (0.6 ppm), confirms the composition of the glycan as HexNAc2Fuc1. The intact glycopeptide ion with a HexNAc loss

(peptide + HexNAcFuc, +2, 1134.58) indicates that the fucose group was attached to the innermost HexNAc. Peptide sequence SVVAPATDGGLNLTSTFLR was deduced based on the abundant backbone fragment ions including both b/y and c/z ions. The peptide, belonging to prostaglandin-H2 D-isomerase, has an N-glycan consensus motif at 12th asparagine (Asn, N) residue, and it has been reported as an N-glycosylation site previously; however, he truncated N-glycan form has never been reported on this site [16]. The presence of a glycosylation site-containing c/z ions c12, c13, c14, c15 with intact N-glycan preserved and c11 ion without N-glycan further helps unambiguously identify the intact glycopeptide in a site-specific manner.

Site-specific glycoform mapping in CSF

As discussed previously, the current depth of site-specific N-glycoproteome in CSF is far less than satisfactory, which will largely hinder glycosylation-based biomarker discovery studies in CSF for various diseases. Here, the developed enhanced glycoproteomics strategy was applied to the in-depth site-specific N-glycoproteome analysis of the CSF sample. In total, 3596 intact N-glycopeptides from 676 N-glycosylation sites and 358 N-glycoproteins were identified (**Fig. 3a**), representing the largest site-specific CSF N-glycoproteome dataset so far.

To obtain an overview of the cellular component, molecular functions and biological processes of the 358 N-glycoproteins identified, Gene Ontology (GO) function enrichment analysis was conducted using DAVID tool (http://david.abcc.ncifcrf.gov/) [39]. The majority of the identified glycoproteins are mainly distributed in extracellular region and membrane or cell surface, agreeing well with the fact that most of N-glycoproteins are membrane or secreted proteins [40]. In terms of

molecular functions, the top six enrichment clusters are heparin binding, integrin binding, serine-type endopeptidase inhibitor activity, calcium ion binding, cell adhesion molecule binding and serine-type endopeptidase activity. Many biological processes that are known to involve glycosylation are enriched, including cell adhesion, platelet degranulation, negative regulation of endopeptidase activity, extracellular matrix organization, axon guidance.

In terms of the glycosylation site distribution among the glycoproteins, more than half of them (54%) carry only one glycosylation site, and the majority (more than 90%) of them have less than or equal to 5 glycosylation sites **(Fig. 3c)**. Only 10 N-glycoproteins carry more than 5 glycosylation sites, which include clusterin, voltage-dependent calcium channel subunit alpha-2/delta-1, prolow-density lipoprotein receptor-related protein 1, galectin-3-binding protein contactin-1, BDNF/NT-3 growth factors receptor, multiple epidermal growth factor-like domains protein 8, neural cell adhesion molecule L1-like protein, neuronal cell adhesion molecule and Ig GFc-binding protein (FcγBP). The results were in accordance with previous findings that these proteins were highly N-glycosylated [41, 42]. Among them, FcγBP was found to carry the most number of N-glycosylation sites, with 11 N-glycosylation sites detected. According to sequence analysis, it could have 33 potential N-glycosylation sites with typical N-glycan consensus motif (NXT/S, X≠P), and 9 N-glycosylation sites were recorded in the Uniprot without any microheterogeneity information available [42]. FcγBP is a secretory mucin-like glycoprotein present widely throughout mucous membranes and in external body fluids [43]. The importance of O-glycosylation has long been recognized in the processing and biological properties of mucins, and the role of N-glycosylation starts to come to light and has been found to support mucin stability, folding, sorting, membrane

trafficking, and secretion recently [44, 45]. Due to a lack of glycosylation microheterogeneity information available at each site, further study into its more specific and precise biological role is hampered [46]. A total of 22 N-glycoforms were identified on the detected 11 N-glycosylation sites, with 4 N-glycoforms on 3719Asn, 3 N-glycoforms on 3339Asn, 4540Asn and 5186Asn, 2 N-glycoforms on 2138Asn and 2518Asn, and 1 N-glycoform on 75Asn, 91Asn, 1743Asn, 2944Asn and 4145Asn, respectively. With the site-specific glycoform information available, it provides an opportunity for a further and more complete biological role investigation for the N-glycosylation as well as FcγBP as a whole glycoprotein.

On the other hand, a total of 3311 unique glycoforms from 676 N-glycosylation sites were detected from these glycoproteins after removing the redundant N-glycopeptides (glycopeptides with the same glycosylation site and glycan but different peptide sequence length were combined). On average, each N-glycosylation site carried ~4.9 glycoforms and each N-glycoprotein carried ~9.2 glycoforms, suggesting a highly diverse microheterogeneity. More than half of the glycosylation sites (62%) carried less than or equal to 2 glycoforms, whereas there were 12% glycosylation sites with more than 10 glycoforms and 1.5% glycosylation sites with more than 50 glycoforms (**Fig. 3d**). For example, 68 glycoforms were identified at 93Asn on alpha-1-acid glycoprotein 2 (AGP-2), including 58 complex, 6 hybrid, 2 high-mannose and 2 truncated N-glycans. AGP-2 is an acute-phase glycoprotein containing 45% carbohydrate and was once considered to be the protein with the highest carbohydrate content [47]. It mainly contains complex N-glycans and their variation (branched, sialylated and fucosylated) has been shown to be sensitive to various pathophysiological conditions [48, 49]. There were 40 sialylated, 31 fucosylated, and 30

highly-branched (tetra-antennary or more) N-glycoforms detected among the 58 complex N-glycoforms, which indicates AGP-2 has plenty of flexibility altering the N-glycosylation pattern in response to different physiological or external stimuli.

Most glycoforms detected were complex N-glycans (76%), with a minor percent of high-mannose (7%) and hybrid N-glycans (7%) (**Fig. 3b**). Such distinct difference in their abundance results from the N-glycan biosynthesis pathway and their different molecular function. Synthesized early in the glycan biosynthesis pathway, high-mannose glycans plays important role in assisting protein folding in the endoplasmic reticulum (ER) and also protects the proteins against degradation during intracellular transport [50, 51]. However, a majority of these high-mannose N-glycans will be further trimmed by mannosidases in cis-Golgi to give Man5GlcNAc2, a key intermediate in the pathway to hybrid and complex N-glycans [52]. A small part of these high-mannose N-glycans are not fully processed to Man5GlcNAc2 escaping further modification, resulting in a small percentage of mature membrane or secreted glycoprotein carrying high-mannose N-glycans, in accordance with the 7% high-mannose glycoforms detected in our study. In the medial-Golgi and trans-Golgi, further modification will generate an extensive array of mature hybrid and complex N-glycans, differing in branch number, composition, and capping arrangements and so on, fulfilling the needs for various functions needed for different biological processes [52]. In fact, constituting the 76% complex N-glycans detected in the present study, 25% of them are bi-antennary, 29% tri-antennary and 22% tetra-antennary or more branches. Besides, as two of the most common and important “capping” reactions to elongate the glycan branch, sialylation and fucosylation plays a key role both in biological process such as cellular recognition,

cell adhesion, cell signaling, as well as in altered glycosylation associated disease development [9]. In CSF, 47% and 51% glycoforms detected in our study were fucosylated and sialylated, respectively. Such a high percentage of sialylated glycopeptides detected in our study resulted from the sequential enrichment using HILIC and boronic acid enrichment, especially the preferential enrichment of sialylated glycopeptides by HILIC. Besides the majority of glycoforms detected in the main N-glycan synthesis pathway, there were 10% paucimannosidic glycoforms (Fuc0-1Man0-3GlcNAc1-2) detected in the unconventional truncation pathway. Different from the well-defined N-glycan synthesis pathway, this kind of protein N-glycosylation, also referred to as paucimannosylation, was discovered initially in plants and invertebrates [53, 54]. Growing evidence have shown that paucimannosylation are also present in mammals, including mouse embryonic neural stem cells, human buccal epithelial cells, human colorectal cancer epithelial cells *etc.* [55-59]. Although paucimannosylation in the extracellular environment was believed to be a feature of cells to communicate within the immune system and altered expression of paucimannosidic epitopes was implicated in cancer, their exact physiological functions and roles in diseases remains largely unknown [60]. The discovery of these truncated N-glycans in CSF proteins lays a foundation for their further biological role studies.

Glycosylation alteration in AD

As discussed previously, numerous evidence has demonstrated glycosylation alteration is implicated in the pathophysiological development of AD. To get a landscape picture of glycosylation pattern of CSF in AD subjects, CSF samples pooled from 16 AD subjects were

subject to the same in-depth N-glycoproteome analysis. To minimize the variations brought by sample preparation, the AD CSF samples were processed simultaneously with control CSF samples. In total, 3683 N-glycopeptides, 718 N-glycosylation sites and 369 N-glycoproteins were identified in AD CSF (**Fig. 4**), which is quite comparable with the coverage in control CSF. Venn diagram analysis between AD and control indicates that extensive glycoforms changes and glycosylation on/off switch exist during AD development, with 58% overlapping for N-glycopeptides, 68% for N-glycosylation sites, and 77% for N-glycoproteins (**Fig. 4**). This results also suggest that an in-depth glycoproteomics study that focus on characterization of the glycofomrs changes is necessary before any quantitative glycoproteomics study is conducted for any site-specific glycoproteomic study. Based on the degree of branching, sialylation, fucosylation and composition complexity, these glycoforms were divided into 12 categories. As shown in **Fig. 4**, compared to control CSF, the fucosylated (-6.8%), sialylated (-2.4%), fucose only (-5.0%) and complex type (-2.2%) in AD CSF has a decrease over than 2%, while only paucimannosidic type has an increase over than 2% percent (+2.7%).

For the detected 463 N-glycosylation sites shared by control and AD, a heat map analysis was conducted to give a bird's view of the landscape glycosylation alterations in AD **(Fig. 5)**. At first glance, the overall glycoforms distributions may appear the same, but the difference in the number of various glycoforms on each site started to show when looking more closely. These spots marked on the heat map represented the site-specific glycoform difference between control and AD and may act as promising glycosylation-based biomarker candidates. To get a more clear idea of the identify of these altered N-glycoproteins/N-glycosylation sites and their glycoform changes, a two-

dimensional plot depicting the changes in the number of glycoforms as a function of the number of glycoforms in AD was constructed (**Supplemental Fig. S5**). These interesting glycoprotein candidates with glycosylation pattern changes are further discussed later. Besides the shared N-glycoproteins/N-glycosylation sites detected in both control and AD, there were 213 unique N-glycosylation sites detected for control and 215 unique N-glycosylation sites for AD. And 81 unique N-glycoproteins were detected for control and 92 unique N-glycoproteins for AD, respectively**.** Such N-glycoproteins could also be potential markers for AD as glycosylation on/off switching mechanism exists to regulate different biological process or disease states [52]. Selective interesting potential biomarker candidates with altered glycosylation patterns are discussed below.

Discussion

The current knowledge of site-specific protein glycoforms in CSF is quite limited due to the inherent structure complexity of intact glycopeptide and analytical technologies or workflows lagging behind other PTMs studies. Benefiting from the improved workflow, including optimized sequential glycopeptide enrichment and intact N-glycopeptide characterization enabled by EThcD, thousands of intact N-glycopeptides from CSF were identified. In order to map the site-specific glycoforms of proteins in human CSF, an in-depth glycoproteomics analysis was conducted using the CSF samples pooled together from 16 healthy subjects. The identified 3596 intact glycopeptides from 358 N-glycoproteins represents the largest site-specific N-glycoproteome dataset so far. As many studies have shown the glycosylation alteration is implicated in AD development, it will be quite interesting to get a picture of the N-glycoproteome landscape of CSF

in AD and compare it with healthy control.

Unlike in bottom-up proteomics study, where different proteoforms contributed by PTMs were not considered and quantitation was at the protein level, glycoproteomics study targets at each individual intact glycopeptide, which involves the characterization of various glycoforms at each site [61]. Due to this reason, there are two aspects for glycosylation alterations in glycoproteomics study, the abundance changes of the same glycoform at the same site and the glycoform changes at the same site, and the latter includes increased/decreased sialylation, fucosylation, branching *etc.* [9]. Therefore, for any glycoproteomics-driven biomarker discovery study, an in-depth glycoproteomics study is necessary in order to get an idea what are the glycoforms out there and to what extent the glycoform changes are involved in the disease state and control state before a quantitative study is conducted.

Thus, an in-depth qualitative glycoproteomics study was conducted focusing on exploring the glycoform changes in AD patients using the CSF samples pooled from 16 age-matched AD patients. Although ideally an in-depth glycoproteomics study conducted in individual control and AD subject would allow us to account for any individual variation, the limited CSF sample amount available, instrument/labor time and financial constraints are the limiting factors that make such experiments difficult to carry out [62]. Additionally, pooling the samples together will increase the chance of the detection of low abundant glycoproteins. In total, a comparable total number of intact N-glycopeptides were identified compared to control CSF, but with a very different percentage constitute of different categories of glycoforms, indicating AD was affecting the body by manipulating glycoform changes in a subtle and delicate way instead of massive total glycoform

number changes. There are a few glycoproteins showing altered glycosylation patterns. Here, some of the interesting targets were selected for further discussion.

As a member of the serine protease inhibitor family of acute phase proteins, alpha-1-antichymotrypsin (ACT) is predominantly synthesized in the liver, and also produced in the brain by the astrocytes [63]. Studies have shown increased ACT serum and CSF in AD patients, indicating ACT may act as a biomarker for early AD diagnosis [64]. *In vitro* experiments have demonstrated that ACT binds to Aβ peptides and promotes the assembly of the Aβ peptides into amyloid filaments [65, 66], and later *in vivo* studies in transgenic AD mouse models confirmed ACT is an integral component of the amyloid deposits and accelerates amyloid plaque formation [67, 68]. Additionally, ACT has also been shown to induce tau phosphorylation in neurons and subsequent neuronal cells apoptosis [69]. ACT is an N-glycoprotein with an estimated 24% content of carbohydrates content distributed among the six potential N-glycosylation sites [70]. The glycan moiety compositions have been partially revealed by previous studies of the extracted plasma ACT using affinity immune-electrophoresis and NMR spectroscopy [71, 72], showing evidence for disialyl diantennary, trisialyl triantennary, disialylated triantennary structures. Glycosylation patterns of accurate phase proteins (i.e. ACT) in response to chronic inflammatory diseases (i.e. AD) has been extensively studied as potential biomarkers and reviewed recently [73]. In fact, altered glycosylation profiles (reduced terminal GlcNac and sialic acid) have been detected in the purified plasma ACT in AD patients [74]. However in the previous study, lectin-assisted glycan array analysis was used, which cannot provide exact microheterogeneity alterations information for each site. No studies to date have explored the glycosylation alterations in CSF ACT. Benefiting from the site-specific

glycosylation analysis of the workflow, our study was able to show the glycosylation micro-heterogeneity alterations on each site. In total, 4 N-glycosylation sites (Asn106, Asn127, Asn186, Asn271) were detected for ACT in both AD patients and healthy control. The results show that there is an increase in the complex N-glycans, especially sialylated glycans, on site 106 and 186, while a comparable number of N-glycans were detected on site 127 and 271 **(Fig. 6).**

Another interesting glycoprotein target with altered glycosylation in AD is ephrin A3, a membrane ligand for Eph receptors. Although ephrin A3 has been shown to have 3 potential N-glycosylation sites (Asn38, Asn67, Asn100) by Uniprot sequence analysis, there is no experimental evidence so far. In accordance with the Uniprot sequence analysis, our results identified Asn38 as the glycosylation site in CSF ephrin A3. A total of 12 glycoforms were identified on this site, while only 4 gycoforms were identified in AD, with a decreased degree of glycosylation, especially fucosylation and sialylation. The Eph receptors belong to the superfamily of transmembrane Tyr kinase receptors, and the Eph/ephrin pathway mediates short-distance cell-cell communication after activation [73] and regulates various developmental processes, including cardiovascular and skeletal development, as well as axon guidance, synapse formation, maintenance and plasticity in the nervous system [75]. Studies have shown dysregulation of Eph/ephrin signaling could lead to synaptic deficits associated with AD and suggest that Eph/ephrin signaling could act as a target for new therapeutic opportunities for AD [76, 77]. Glycosylation of ephrin has been shown to play important role in the Eph/ephrin signaling. A study into the interaction between ephrin A1 and EphA2 demonstrates that deglycosylation of ephrin A1 decreased its binding affinity to EphA2 and did not activate the downstream signaling pathways

[78]. Analysis of Eph/ephrin crystal structures revealed an interaction between the ligand's carbohydrates and two residues of EphA2 receptor [76, 78]. Our results indicate that decreased ephrin glycosylation in AD may lead to dysregulated Eph/ephrin signaling, and thus contribute to AD progression.

A decreased glycosylation, mainly fucosylation, of glycoprotein carnosinase CN1 was also found in AD. Studies of human plasma carnosinase CN1 have shown there were two potential N-glycosylation sites (Asn322, Asn382) [79], with Asn382 identified in our study. Carnosinase CN1 is a secreted dipeptidase glycoprotein expressed predominantly in the liver and brain, and selectively secreted by brain cells into CSF, which catalyzes the hydrolysis of the dipeptides carnosine (β-alanyl-L-histidine) [80]. Increased levels of CSF carnosinase CN1 have been found in normal aging, while decreased CSF levels was found in AD [81]. *In vitro* experiments has shown that N-glycosylation is essential for appropriate secretion and enzyme activity [82]. Our results indicate that decreased N-glycosylation of carnosinase CN1 may contribute to the decreased level of CSF carnosinase CN1 and may act as a marker for AD. Another interesting candidate is voltage-dependent T-type calcium channel subunit alpha-1H (CaV3.2), which was detected glycosylated only in AD. Uniprot sequence analysis shows three potential N-glycosylation sites (Asn192, Asn271, Asn1466) exist, and Asn271 was detected in AD with 18 N-glycoforms. As a calcium channel, CaV3.2 is expressed throughout the nervous system and controls cellular excitability and synaptic transmission. Studies have shown that changes in the N-glycosylation pattern can affect ion channel function, leading to neurological disorders [83, 84].

These glycoprotein candidates with altered glycosylation discussed above were representative

examples that may potentially correlate to AD progression. The current exploratory glycosylation-based biomarker study focuses on in-depth mapping the representative and averaged glycosylation landscape of CSF proteins in healthy control and AD. Such a comparison will shed light upon the overall average, dominant glycosylation difference and similarities, and also some of the interesting glycoprotein candidates with specific glycosylation patter alterations in AD. To select the interesting targets for discussion, there were several considerations to minimize the effects of detection bias in complex sample. Strict criteria were set to pick out those interesting N-glycoprotein/ N-glycosite candidates, including at least 4 glycoforms changes for these shared N-glycoproteins/ N-glycosites and at least 3 glycoforms changes for these N-glycoproteins/ N-glycosites detected only in AD or control. Other factors are also considered including the increased/decreased percentage in terms of the absolute number of detected glycoforms, the reported association with AD or ND, and their roles in CNS biological processes. Still, such list is merely a preliminary exploration and further investigations are needed to narrow down or provide a more complete list of potential interesting glycosylation-based biomarker candidates in AD. Future studies includes conducting a high-throughput quantitative studies using 12-plex DiLeu isobaric tags developed in our lab [85], which could allow us to take the individual patient-to-patient variation into account and validate the results described here.

Acknowledgements

This research was supported in part by the National Institutes of Health (NIH) grants R21AG055377, R01 DK071801, and R56 MH110215. The Orbitrap instruments were purchased

through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges a Vilas Distinguished Achievement Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. We thank Dr. Marshall Bern from Protein Metrics for providing access to Byonic software package.

References:

1. Segal, M., Extracellular and cerebrospinal fluids. *J. Inherited Metab. Dis.* **1993,** 16, (4), 617-638.
2. McComb, J. G., Recent research into the nature of cerebrospinal fluid formation and absorption. *J. Neurosurg.* **1983,** 59, (3), 369-383.
3. Regeniter, A.; Kuhle, J.; Mehling, M.; Möller, H.; Wurster, U.; Freidank, H.; Siede, W. H., A modern approach to CSF analysis: pathophysiology, clinical application, proof of concept and laboratory reporting. *Clin. Neurol. Neurosurg.* **2009,** 111, (4), 313-318.
4. Abdi, F.; Quinn, J. F.; Jankovic, J.; McIntosh, M.; Leverenz, J. B.; Peskind, E.; Nixon, R.; Nutt, J.; Chung, K.; Zabetian, C., Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. *J. Alzheimer's Dis.* **2006,** 9, (3), 293-348.
5. Zhang, J., Proteomics of human cerebrospinal fluid–the good, the bad, and the ugly. *PROTEOMICS-Clinical Applications* **2007,** 1, (8), 805-819.
6. Fonteh, A. N.; Harrington, R. J.; Huhmer, A. F.; Biringer, R. G.; Riggins, J. N.; Harrington, M. G., Identification of disease markers in human cerebrospinal fluid using lipidomic and proteomic methods. *Dis. Markers* **2006,** 22, (1-2), 39-64.
7. Guldbrandsen, A.; Vethe, H.; Farag, Y.; Oveland, E.; Garberg, H.; Berle, M.; Myhr, K.-M.; Opsahl, J. A.; Barsnes, H.; Berven, F. S., In-depth characterization of the cerebrospinal fluid (CSF) proteome displayed through the CSF proteome resource (CSF-PR). *Mol. Cell. Proteomics* **2014,** 13, (11), 3152-3163.
8. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.
9. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nature reviews. Cancer* **2015,** 15, (9), 540.
10. Hwang, H.; Zhang, J.; Chung, K. A.; Leverenz, J. B.; Zabetian, C. P.; Peskind, E. R.; Jankovic, J.; Su, Z.; Hancock, A. M.; Pan, C., Glycoproteomics in neurodegenerative diseases. *Mass Spectrom. Rev.* **2010,** 29, (1), 79-125.
11. Schedin‐Weiss, S.; Winblad, B.; Tjernberg, L. O., The role of protein glycosylation in Alzheimer disease. *The FEBS journal* **2014,** 281, (1), 46-62.
12. Li, X.; Lu, F.; Wang, J. Z.; Gong, C. X., Concurrent alterations of O‐GlcNAcylation and phosphorylation of tau in mouse brains during fasting. *Eur. J. Neurosci.* **2006,** 23, (8), 2078-2086.
13. Liu, F.; Iqbal, K.; Grundke-Iqbal, I.; Hart, G. W.; Gong, C.-X., O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, (29), 10804-10809.
14. Yuzwa, S. A.; Shan, X.; Macauley, M. S.; Clark, T.; Skorobogatko, Y.; Vosseller, K.; Vocadlo, D. J., Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **2012,** 8, (4), 393-399.
15. Palmigiano, A.; Barone, R.; Sturiale, L.; Sanfilippo, C.; Bua, R. O.; Romeo, D. A.; Messina, A.; Capuana, M. L.; Maci, T.; Le Pira, F., CSF N-glycoproteomics for early diagnosis in

Alzheimer's disease. *J. Proteomics* **2016,** 131, 29-37.
16. Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G., Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **2009,** 6, (11), 809-811.
17. Goyallon, A.; Cholet, S.; Chapelle, M.; Junot, C.; Fenaille, F., Evaluation of a combined glycomics and glycoproteomics approach for studying the major glycoproteins present in biofluids: Application to cerebrospinal fluid. *Rapid Commun. Mass Spectrom.* **2015,** 29, (6), 461-473.
18. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** 6, (5), 359.
19. Cunningham, R.; Wang, J.; Wellner, D.; Li, L., Investigation and reduction of sub‐microgram peptide loss using molecular weight cut‐off fractionation prior to mass spectrometric analysis. *J. Mass Spectrom.* **2012,** 47, (10), 1327-1332.
20. Wang, J.; Cunningham, R.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L., Label‐free quantitative comparison of cerebrospinal fluid glycoproteins and endogenous peptides in subjects with Alzheimer's disease, mild cognitive impairment, and healthy individuals. *PROTEOMICS-Clinical Applications* **2016,** 10, (12), 1225-1241.
21. Zhang, C.; Ye, Z.; Xue, P.; Shu, Q.; Zhou, Y.; Ji, Y.; Fu, Y.; Wang, J.; Yang, F., Evaluation of different N-glycopeptide enrichment methods for N-glycosylation sites mapping in mouse brain. *Journal of proteome research* **2016,** 15, (9), 2960-2968.
22. Pan, Y.; Bai, H.; Ma, C.; Deng, Y.; Qin, W.; Qian, X., Brush polymer modified and lectin immobilized core–shell microparticle for highly efficient glycoprotein/glycopeptide enrichment. *Talanta* **2013,** 115, 842-848.
23. Zielinska, D. F.; Gnad, F.; Wiśniewski, J. R.; Mann, M., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010,** 141, (5), 897-907.
24. Deeb, S. J.; Cox, J.; Schmidt-Supprian, M.; Mann, M., N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Molecular & cellular proteomics* **2014,** 13, (1), 240-251.
25. Zielinska, D. F.; Gnad, F.; Schropp, K.; Wiśniewski, J. R.; Mann, M., Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Molecular cell* **2012,** 46, (4), 542-548.
26. Wang, X.; Yuan, Z.-F.; Fan, J.; Karch, K. R.; Ball, L. E.; Denu, J. M.; Garcia, B. A., A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol. Cell. Proteomics* **2016,** 15, (7), 2462-2475.
27. Lee, L. Y.; Moh, E. S.; Parker, B. L.; Bern, M.; Packer, N. H.; Thaysen-Andersen, M., Toward automated N-glycopeptide identification in glycoproteomics. *J. Proteome Res.* **2016,** 15, (10), 3904-3915.
28. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1999,** 1473, (1), 4-8.
29. Zhang, Y.; Zhang, C.; Jiang, H.; Yang, P.; Lu, H., Fishing the PTM proteome with chemical approaches using functional solid phases. *Chem. Soc. Rev.* **2015,** 44, (22), 8260-8287.

30. Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R., Glycopeptide enrichment and separation for protein glycosylation analysis. *J. Sep. Sci.* **2012,** 35, (18), 2341-2372.
31. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **2014,** 13, (12), 6176-6186.
32. Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M., Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Analytical Chemistry* **2010,** 82, (13), 5598-5609.
33. Sparbier, K.; Koch, S.; Kessler, I.; Wenzel, T.; Kostrzewa, M., Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. *Journal of biomolecular techniques: JBT* **2005,** 16, (4), 407.
34. Thaysen-Andersen, M.; Larsen, M. R.; Packer, N. H.; Palmisano, G., Structural analysis of glycoprotein sialylation–part I: pre-LC-MS analytical strategies. *Rsc Advances* **2013,** 3, (45), 22683-22705.
35. Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. M.; Mechtler, K.; Heck, A. J.; Mohammed, S., Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **2013,** 12, (3), 1520-1525.
36. Frese, C. K.; Altelaar, A. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012,** 84, (22), 9668-9673.
37. Marino, F.; Bern, M.; Mommen, G. P.; Leney, A. C.; van Gaans-van den Brink, J. A.; Bonvin, A. M.; Becker, C.; van Els, C. c. A.; Heck, A. J., Extended O-GlcNAc on HLA class-I-bound peptides. *J. Am. Chem. Soc.* **2015,** 137, (34), 10922-10925.
38. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.-W.; Kent, K. C.; Li, L., Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass Spectrom.* **2017,** 28, (9), 1751-1764.
39. Dennis, G.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A., DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **2003,** 4, (9), R60.
40. Roth, J., Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **2002,** 102, (2), 285-304.
41. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M., UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2004,** 32, (suppl_1), D115-D119.
42. Farriol‐Mathis, N.; Garavelli, J. S.; Boeckmann, B.; Duvaud, S.; Gasteiger, E.; Gateau, A.; Veuthey, A. L.; Bairoch, A., Annotation of post‐translational modifications in the Swiss‐Prot knowledge base. *Proteomics* **2004,** 4, (6), 1537-1550.
43. Kobayashi, K.; Ogata, H.; Morikawa, M.; Iijima, S.; Harada, N.; Yoshida, T.; Brown, W.; Inoue, N.; Hamada, Y.; Ishii, H., Distribution and partial characterisation of IgG Fc binding protein

in various mucin producing cells and body fluids. *Gut* **2002,** 51, (2), 169-176.
44. Parry, S.; Hanisch, F. G.; Leir, S.-H.; Sutton-Smith, M.; Morris, H. R.; Dell, A.; Harris, A., N-Glycosylation of the MUC1 mucin in epithelial cells and secretions. *Glycobiology* **2006,** 16, (7), 623-634.
45. Taniguchi, T.; Woodward, A. M.; Magnelli, P.; McColgan, N. M.; Lehoux, S.; Jacobo, S. M. P.; Mauris, J.; Argüeso, P., N-glycosylation Affects the Stability and Barrier Function of the MUC16 Mucin. *J. Biol. Chem.* **2017**, jbc. M116. 770123.
46. Chugh, S.; Gnanapragassam, V. S.; Jain, M.; Rachagani, S.; Ponnusamy, M. P.; Batra, S. K., Pathobiological implications of mucin glycans in cancer: Sweet poison and novel targets. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **2015,** 1856, (2), 211-225.
47. Fournier, T.; Medjoubi-N, N.; Porquet, D., Alpha-1-acid glycoprotein. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2000,** 1482, (1), 157-171.
48. Dente, L.; Rüther, U.; Tripodi, M.; Wagner, E. F.; Cortese, R., Expression of human alpha 1-acid glycoprotein genes in cultured cells and in transgenic mice. *Genes Dev.* **1988,** 2, (2), 259-266.
49. Imre, T.; Kremmer, T.; Heberger, K.; Molnár-Szöllősi, É.; Ludanyi, K.; Pocsfalvi, G.; Malorni, A.; Drahos, L.; Vekey, K., Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *J. Proteomics* **2008,** 71, (2), 186-197.
50. Yamaguchi, H.; Uchida, M., A chaperone-like function of intramolecular high-mannose chains in the oxidative refolding of bovine pancreatic RNase B. *The Journal of Biochemistry* **1996,** 120, (3), 474-477.
51. Morishima, S.; Morita, I.; Tokushima, T.; Kawashima, H.; Miyasaka, M.; Omura, K.; Murota, S., Expression and role of mannose receptor/terminal high-mannose type oligosaccharide on osteoclast precursors during osteoclast formation. *J. Endocrinol.* **2003,** 176, (2), 285-292.
52. Ajit, V.; Richard, C.; Jeffrey, E.; Hudson, F.; Gerald, H.; Jamey, M., Essentials of glycobiology. *Cold Spring Harbor Laboratory Press, New York* **2009**.
53. Sarkar, M.; Leventis, P. A.; Silvescu, C. I.; Reinhold, V. N.; Schachter, H.; Boulianne, G. L., Null mutations in Drosophila N-acetylglucosaminyltransferase I produce defects in locomotion and a reduced life span. *J. Biol. Chem.* **2006,** 281, (18), 12776-12785.
54. Schachter, H., Paucimannose N-glycans in Caenorhabditis elegans and Drosophila melanogaster. *Carbohydr. Res.* **2009,** 344, (12), 1391-1396.
55. Yagi, H.; Saito, T.; Yanagisawa, M.; Robert, K. Y.; Kato, K., Lewis X-carrying N-glycans regulate the proliferation of mouse embryonic neural stem cells via the Notch signaling pathway. *J. Biol. Chem.* **2012,** 287, (29), 24356-24364.
56. Joosten, C. E.; Cohen, L. S.; Ritter, G.; Batt, C. A.; Shuler, M. L., Glycosylation profiles of the human colorectal cancer A33 antigen naturally expressed in the human colorectal cancer cell line SW1222 and expressed as recombinant protein in different insect cell lines. *Biotechnol. Prog.* **2004,** 20, (4), 1273-1279.
57. Sethi, M. K.; Thaysen-Andersen, M.; Smith, J. T.; Baker, M. S.; Packer, N. H.; Hancock, W. S.; Fanayan, S., Comparative N-glycan profiling of colorectal cancer cell lines reveals unique bisecting GlcNAc and α-2, 3-linked sialic acid determinants are associated with membrane

proteins of the more metastatic/aggressive cell lines. *J. Proteome Res.* **2013,** 13, (1), 277-288.
58. Balog, C. I.; Stavenhagen, K.; Fung, W. L.; Koeleman, C. A.; McDonnell, L. A.; Verhoeven, A.; Mesker, W. E.; Tollenaar, R. A.; Deelder, A. M.; Wuhrer, M., N-glycosylation of colorectal cancer tissues a liquid chromatography and mass spectrometry-based investigation. *Mol. Cell. Proteomics* **2012,** 11, (9), 571-585.
59. Everest-Dass, A. V.; Jin, D.; Thaysen-Andersen, M.; Nevalainen, H.; Kolarich, D.; Packer, N. H., Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by Candida albicans. *Glycobiology* **2012,** 22, (11), 1465-1479.
60. Thaysen-Andersen, M.; Venkatakrishnan, V.; Loke, I.; Laurini, C.; Diestel, S.; Parker, B. L.; Packer, N. H., Human neutrophils secrete bioactive paucimannosidic proteins from azurophilic granules into pathogen-infected sputum. *J. Biol. Chem.* **2015,** 290, (14), 8789-8802.
61. 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* **2016,** 15, (6), 1773-1790.
62. Diz, A. P.; Truebano, M.; Skibinski, D. O., The consequences of sample pooling in proteomics: an empirical study. *Electrophoresis* **2009,** 30, (17), 2967-2975.
63. Kalsheker, N. A., α1-antichymotrypsin. *The international journal of biochemistry & cell biology* **1996,** 28, (9), 961-964.
64. Licastro, F.; Parnetti, L.; Morini, M. C.; Davis, L. J.; Cucinotta, D.; Gaiti, A.; Senin, U., Acute Phase Reactant [alpha] 1Antichymotrypsin Is Increased in Cerebrospinal Fluid and Serum of Patients with Probable Alzheimer Disease. *Alzheimer Dis. Assoc. Disord.* **1995,** 9, (2), 112-118.
65. Ma, J.; Yee, A.; Brewer, H. B.; Das, S.; Potter, H., Amyloid-associated proteins α1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer β-protein into filaments. *Nature* **1994,** 372, (6501), 92-94.
66. Ma, J.; Brewer, H. B.; Potter, H., Alzheimer Aβ neurotoxicity: promotion by antichymotrypsin, apoE4; inhibition by Aβ-related peptides. *Neurobiol. Aging* **1996,** 17, (5), 773-780.
67. Nilsson, L. N.; Bales, K. R.; DiCarlo, G.; Gordon, M. N.; Morgan, D.; Paul, S. M.; Potter, H., α-1-Antichymotrypsin promotes β-sheet amyloid plaque deposition in a transgenic mouse model of Alzheimer's disease. *J. Neurosci.* **2001,** 21, (5), 1444-1451.
68. Mucke, L.; Yu, G.-Q.; McConlogue, L.; Rockenstein, E. M.; Abraham, C. R.; Masliah, E., Astroglial expression of human α 1-antichymotrypsin enhances Alzheimer-like pathology in amyloid protein precursor transgenic mice. *The American journal of pathology* **2000,** 157, (6), 2003-2010.
69. Padmanabhan, J.; Levy, M.; Dickson, D. W.; Potter, H., Alpha1-antichymotrypsin, an inflammatory protein overexpressed in Alzheimer's disease brain, induces tau phosphorylation in neurons. *Brain* **2006,** 129, (11), 3020-3034.
70. LAINE, A.; DAVRIL, M.; RABAUD, M.; VERCAIGNE‐MARKO, D.; HAYEM, A., Human serum α 1‐antichymotrypsin is an inhibitor of pancreatic elastases. *The FEBS Journal* **1985,** 151, (2), 327-331.
71. Hachulla, E.; Laine, A.; Hayem, A., Alpha 1-antichymotrypsin microheterogeneity in crossed immunoaffinoelectrophoresis with free concanavalin A: a useful diagnostic tool in inflammatory

syndrome. *Clin. Chem.* **1988,** 34, (5), 911-915.
72. LAINE, A.; HACHULLA, E.; STRECKER, G.; MICHALSKI, J. C.; WIERUSZESKI, J. M., Structure determination of the glycans of human‐serum α1‐antichymotrypsin using 1H‐NMR spectroscopy and deglycosylation by N‐glycanase. *The FEBS Journal* **1991,** 197, (1), 209-215.
73. Kania, A.; Klein, R., Mechanisms of ephrin-Eph signalling in development, physiology and disease. *Nature reviews Molecular cell biology* **2016,** 17, (4), 240-257.
74. Ianni, M.; Manerba, M.; Di Stefano, G.; Porcellini, E.; Chiappelli, M.; Carbone, I.; Licastro, F., Altered glycosylation profile of purified plasma ACT from Alzheimer's disease. *Immun. Ageing* **2010,** 7, (1), S6.
75. Arvanitis, D.; Davy, A., Eph/ephrin signaling: networks. *Genes Dev.* **2008,** 22, (4), 416-429.
76. Chen, Y.; Fu, A. K.; Ip, N. Y., Eph receptors at synapses: implications in neurodegenerative diseases. *Cell. Signal.* **2012,** 24, (3), 606-611.
77. Barquilla, A.; Pasquale, E. B., Eph receptors and ephrins: therapeutic opportunities. *Annu. Rev. Pharmacol. Toxicol.* **2015,** 55, 465-487.
78. Ferluga, S.; Hantgan, R.; Goldgur, Y.; Himanen, J. P.; Nikolov, D. B.; Debinski, W., Biological and structural characterization of glycosylation on ephrin-A1, a preferred ligand for EphA2 receptor tyrosine kinase. *J. Biol. Chem.* **2013,** 288, (25), 18448-18457.
79. Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D., Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *Journal of proteome research* **2005,** 4, (6), 2070-2080.
80. Teufel, M.; Saudek, V.; Ledig, J.-P.; Bernhardt, A.; Boularand, S.; Carreau, A.; Cairns, N. J.; Carter, C.; Cowley, D. J.; Duverger, D., Sequence identification and characterization of human carnosinase and a closely related non-specific dipeptidase. *J. Biol. Chem.* **2003,** 278, (8), 6521-6531.
81. Barrack, S., *Biochemical biomarkers in Alzheimer's disease*. iMedPub: 2013.
82. Riedl, E.; Koeppel, H.; Pfister, F.; Peters, V.; Sauerhoefer, S.; Sternik, P.; Brinkkoetter, P.; Zentgraf, H.; Navis, G.; Henning, R. H., N-Glycosylation of carnosinase influences protein secretion and enzyme activity. *Diabetes* **2010,** 59, (8), 1984-1990.
83. Baycin-Hizal, D.; Gottschalk, A.; Jacobson, E.; Mai, S.; Wolozny, D.; Zhang, H.; Krag, S. S.; Betenbaugh, M. J., Physiologic and pathophysiologic consequences of altered sialylation and glycosylation on ion channel function. *Biochem. Biophys. Res. Commun.* **2014,** 453, (2), 243-253.
84. Ondacova, K.; Karmazinova, M.; Lazniewska, J.; Weiss, N.; Lacinova, L., Modulation of Cav3. 2 T-type calcium channel permeability by asparagine-linked glycosylation. *Channels* **2016,** 10, (3), 175-184.
85. Frost, D. C.; Greer, T.; Li, L., High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Analytical chemistry* **2014,** 87, (3), 1646-1654.

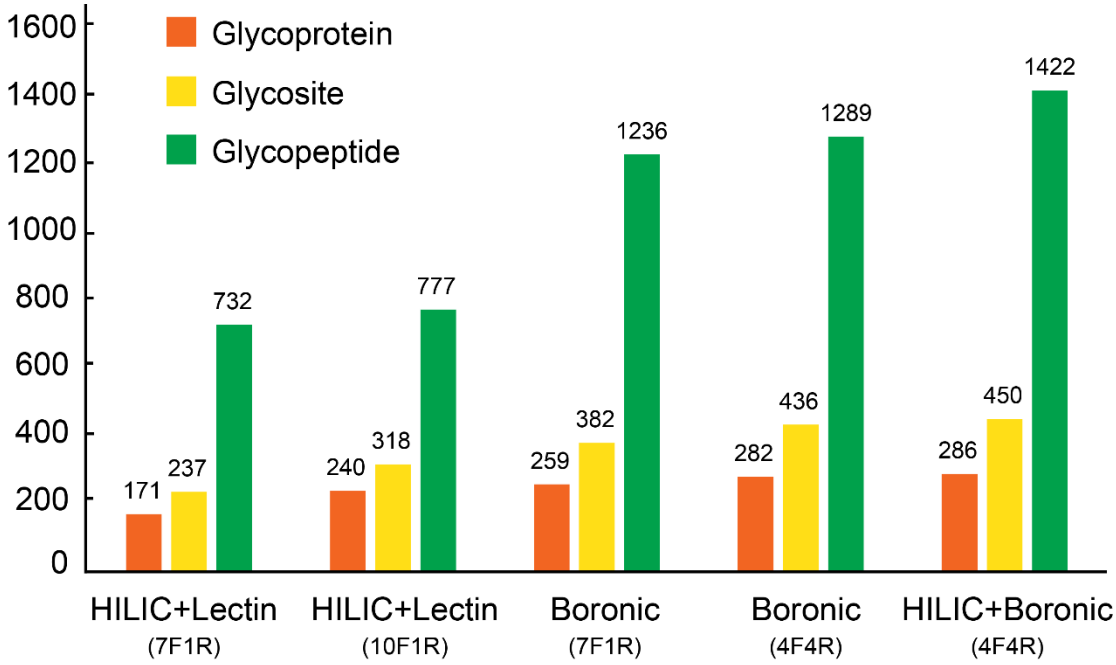


Figure 1. N-glycopeptide enrichment strategy optimization. Different enrichment strategies including either single enrichment or sequential enrichment were employed to enrich N-glycopeptides from tryptic peptides derived from 200 μg proteins from PANC1 cells. The number of N-glycoroteins, N-glycopeptides and N-glycosylation sites of each strategy were compared. The combination of HILIC and boronic acid enrichment gives the highest coverage. (F: fractions, R: technical replicates)

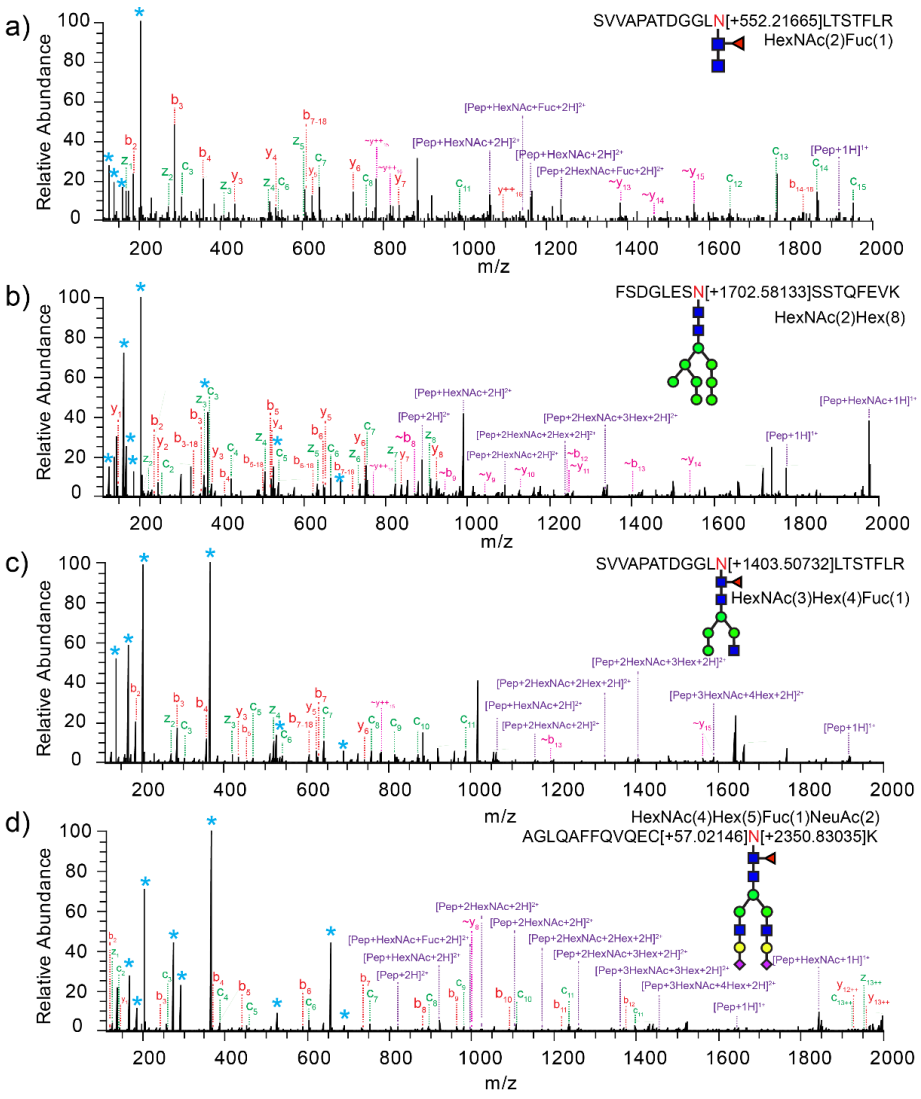


Figure 2. Representative EThcD spectra of intact N-glycopeptides with different categories of N-glycan attached were shown, including truncated N-glycan (a), high-mannose N-glycan (b), hybrid N-glycan (c), complex N-glycan (d). Peptide fragment ions are denoted as b/y, c/z ions using the most commonly used nomenclature, and ~b/y ions denote the loss of labile glycan. (Asterisk: glycan oxonium ion and glycan fragment ions, "pep" : peptide sequence, HexNAc: acetylglucoseamine, Hex: hexose, Fuc: fucose, NeuAc: sialic: N-acetylneuraminic acid)

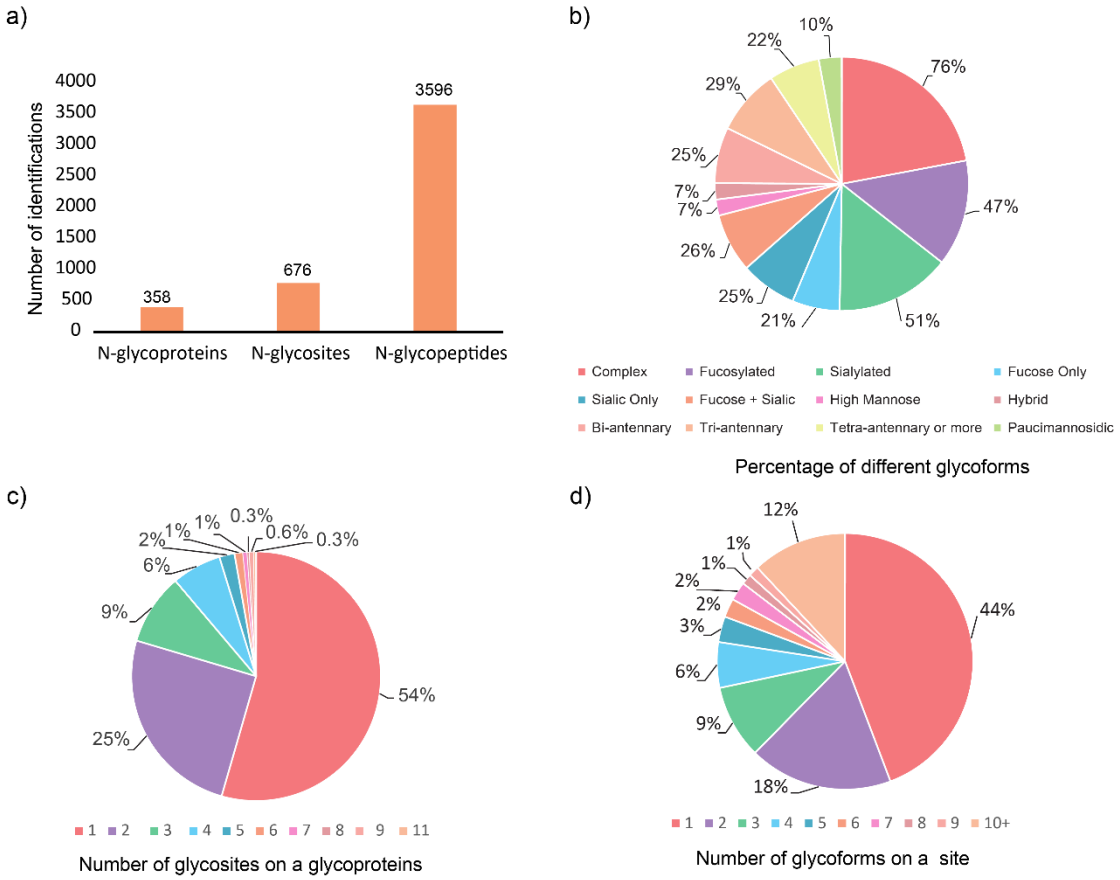


Figure 3. Site-specific N-glycoproteome profiling in CSF samples from healthy control. a) The number of N-glycoproteins, N-glycopeptides and N-glycosylation sites identified. b) The percentage of different N-glycoforms in CSF. c) N-glycosylation sites distribution among glycoproteins. d) N-glycoforms distribution among glycosylation sites.

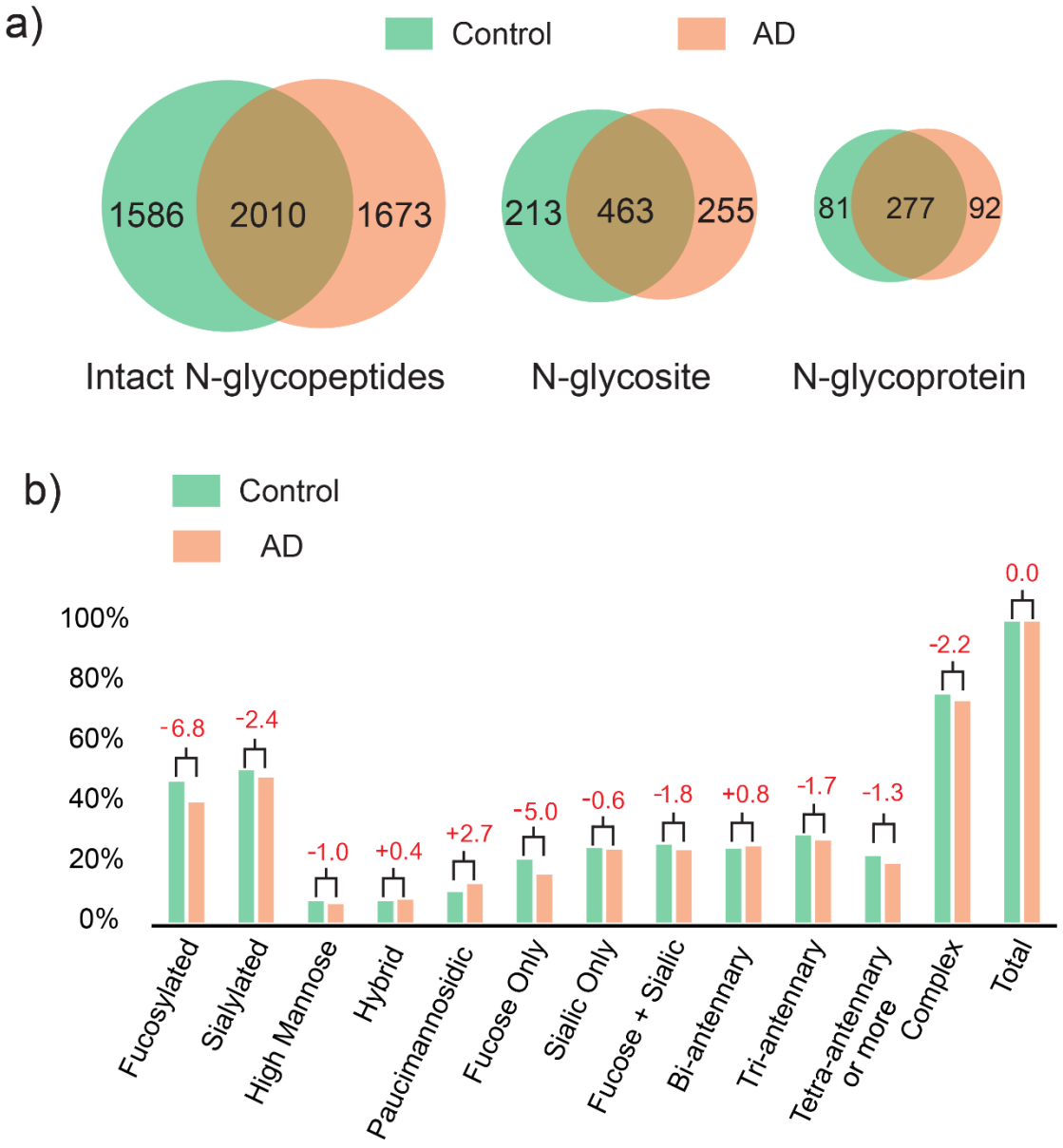


Figure 4. Comparison of CSF N-glycoproteome between healthy control and AD. a) Venn diagram analysis N-glycoproteins, N-glycopeptides and N-glycosylation sites identified between healthy control and AD. b) Comparison of different categories of glycoforms between healthy control and AD.

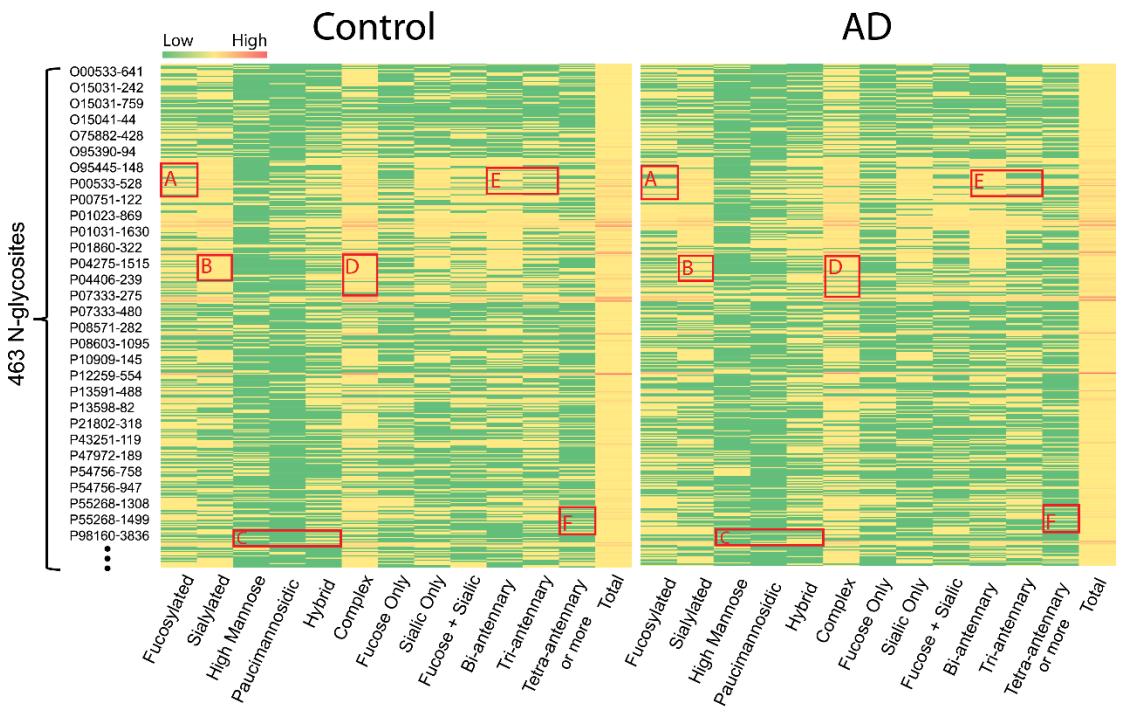


Figure 5. Comparative heat map analysis of N-glycoproteome between healthy control and AD. The horizontal axis represents 13 categories of glycoforms, and vertical axis represents specific N-glycosylation site (Portein Uniprot ID + glycosylation site position). The number of a certain category of glycoform on each site is color coded. Different marked region (A-F) represents glycosylation alteration of different categories glycoforms (A: Fucosylation, B: Sialylation, C: High-mannose, paucimannosidic, hybrid, D: Complex, E: Bi-antennary, Tri-antennary, F: Tetra-antennary). The number of glycoforms is color-coded. Green represents a low number and red represents a high number.

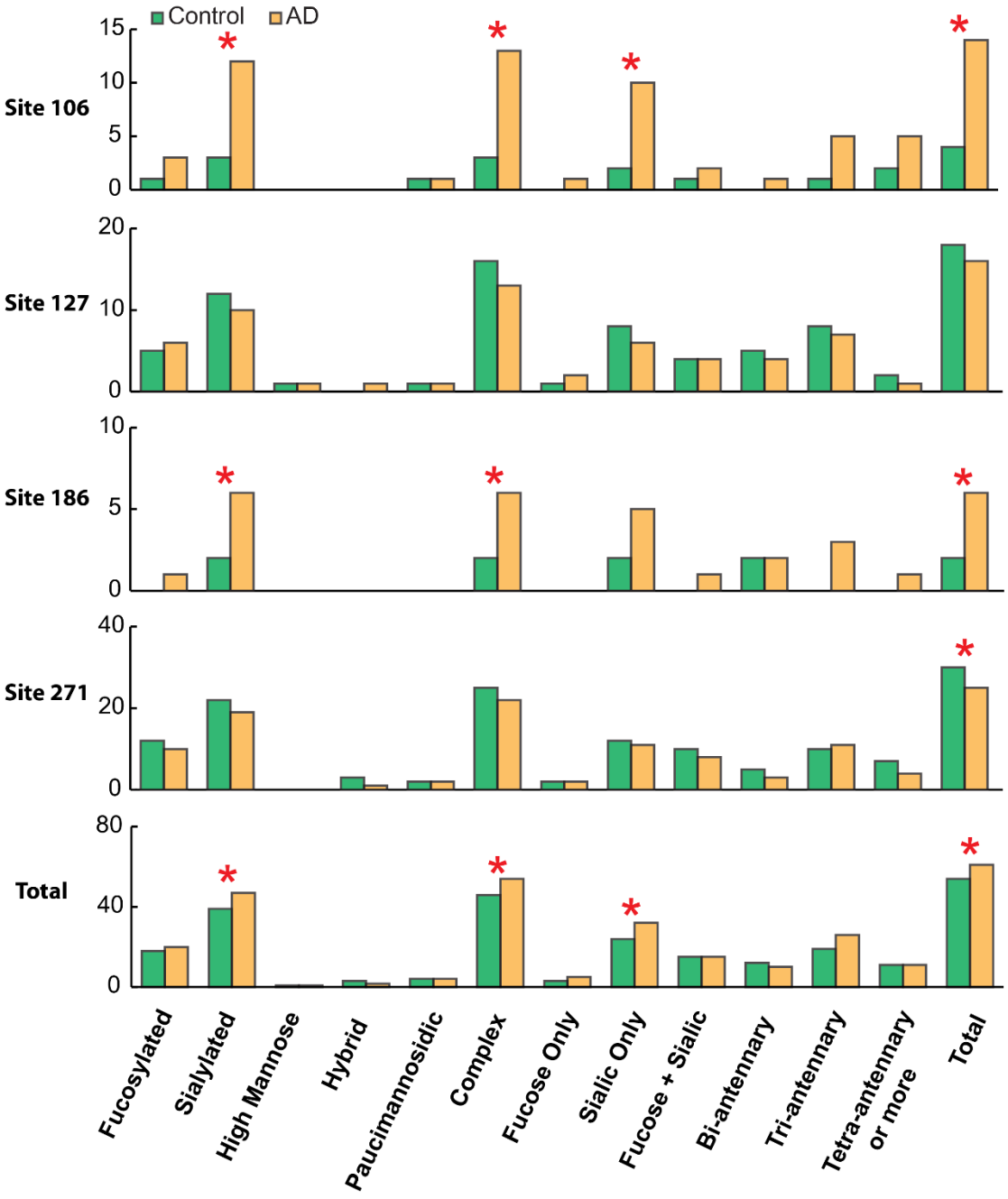


Figure 6. Glycosylation pattern changes for alpha-1-antichymotrypsin (ACT) in AD. The changes of the number of N-glycoforms for each site and the total glycoforms were shown. (Asterisk denotes changes ≥ 4)

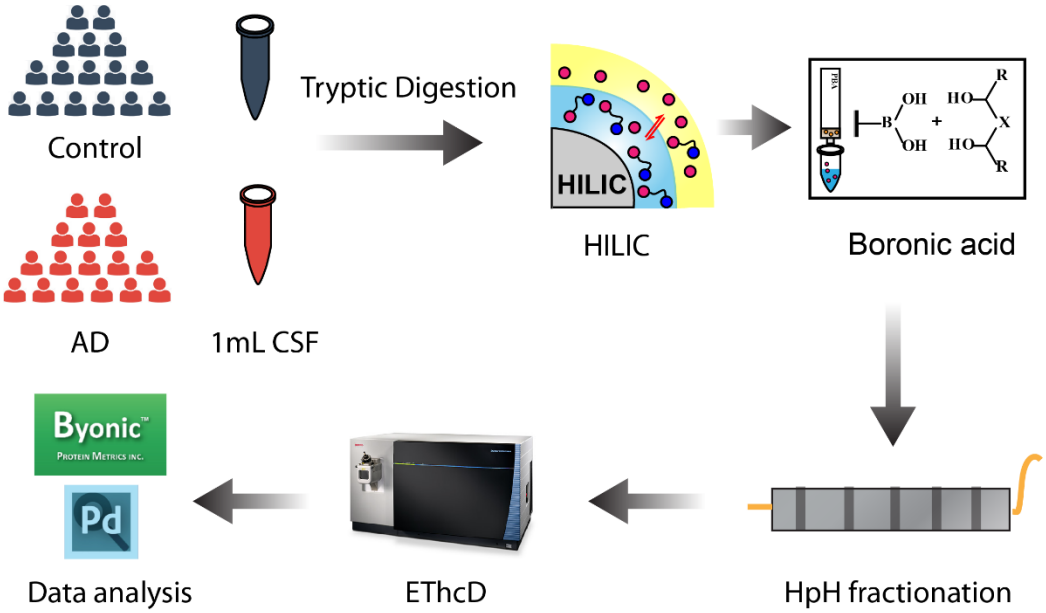


Figure S1. Overall experiment design processing CSF samples from healthy control and AD patients.

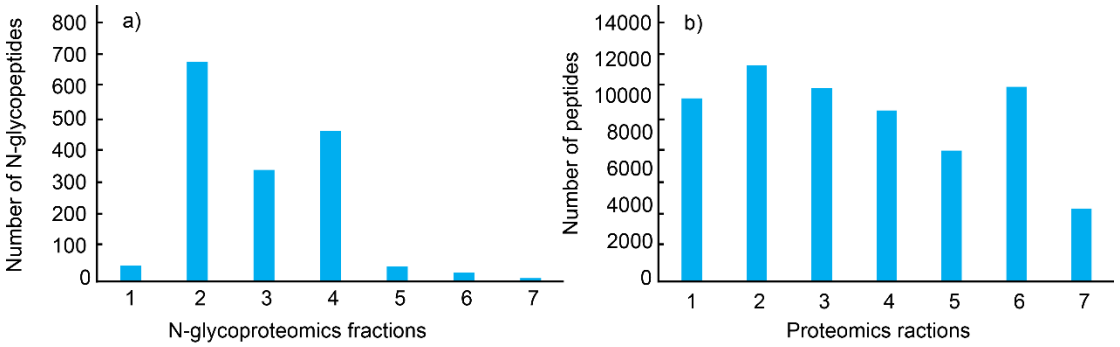


Figure S2. High-pH-fractionation comparison between glycoproteomics a) and proteomics b).

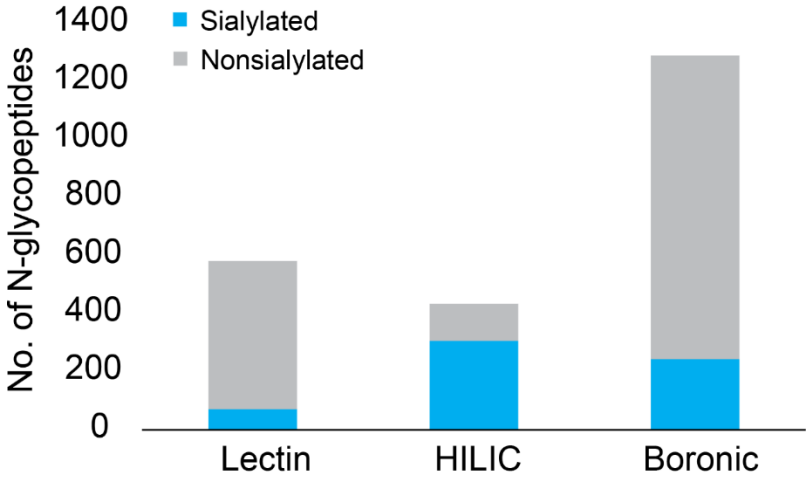


Figure S3. The enriched sialyalted and nonsialylated glycopeptides percentage comparison by lectin, HILIC and boronic acid methods.

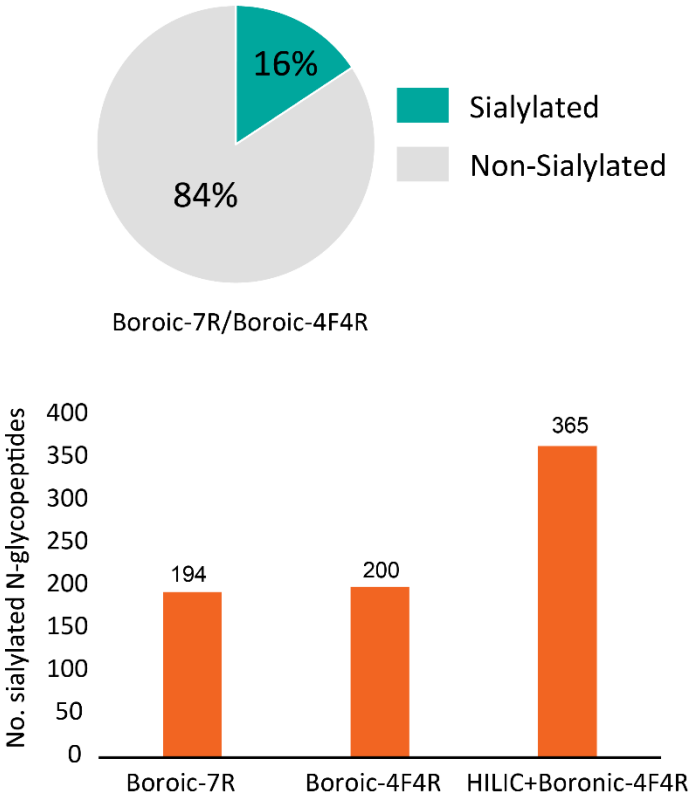


Figure S4. The increase of sialylated glycopeptides after combining HILIC with boronic acid enrichment from 200 μg PANC1 cells proteins. (F: fractions, R: technical replicates)

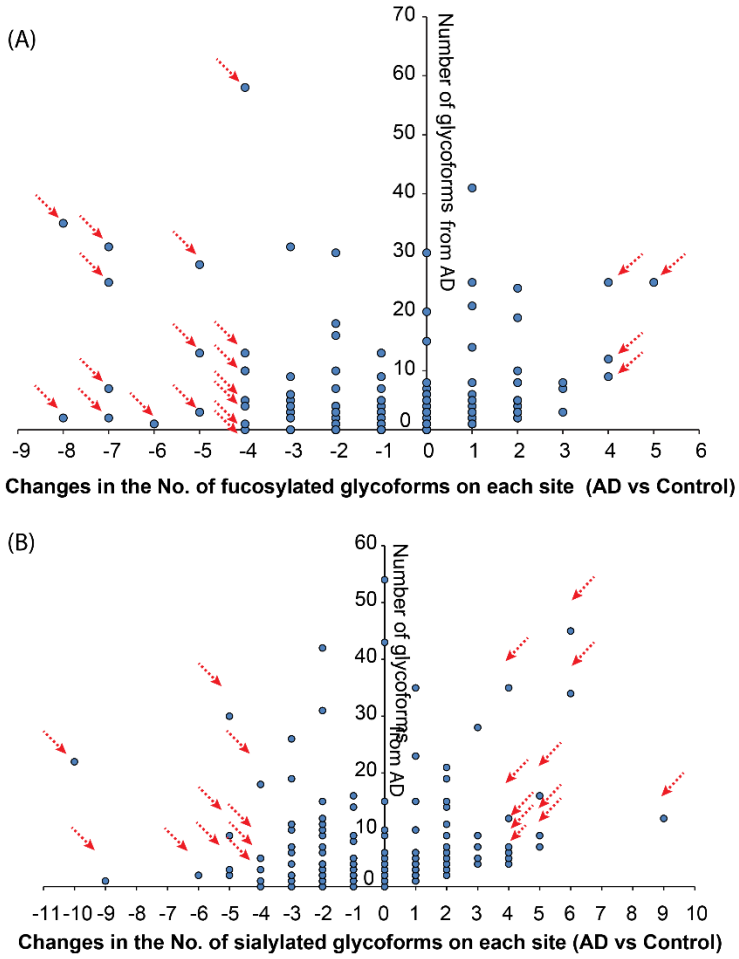


Figure S5. A two-dimensional plot depicting the changes in the number of fucosylated (A) and sialylated (B) glycoforms on each site (AD vs. Control) as a function of the number of glycoforms from AD.

Table S1. Subjects information of 16 healthy control and 16 AD patients.

Groups	No. of Subjects	Male, No. (%)	Female, No. (%)	Age, mean (SD)	APOE4 positive, No. (%)
Normal Control	16	8 (50%)	8 (50%)	73.0 (5.3)	3 (19%)
AD	16	8 (50%)	8 (50%)	73.2 (8.6)	10 (63%)

Chapter 8

In-depth site-specific O-glycosylation analysis of glycoproteins and endogenous peptides in cerebrospinal fluid (CSF) from healthy individual, mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients

Adapted from **Chen, Z.**, Yu, Q., Johnson, J., Shipman, R., Zhong, X., Huang, J., Zetterberg, H., Asthana, S., Carlsson, C., Okonkwo, O., Li, L. *To be submitted to J. Proteome Res. (2018)*

Author contribution: study was designed by Chen, Z., Zetterberg, H., Asthana, S., Carlsson, C., Okonkwo, O., Li, L; experiment was performed by Chen, Z., Yu, Q., Johnson, J., Shipman, R.,; data was analyzed by Chen, Z., Johnson, J., Shipman, R., Zhong, X., Huang, J.,; manuscript was written by Chen, Z. and edited by Yu Q., Zetterberg, H., Okonkwo, O. and L. Li.

Abstract

Site-specific O-glycoproteome mapping in complex biological system provides molecular basis for understanding the structure-function relationships of glycoproteins and their roles in physiological and pathological processes. Previous O-glycoproteome analysis in cerebrospinal fluid (CSF) focused on sialylated glycoforms, and many CSF glycoproteins have not been characterized comprehensively with respect to their O-glycosylation. In order to provide an unbiased O-glycosylation profiling, we have developed an integrated strategy combining universal boronic acid enrichment, high-pH fractionation, and electron-transfer and higher-energy collision dissociation (EThcD) for improved intact O-glycopeptide analysis. This strategy was applied to analyze O-glycoproteome in CSF, leading to identifications of 308 O-glycopeptides from 110 O-glycoproteins, covering both sialylated and non-sialylated glycoforms. To our knowledge, this is the largest number of O-glycoproteins and O-glycosites reported for CSF so far, including 154 novel O-glycosites. Due to a lack of peptidomics workflow that could incorporate glycosylation analysis, the glycosylation state of CSF endogenous peptides has not been comprehensively studied. Here, we developed a peptidomics workflow that utilizes the EThcD fragmentation and a three-step database searching strategy, allowing both N-glycosylation and O-glycosylation, as well as other common peptide PTMs, to be analyzed at the same time. Interestingly, among the 1492 endogenous peptides identified, 95 of them were O-glycosylated and only 1 N-glycosylated peptide was found, indicating CSF endogenous peptides were preferentially O-glycosylated. By referring to human neuropeptide database, 15 of them were actually O-glycosylated neuropeptides deriving from ProSAAS and secretogranin-1. O-glycoproteome and endogenous peptidome PTMs

analysis were also conducted in MCI and AD patients to give a landscape picture of glycosylation in these disease states. The results showed that a decreased fucosylation trend was found in MCI and AD, suggesting its potential relation to the progression of AD.

Introduction

Predominantly produced in choroid plexuses, human cerebrospinal fluid (CSF) circulates within the ventricles of the brain and surrounds the brain and spinal cord.[1] The averaged total volume of CSF is estimated to be about 150 ml in adults, and daily volume of CSF produced varies between 400 to 600 ml.[2] The functional role of CSF, though remaining to be defined precisely, includes mechanical protection of central nervous system (CNS), homeostasis of the interstitial fluid in the brain, and regulation of neuronal functioning.[3, 4] Being in direct contact with CNS, CSF reflects the ongoing physiological or pathological state of CNS most directly.[5, 6] CSF contains metabolites, peptides, proteins, enzymes, and hormones *etc.* that play important roles in many biological processes in CNS.[7] Changes in the compositions of CSF acts as a sign of pathological alterations in the CNS, and thus provides great opportunity for biomarker discovery in neurological diseases.

As one of the most complex forms of protein post-translational modifications (PTMs), glycosylation acts as a key regulatory mechanism controlling protein folding, molecular trafficking, cell adhesion, receptor activation and signal transduction.[8] According to the different amino acids glycans attached to, glycosylation could be classified into two major classes, N-glycosylation and O-glycosylation. Biosynthesis of N-glycosylation is initiated by transfer of a pre-assembled 14 monosaccharide complex glycan to asparagine residue (Asn) within the consensus motif (Asn-X-

Ser/Thr, X≠P) containing polypeptides followed by sequential addition or removal of certain monosaccharide in a well-defined process.[9] In contrast, O-glycosylation synthesis involves attachment of a single monosaccharide to the serine/threonine (Ser/Thr) residue of a polypeptide without any definable peptide consensus motif and latter transfer of many diverse monosaccharide residues.[10, 11] As a result, a higher extent of occupancy and a larger degree of structure complexity and diversity were found for O-glycosylation.

In general, O-glycosylation could be categorized into mucin and non-mucin types according to the monosaccharide residue directly linked to polypeptides. The attached monosaccharide residue is N-acetylgalactosamine (GalNAc/HexNAc) in mucin type, whereas the attached residue can be N-acetylglucosamine (GlcNAc/HexNAc), fucose (Fuc), galactose (Gal/Hex), mannose, glucose in the non-mucin type.[9] The clustered regions of O-GalNAcylation on mucins has been shown to provide protection from proteolysis as well as unique rheological properties.[12] Mucin type O-glycosylation on cell surface and secreted proteins has also been shown to modulate recognition, adhesion, communication between cells and also their surrounding environments.[13] As a nutrient- and stress-responsive modification, non-mucin type O-GlcNAcylation is extensively involved in the spatiotemporal regulation of diverse cellular processes, including transcription, epigenetic modifications and cell signaling dynamics.[14] Apart from modifying protein, O-glycosylation could also happen on the endogenous peptides, including neurotransmitters and hormones. In fact, it has been well recognized that after initial peptidase cleavages, endogenous peptides could undergo further post-translational modifications such as amidation, acetylation, phosphorylation, sulfation, and N-/O-glycosylation.[15, 16] As an example, it has been reported that

there was an extensive N-/O-glycosylation of gonadotropin, a glycoprotein polypeptide hormones.[17] Recently, our group also discovered that mouse insulin-1B and -2B chains, human insulin-B chain were O-glycosylated, and multiple O-glycoforms were identified.[18] These findings suggest that O-glycosylation may play a crucial role in the folding, trafficking, metabolic clearance and biological activity of neurotransmitters and hormones and it is import to characterize their glycosylation state.

The CSF proteome and endogenous peptidome has been extensively characterized, but there are only few reports on the O-glycosylation study. One study reported 39 O-glycopeptides from 22 CSF O-glycoproteins,[19] and another study identified 43 O-glycopeptides from 28 CSF O-glycoproteins.[20] Utilizing enrichment approach based on hydrazide chemistry after oxidizing sialic acid (NeuAc), both of the two studies suffered from the weakness that sialic acid information was lost and only a subset of O-glycoproteome (sialylated O-glycopeptides) were characterized. As a result, the current O-glycoproteome depth is far less than satisfactory, hindering the design of studies to explore disease-related O-glycosylation alterations. For endogenous peptidome study in CSF, one study identified 730 endogenous peptides, including 138 peptides with PTMs such as acetylation, amidation, phosphorylation, Gln to pyro-Glu conversion.[21] But none of the identified peptides is reported to be glycosylated. In another study with 563 endogenous peptides identified,[22] Zougman et al. noted the presence of glycan oxonium ions in some spectra, indicating the existence of glycosylated endogenous peptides in CSF. By lowering the HCD collision energy, labile glycan moiety could be partially preserved, and 28 O-glycopeptides were successfully identified. However, only two O-glycan compositions were found. Considering that the two O-glycan

modifications set in the search engine were found by merely incidental observation and the large diversity of O-glycans out there, it's highly possible that there are more O-glycosylated endogenous peptides in CSF, even N-glycosylation. Therefore, a systemic approach needs to be developed, which features a glycosylation-centered analytical strategy and searching strategy to take vast categories of N-/O-glycans into account.

In the present study, we first optimized the boronic acid-based enrichment strategy after PNGase F pretreatment to efficiently enrich both sialylated and non-sialylated O-glycopeptides in CSF. Combined with high-pH (HpH) fractionation and intact O-glycopeptide characterization enabled by EThcD, 987 intact O-glycopeptides were identified using 200 μg tryptic peptides from PANC1 cells. The optimized approach was then applied to CSF O-glycoproteome study, 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins were confidently identified. About 30% of the O-glycosites identified contained at least two O-glycoforms, revealing the microheterogeneity of O-glycosylation. Up until now, the dataset represents the largest site-specific O-glycoproteome reported for CSF, including 154 novel O-glycosites. For endogenous peptides analysis, we developed a peptidomics workflow that combines CSF endogenous peptides extraction by 10k molecular weight cut-off (MWCO), EThcD fragmentation, and a three-step database searching strategy for comprehensive PTMs analysis. In this workflow, the use of EThcD enables the preserving of labile PTMs including glycosylation and phosphorylation, facilitating the accurate site localization. The three-step database searching strategy using the Byonic as the search engine allows a comprehensive PTMs analysis of endogenous peptides, including both N-/O-glycosylation, phosphorylation, amidation, acetylation, Gln to pyro-Glu conversion. To our

knowledge, the peptidomics workflow developed here is the first workflow that enables a thorough and systemic analysis of N-/O-glycosylation state of endogenous peptides. In total, 1492 enodgenous peptides were identified, and 370 of them underwent post-translational modifications including O-glycosylation, phosphorylation, acetylation, amidation, Gln to pyro-Glu conversion. Among them, 95 endogenous peptides were O-glycosylated, and 15 of them were O-glycosylated neuropeptides. Interestingly, only 1 N-glycosylated peptide was found, indicating endogenous peptides were preferentially O-glycosylated. To provide a landscape picture of the O-glycosylation state of CSF glycoproteins in MCI and AD patients, in-depth O-glycoprofiling experiments were also conducted in parallel, showing a decreased fucosylation in MCI and AD. The PTMs analysis of endogenous CSF peptides in MCI and AD patients showed an increased percentage of PTMs modified peptides, including O-glycosylation, Gln->pyro-Glu and acetylation. Furthermore, the developed strategy here is readily applicable for site-specific O-glycosylation analysis of glycoproteins or endogenous peptides in other complex biological systems.

Experimental Procedures

Chemicals and materials

Dithiothreitol (DTT), PNGase F, sequencing grade trypsin were from Promega (Madison, WI). Optimal LC/MS grade acetonitrile (ACN), methanol (MeOH) and water were from Fisher Scientific (Pittsburgh, PA). Tris base, urea (UA), sodium chloride and ammonium bicarbonate (ABC) were obtained from Fisher Scientific (Pittsburgh, PA). Formic acid (FA), 10% Sodium dodecyl sulfate solution (SDS), trifluoroacetic acid (TFA), dimethyl sulfoxide (DMSO) were

purchased from Sigma-Aldrich (St. Louis, MO). C18 OMIX tips and Phenylboronic acid (PBA) solid phase extraction cartridges were obtained from Agilent (Santa Clara, CA). Microcon filters YM-30 (30 kDa) and amicon Ultra-0.5 mL centrifugal filters (10 kDa) were purchased from Merck Millipore (Billerica, MA). PANC-1 pancreatic ductal adenocarcinoma cells were from ATCC (Manassas, VA).

CSF samples

48 enrollees in the Wisconsin Alzheimer's Disease Research Center (ADRC) participated in this study. The subjects comprised of 16 cognitively normal individuals who enrolled in the Wisconsin ADRC at late middle age, 16 individuals with MCI and 16 individuals with AD dementia. Detailed subjects information can be found in **Supplemental Table S1**. All MCI and AD participants were diagnosed via applicable clinical criteria in standardized and multidisciplinary consensus conferences. [23, 24] And cognitive normalcy was determined based on intact cognitive performance by a comprehensive battery of neuropsychological tests, lack of functional impairment, and absence of neurological or psychiatric conditions that might impair cognition.[25, 26] CSF was collected by lumbar puncture of individuals under written informed consent. The University of Wisconsin Institutional Review Board approved all study procedures. Each enrollee provided a signed informed consent form before participation. CSF aliquots from each of the 16 individuals at each stage were combined into a pool of 1 mL for control, MCI and AD subjects.

PANC1 cells

PANC1 pancreatic ductal adenocarcinoma cells were maintained in complete media of DMEM/Ham's F-12 (1:1) (ATCC) supplemented with 10% fetal bovine serum (Hyclone) and 1%

antibiotic-antimycotic solution (Cellgro). Cell culture flasks were incubated in the incubator containing 5% CO_2 and 98% humidity. Cells were harvested once 80% confluence was achieved, and cells with a maximum of 15 passages were used. Cell pellets were washed twice with phosphate-buffered saline, flash frozen in dry ice, and stored at -80 ℃.

Protein extraction and digestion from PANC1 cells

Protein extraction and trypsin digestion was performed based on previously reported filter-aided sample preparation (FASP) protocol[27] with some modifications. Cell pellets were lysed by sonication in a solution containing digest buffer (4% SDS, 100 mM Tris/Base pH 8.0). To determine the protein concentration, the bicinchoninic acid assay (BCA assay) was conducted. Briefly, the proteins was thawed and centrifuged at 16, 000g for 5 min. Take out 200 μg protein to the vial and add 1 M DTT into digest buffer to make DTT final concentration 0.1 M. Incubate the sample at 95℃ for 3 min to reduce disulfide bonds. 200 μL of UA buffer (8 M UA in 100 mM Tris/Base) was added into the vial and transferred onto the 30 kDa filter. The filter was centrifuged at 14, 000g for 15 min. Add another 200 μL of UA buffer was to the sample and centrifuge at 14, 000g for 15 min. Add 100 μL of IAA buffer (0.05 M IAA in UA buffer) onto the filter and gently swirl to mix, then incubate in darkness for 20 min. And the filter was centrifuged at 14, 000g for 10 min. Add 100 μL of UA buffer onto the filter, and centrifuge at 14, 000g for 15 min. Repeat this step for another 2 times to completely exchange SDS with UA buffer. Add 100 μL of ABC buffer (50 mM) onto the filter, and centrifuge at 14, 000g for 15 min. Repeat this step for another 2 times. All the centrifugation is done at 20℃. Then add 10 μL of trypsin and 40 μL of ABC buffer onto the filter. Incubate the filter at 37 ℃ water bath for 18h. After incubation, the filter was transferred

to a fresh collection vial and centrifuged at 14, 000g for 10min. Add 50 μL 0.5 M NaCl solution onto the filter and centrifuge the filter at 14, 000g for 10min. Repeat for once. Add TFA into the vial to make its final concentration 0.25%. Samples were desalted using a SepPak C18 SPE cartridge (Waters, Milford, MA).

CSF sample processing

Separate CSF sample into peptide fraction and protein fraction using 10 kDa molecular weight cut-off (MWCO) following the previous protocol.[28, 29] The peptide fraction was injected for LC-MS/MS analysis after desalting with SepPak C18 SPE cartridge. The protein fraction was dissolved in 8 M urea, reduced (5 mM DTT, 1 h at room temperature) and alkylated (15 mM IAA, 30 min at room temperature in the dark). Alkylation was quenched by incubation in 9 mM DTT by adding a second aliquot of DTT at room temperature. Dilute the samples with 50 mM Tris buffer to make urea below 1 M. Trypsin was added in a 1:50 (w/w) ratio and incubated for 18 h at 37 ℃. Quench the digestion by adding TFA to a final concentration of 0.3%. Finally, the samples were desalted on a C18 SepPak cartridge (Waters, Milford, MA) and dried under vacuum. Tryptic peptides were then incubated with PNGase F to remove N-glycans to avoid detection interference with O-glycosylated peptides. Then the samples were subject to C18 SepPak cartridge to get rid of salts and released N-glycans. Samples were dried down under vacuum and stored under -80 ℃.

Boronic acid enrichment

Boronic acid enrichment was conducted according to previous reported protocol with slight modifications.[24] PBA cartridges were first conditioned with 1 mL of anhydrous DMSO for 3 times. Tryptic peptides were dissolved in 35μL DMSO, loaded onto the cartridge, and incubated in 37℃

for 2 h with both ends of the cartridge sealed. Wash the non-bound peptides away with 1 mL anhydrous ACN, and repeat for another 2 times. Then the O-glycopeptides were eluted after incubation in 600μL of 0.1% TFA in 37℃ for 1 h. Repeat for once. The enriched O-glycopeptides were dried down under vacuum and stored under -80 ℃ before analysis.

High-pH fractionation

Enriched O-glycopeptides were fractionated using a C18 reverse-phase column (2.1 × 150 mm, 5 μm, 100 Å) operating at 0.3 mL/min in high-pH mode. Samples were first reconstituted in 100 μL of 10 mM ammonium formate at pH 10 (mobile phase A). Mobile phase B consisted of 90% CAN and 10 mM ammonium formate at pH=10. O-glycopeptides were eluted with a gradient as follows: 1 % A (0–3 min), 1-35% (3-50 min), 35-60% (50-54 min), 60-70% (54-58 min), and 70-100 % (58-59 min). Seven fractions were collected from 4 min to 62 min. Fractions were dried down under vacuum.

LC-MS/MS analysis

After dissolved in 0.1% FA (mobile phase A), samples were analyzed on the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Dionex UPLC system. Mobile phase B consisted of 0.1% formic acid in ACN. Peptides were loaded and separated on a 75 μm x 15 cm homemade column packed with 1.7 μm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661). The LC gradient was set as follows, 3%-30% A (18-98min), 30%-75% A (100-108 min) and 75%-95% A (108-118min). The mass spectrometer was operated in data dependent (DDA) mode to automatically switch between MS and MS/MS acquisition. An MS1 scan was acquired from 400–1800 (120,000

resolution, $4e^5$ AGC, 100 ms injection time) followed by EThcD MS/MS acquisition of the precursors with the highest charge states in an order of intensity and detection in the Orbitrap (60,000 resolution, $3e^5$ AGC, 100 ms injection time). EThcD was performed with optimized user defined charge dependent reaction time (2+ 50 ms; 3+ 20 ms; 4+ 20 ms; 5+ 20ms; 6 + 9 ms; 7+; 9 ms; 8+ 9ms) supplemented by 33% HCD activation.

Data analysis

All raw data files were searched against UniProt *homo sapien*s reviewed database (08.10.2016, 20, 152 sequences), using PTM-centric search engine Byonic (version 2.9.38, Protein Metrics, San Carlos, CA) incorporated in Proteome Discoverer (PD 2.1). Trypsin was selected as the enzyme and two maximum missed cleavages were allowed. Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.01 Da. Static modifications consisted of carbamidomethylation of cysteine residues (+57.02146 Da). Dynamic modifications consisted of oxidation of methionine residues (+15.99492 Da), deamidation of asparagine and glutamine (+0.98402 Da). Oxidation and deamidation were set as “rare” modification, and O-glycosylation was set as “common” modification through Byonic node. Two rare modification and one common modification were allowed. Human O-glycan database embedded in Byonic, which contains 70 glycan entities, was used. As for glycopeptide FDR control, Byonic default settings was applied that cuts the protein list after the 20th decoy proteins or at the point in the list at which the protein FDR first reaches 1%, whichever cut gives more proteins. After that, Byonic estimated the spectrum-level FDR of the remaining PSMs to the reported proteins which will typically be in the range 0-5%. Only these O-glycopeptides with PSMs FDR ≤ 1% and Byonic score over 50 and

Delta Mod Score over 40 were reported. For endogenous peptides, three consecutive searches were conducted as shown in **Figure S1**. Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.01 Da, and digestion was set to unspecific. The 1st search was using the whole human protein database and total "rare" modification was set to 1, and dynamic modifications includes oxidation of methionine residues (+15.99492 Da), amidation at peptide C-terminal (-0.984016), acetylation at peptide N-terminal and lysine, serine (+42.010565), and Gln to pyro-Glu conversion (-17.026549). The proteins precursors identified in the first search were combined with the reported protein precursors[21, 22] to construct focused protein database. The 2nd search was using the same dynamic modification as first search, except adding N-glycosylation and O-glycosylation as "common" modification, with total common modification set as 1. The human glycan database used is 182 entities for N-glycan and 70 entities for human O-glycan, and focused protein database was used. The 2nd searche yielded a list of identified N-glycans and O-glycans, and these glycans were used to build a focused glycan database. For the 3nd search, dimethylation at peptide N-terminal, lysine, arginine (+28.0313), deamidation at asparagine (+0.984016), methylation at peptide N-terminal, lysine, arginine, glutamine (14.01565), phosphorylation at serine, tyrosine, threonine (+79.966331) were added to dynamic "rare" modifications. N-glycosylation and O-glycosylation as "common" modification, with total common modification set as 1. Focused protein database and focused glycan database were used in the 3rd search. Only these peptides with PSMs FDR $\leqslant$ 1% and Byonic score over 50 were reported. For endogenous peptides with PTMs, a Delta Mod Score over 40 was required. The reported O-glycosylation information was extracted from Uniprot (07.13.2017). The peptide

sequence analysis of glycosylation site-containing peptides was conducted using the on-line tool Weblogo 3 (http://weblogo.threeplusone.com).[30]

Results

Optimization of boronic acid enrichment strategy

To avoid the interference of highly abundant nonglycosylated peptides, enrichment is a key step for the success of glycopeptide detection. Different aspects of the properties of O-glycopeptide could be utilized for enrichment purpose, including lectin recognized glycan motif, metabolic labeling by azido analog and sialic acid binding affinity to titanium dioxide (TiO2) beads.[31] The *cis*-diol groups on sialic acid group could also be converted to aldehydes by periodate oxidation, and the sialyalted glycopeptides could further be enriched based on hydrazide chemistry.[19] In addition, the *cis*-diol groups could react with boronic acids to form reversible covalent bond and later be released under acidic conditions without any side effect.[32, 33] The presence of *cis*-diol groups in all glycoproteins and glycopetpides makes it a universal enrichment method for both N-glycopeptide and O-glycopeptide. Although this method has been applied for large-scale N-glycoproteome study[34] and small-scale analysis of O-GlcNAcylation[24], it has not yet been used for the comprehensive O-glycoproteome analysis in complex biological samples.

Here, we used phenylboronic acid (PBA) solid phase extraction cartridge to extract the O-glycopeptides from complex tryptic peptides sample. PNGase F was first used to remove the N-glycans to avoid its interference with the enrichment process and the later MS detection of O-glycopeptides. Starting from 200 μg tryptic peptides from PANC1 cells, a total of 213 intact O-

glycopeptides were identified (**Figure 1a**). In the spectra, we found that quite many nonglycosylated peptides were co-enriched which may suppress the glycopeptide signal. Thus, to reduce sample complexity and minimize the effects of co-eluted nonglycosylated peptides, off-line high-pH fractionation (HpH) was utilized, which has also been shown highly orthogonal to the following LC-MS/MS analysis using low-pH reversed-phase chromatography.[35] Seven HpH fractions were collected and injected for analysis. The results showed that 229 intact O-glycopeptides were identified, with only a slight improvement (**Figure 1a, supplemental table S2)**. A closer look at the distribution of the number of O-glycopeptides identified showed that O-glycopeptides were mainly distributed in the first four fractions **(Figure 1b)**, whereas the number of peptides identified in a typical proteomics study using the same HpH methods resulted in an evenly distributed peptides number in all fractions. This should be no surprise considering that the increased hydrophilicity of O-glycopeptides would shift the elution to an earlier time frame on a C18 column. Then the first fraction was combined with the last three fractions and the saved instrument time by using less fractions will be used for more technical replicates. Another round of experiment using the same amount of starting material employing four fractions and four technical replicates yielded a total number of 987 intact O-glycopeptides, with a 4-fold increase. Such a large increase mainly result from the four replicates employed for each fraction. It is well-known that two or more technical replicates are needed to get the maximum coverage of peptides in a certain sample due to the randomness of data-dependent analysis (DDA) mode in short-gun proteomics.[36] Such effects would be more significant when it comes to glycoproteomics study, as the co-enriched nonglycosylated peptide would compete with glycopeptide for fragmentation and

the higher ionization efficiency would help them gain advantage over glycosylated peptide. Here, four replicates were used to find an optimal number of replicates. As shown in **Figure 1c**, two replicates only yielded less than 60% coverage, and at least three technical replicates were needed to reach a coverage of more than 90%.

Site-specific O-glycopeptide characterization

Due to the inherent structural complexity of O-glycopeptide, multiple facets exist for site-specific O-glycopeptide analysis, including glycosylation site localization, glycan structure and peptide sequence information. Various analytical technics have been developed to advance the MS-based O-glycopeptide analysis, which has greatly improved the overall performance of large-scale O-glycoproteomics workflow.[31] One focus is the development and optimization of fragmentation techniques. In general, c/z ions produced in ETD-MS/MS provide information on the glycosylation site and the peptide identity, but the abundant unreacted and charge reduced precursors hamper its performance and glycan fragments B/Y ions cannot be obtained in this mode. HCD produces B/Y ions and abundant oxonium ions for glycan identification, as well as b/y ions for peptide sequence information, but it does poorly in glycosylation site localization. In order to obtain comprehensive information, multiple dissociation modes are usually needed to draw a complete picture of O-glycopeptide structure. To this end,

the “hybrid” dissociation method electrontransfer and higher-energy collision dissociation (EThcD), introduced by Heck and co-workers[37] has shown great potential for O-glycopepitde analysis. Its superior performance in labile PTM analysis was first demonstrated in phosphopeptide

analysis, with improved site localization, richer and more informative peptide backbone spectra compared to ETD and HCD.[38] When applied to O-glycopeptide analysis, EThcD can produce rich fragment ions information for glycan moiety, peptide backbone and glycosylation site identification in one spectrum, enabling enhanced site-specific O-glycopeptide analysis.[39-42]

As an example, EThcD spectrum of the identified O-glycopeptide VHENENIGTTEPGEHQEAK was shown in **Figure 2a**. In the low mass region, the glycan signature oxonium ions, including 138.06 (HexNAc-$2H_2O$-CH_2O), 168.06 (HexNAc-$2H_2O$), 186.08 (HexNAc-H_2O) and 204.09 (HexNAc), indicates the spectrum belongs to a glycopeptide. The glycan fragments 274.05 (NeuAc) and 292.11 (NeuAc-18) suggests the presence of sialic acid. Based on these information, along with fragments HexNAcHex (366.14), HexNAcHexNeuAc (657.23) and accurate precursor mass match (1.2 ppm), the O-glycan composition can be assigned as HexNAcHexNeuAc, a sialylated Tf antigen (**Figure 2a)**. The peptide sequence VHENENIGTTEPGEHQEAK were deduced based on the wealthy backbone fragments b/y ions and c/z ions. Two possible O-glycosylation sites, 9th and 10th threonine, exist on the peptide sequence, and the c2−c10 ions enabled unambiguous localization of the glycan at 9th threonine, which agrees well with previous reports.[19, 20] Besides, abundant glycopeptide peaks with total or partial glycan moiety loss (pep, pep+HexNAc, pep+HexNAcHex) were also detected, further confirming the assignment. In the same way, the O-glycosites of endogenous O-glycopeptide AAVGTSAAPVPSDNH, with 3 potential O-glycosites (5th T, 6th S, 12th S), were also unambiguously localized at 6th serine residue using the c3-c7 ions, and peptide sequence and O-glycan compositions were deduced using c/z, b/y and glycan fragment ions (**Figure 2b)**. This O-

glycopeptide originates from Apoliprotein E, and the O-glycosite at 6th serine has been tentatively identified and reported as a novel O-glycosite in the previous study.[19] In the previous report, although the author narrowed down the O-glycosite to the 5th threonine and 6th serine, the accurate site was not able to be assigned due to a lack of informative fragments generated by electron-capture dissociation (ECD). Furthermore, limited by the enrichment method (sialic acid was periodate-oxidized) used in their study, the sialylation information of the O-glycan was lost. The EThcD spectrum in the current study not only helps us precisely pinpoint this novel O-glycosite, but also explicitly assign the O-glycan composition as HexNAcHexNeuAc(2), fully preserving sialylation information. Similar to O-glycosylation, phosphorylation is also quite labile. Benefitting from EThcD, the detection of c4 ion without phosphorylation group and c5 ion with phosphorylation group help us unambiguously localize the phosphorylation site at the 5th serine residue. Thus, along with other b/y, c/z ions and 98 Da neutral loss, the endogenous phosphorylated peptide VDPKSKEEDKH could be confidently identified as shown in **Figure 2c.**

Site-specific O-glycoproteome analysis in CSF

As discussed previously, the current depth of site-specific O-glycoproteome in CSF is not satisfactory. One of our goal in the present study is to improve the current O-glycoproteome coverage by developing a more robust analytical strategy. Benefiting from the optimized O-glycopeptide enrichment strategy, along with robust hybrid fragmentation EThcD techniques, we were able to identify 308 intact O-glycopeptides from 181 O-glycosites and 110 O-glycoproteins, which is a large increase compared to previous reports (**Figure 3a**). After combining O-

glycopeptides with the same glycosite and glycan composition but different peptide sequence length, there were 292 unique O-glycoforms in total (**Figure 3a**). Overall, the majority of O-glycoproteins (72%) carry only one single O-glycosite, and 25% carry 2-3 O-glycosites, with a slight percentage (5%) more than 4 sites (**Figure 3b**). There are around 30% O-glycosites identified carrying more than 2 O-glycan compositions, indicating the microheterogeneity of O-glycosylation (**Figure 3c**).

Among different kinds of O-glycoforms, mucin-type core 1 (Galβ1→3GalNAc) is known to be the major constituent of O-glycans.[43] In the present study, around 48.3% (141/292) of O-glycoforms had core 1 glycoforms (including $HexNAc_{(1)}Hex_{(1)}$, $HexNAc_{(1)}Hex_{(1)}Fuc_{(1)}$, $HexNAc_{(1)}Hex_{(1)}Fuc_{(1)}NeuAc_{(1)}$, $HexNAc_{(1)}Hex_{(1)}NeuAc_{(1)}$, $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$, $HexNAc_{(1)}Hex_{(1)}NeuAc_{(3)}$, $HexNAc_{(1)}Hex_{(2)}Fuc_{(1)}$). This O-glycosylation feature also agrees quite well with the core 1 O-glycoform percentage in human serum, with 46.4% reported.[44] The sialylated core 1 glycoforms, mainly $HexNAc_{(1)}Hex_{(1)}NeuAc_{(1)}$ and $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$, were the domain compositions (~67%), and the same phenomenon was also observed in human serum. This is probably because more than 80% of CSF proteins originates from the plasma filtrate, preserving a similar glycosylation pattern.[19, 44, 45] Such sialylation signature information obtained here could never be obtained if O-glycopeptide enrichment was based on sialic acid oxidation, which was the case in previous studies.[19, 20] Overall, the sialylated O-glycoforms accounted for 51% of the total O-glycoforms, suggesting sialylation was a common modification on O-glycans and an enrichment approach that can preserve full sialylation information is highly desirable. Interestingly, acting also as a capping unit to elongate the glycan branch, fucosylated core 1 O-

glycoforms only accounted for 7% and total fucosylated O-glycoforms accounted for 29%, a lot smaller percentage compared to sialylated forms. As shown in **Figure 3d**, the number of O-glycoforms with 2 sialic acids largely exceeded the number of O-glycoforms with 2 fucoses, while the number of O-glycoforms with 1 or 3 sialic acids/fucoses were close.

Peptide sequence analysis of CSF O-glycopeptides

Unlike N-glycosylation, there is no consensus motif for O-glycosylation; whereas, some O-glycosylated peptide sequences might be preferred than others. Many preferred structure motifs have been proposed placing significant effects of the amino acids adjacent to the glycosite, including proline, alanine, serine, threonine, glycine, charged residue and amino acids with small side chains.[46-48] Among them, the effects of adjacent proline residues on the O-glycosylation biosynthesis is perhaps the most extensively studied. O'Connell et al[48] has shown the proline from −6 to +4 positions facilitates the process of glycosylating the proteins, agreeing well with another study that proposed proline residue from −4 to +4 positions are important for O-glycosylation.[47] More specifically, another study shows that the proline at -1 and +1 sites enhances O-glycosylation on recombinant erythropoietin.[49] Using a statistical analysis based on 174 O-glycosites, Wilson et al[50] has shown a high frequency of occurrence at +3 and -1 positions relative to the glycosite. Thus, we performed a proline frequency analysis of the ±10 residues surrounding the 181 identified O-glycosites. Note that such proline frequency analysis results may differ depending on the origins of the O-glycosites dataset, and the proline frequency may vary from different tissues used. Despite of being tissue-specific, we did observe that the proline residue was sequence conserved at the -1

and +3 positions based on the 181 O-glycosites identified in CSF (**Figure 4a, c**), which agrees quite well with previous global statistical analysis[50] and report in CSF.[20] As increasing amount of new O-glycosites information were added to the known database each day, we also conducted the proline frequency analysis of the experimentally verified 435 O-glycosites from Uniprot (2017_12). The results revealed that proline showed the highest frequency from -4 to +7 positions, except at +1 position only next to threonine (**Figure 4b, d**). The conservative proline frequency was most pronounced at -1 and +3 positions in accordance with our findings. Such observation indicates that although a complete picture of conservative sequence may not be fully preserved in a specific tissue carrying a subset of global O-glycosites, the most pronounced conservative sequence was able to be preserved. These conservative sequence features that are unique to CSF may represent CSF-specific features affected by factors such as the origins of the glycoproteins, different biosynthesis routes of these glycoproteins, glycotransferase enzyme actives etc. One of these CSF-specific features is the higher frequency of glutamic acid (E) adjacent to the O-glycosite compared to the results from global O-glycosites database, especially at +1 and +2 positions. Previous study has shown O-glycosylation was markedly reduced when glutamic acid residue was substituted at position -1 and +3, but glutamic acid replacement at +1 and +2 had no effect,[51] which may explain a higher glutamic acid frequency at +1 and +2 positions rather than other positions.

Selected examples of CSF O-glycoproteins

Selected identified O-glycoprotein examples were discussed below. In the previous CSF O-glycoproteome study, Halim et al. discussed the situation of O-glycosylation identified at Thr

residues of the N-glycosylation Asn-X-Ser/Thr consensus sequence.[20] To make a comparison with their results, we will start out discussing these identified O-glycopeptides containing this special O-glycosylation site. As a first example, the GNLTGAPGQR peptide from endothelin B receptor-like protein 2 (ETBR2) was found to be O-glycosylated at 4th Thr residue, and the 2nd asparagine (Asn) underwent deamination **(Table 1)**. As enzyme PNGase F was used in our study to first get rid of N-glycans, any Asn residue with an N-glycan attached would go through deamidation. It should also be noted that chemical deamidation may also occur during sample preparation process,[52] so further experimental evidence is needed before the deamidated Asn can be assigned to an N-glycosite. Sequence analysis showed that the Asn residue is in the Asn-X-Ser/Thr N-glycosylation consensus motif and it has been annotated as N-glycosylation site in Uniprot. In addition, in our previous CSF N-glycoproteome study (data in a separate manuscript), the 2nd Asn residue has been confidently identified as N-glycosite. Thus it confirmed that the deamidation at 2nd Asn residue was truly due to the release of N-glycans by PNGase F. Such finding is in agreement with Halim et al report.[20] In addition, our study was able to identify the intact O-glycan compositions as $HexNAc_{(2)}Hex_{(2)}$ and $HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}NeuAc_{(1)}$. In the same protein ETBR2, they also identified O-glycopeptide VSGGAPLHLGR from ETBR2, which is different from the proposed signal sequence cleavage at 4th Gly. In our study, the same O-glycopeptide with $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$ attached was identified supporting the previous finding.[20] A second example was the O-glycopeptide LPTTVLNATAK from protein YIPF3, with O-glycosylation at 9th Thr and demidation at 7th Asn. Both previous reports and our N-glycoproteome study confirmed 7th Asn as a true N-glycosite, which is also in accordance with Halim et al. report.[20] Two sialylalted

O-glycoforms $HexNAc_{(1)}Hex_{(1)}NeuAc_{(1)}$ and $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$ were identified. A third example of O-glycosylation within the N-glycosylation consensus motif that hasn't been reported before are the two O-glycopeptides ELPGVCNETMMALWEECK and LANLTQGEDQYYLR from Apolipoprotein J (also known as clusterin) **(Table 1).** The deamidation at Asn residue in the two peptide sequences, along with our previous N-glycoproteome study results and literature report,[19, 53, 54] confirmed their identity as N-glycosites. The reason that we were able to detect these two sites was related to the different O-glycoforms attached to the two sites on apoliprotein J. In contrary to the sialylated O-glycoforms detected in the first two examples, the O-glycoforms detected here were nonsialyated forms $HexNAc_{(3)}Fuc_{(1)}$ and $HexNAc_{(6)}Hex_{(5)}Fuc_{(3)}$. The non-sialylated O-glycopeptide cannot be enriched by the previous enrichment method based on sialic acid oxidation. Whereas, the boronic acid enrichment method used in our study was able to capture the non-sialylated O-glycoforms, which enables the detection of this nonsialylated O-glycopeptide.

As structural components of lipoprotein particles, apoliproteins play important roles in maintaining their structure and regulating their metabolism and enzyme activities. Studies have shown apoliproteins are often modified by O-glycosylation, including apoliprotein E, A-I, A-II and C-III.[55-57] In our study, four apoliproteins, apoliprotein D, E, A-I, and J were found to be O-glycosylated **(Table 1)**. Although N-glycosylation of apoliprotein D (APOD) has been extensively studied at 65Asn and 98Asn and many potential O-glycosylation sites (8 serine residues and 10 threonine residues) existed in its peptide sequence, O-glycosylation of APOD has not been reported before. The APOD-derived tryptic peptides VLNQELRADGTVNQIEGEATPVNLTEPAK and ADGTVNQIEGEATPVNLTEPAK were

found to be O-glycosylated at 86Thr, and deamidation was also found at 96Asn in accordance with the expected outcome of the PNGase F pre-treatment, indirectly supporting the assignment. A total of three O-glycoforms were detected at this site, and all of them were fucosylated ($HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}$, $HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}NeuAc_{(2)}$, $HexNAc_{(4)}Hex_{(4)}Fuc_{(3)}$). For apoliprotein E (APOE), two O-glycosites at 36th Thr and 212th Thr were identified with 11 O-glycoforms and 2 O-glycoforms detected at each site respectively. Using lectin-based isolation method, Cubedo et al reported the O-glycosylation of serum apoliprotein A-I, but the exact O-glycosite or O-glycoform information was lacking.[58] In our study, the O-glycopeptide EQLGPVTQEFWDNLEK from apoliprotein A-I was identified, with 1 O-glycoform detected ($HexNAc_{(3)}Hex_{(2)}Fuc_{(1)}$). Despite the fact that APOJ has been shown carrying several PTMs such as N-glycosylation, ubiquitination and phosphorylation, O-glycosylation hasn't been reported before.[59-62] In our study, three O-glycosites 105Thr, 210Ser, 376Thr with 5 O-glycoforms were found. Only 210Ser carried a sialylated O-glycoform, and the rest of them carried non-sialylated O-glycoforms.

Comprehensive PTMs analysis of CSF endogenous peptides

Although deemed as a common PTM for endogenous peptides, glycosylation was not considered as a possible dynamic modification in most of current peptidomics workflow. This is largely because unlike other PTMs such as phosphorylation, glycosylation requires its own glycan database that contains hundreds of glycan compositions, which requires special software platforms with the ability to *de novo* glycopeptide spectra.[63] In addition, glycosylation is a quite labile PTM and can be easily lost in the most commonly used CID/HCD-based peptidomics workflow. As a

result, to our best knowledge currently there is no peptidomics workflow that takes the glycosylation into account, along with other common PTMs. To solve this two problems, EThcD was utilized to preserve the labile glycosylation and the raw data will be analyzed by PTM-centric Byonic searching engine to enable site-specific glycoprofiling.[64] We first tried to search the raw data directly by setting all the possible PTMs (glycosylation, phosphorylation, acetylation etc.) and using the whole human protein and glycan database, which leaded to computational explosion. Then a constructed "focused" CSF endogenous peptides database based on literature was used but also failed, indicating that the size of glycan database may be too large. Therefore, we developed a 3-step searching strategy (**Figure S1**) to facilitate the search and details could be found in "Experimental Procedures" section. The key idea is that focused protein and glycan database would be constructed by the first and second wild searches, and would be used to facilitate the final search, where all the possible PTMs (including N-glycosylation and O-glycosylation) could be searched all at once.

For endogenous peptides analysis in CSF, our lab previously developed an optimized 10kDa MWCO-based protocol to achieve an optimal recovery rate.[28, 29] Here, we compared 10kDa MWCO-based protocol developed in our lab with another 30kDa MWCO-based protocol reported previously in terms of their recovery rate.[65] As shown in **Figure 5 (Supplemental Table S3)**, 10k-based protocol outperformed 30k-based protocol in both the total number of identified endogenous peptides, peptides without PTMs and peptides with PTMs. Thus, the 10k-based protocol was used to separate CSF samples into peptide fraction and protein fraction (**Figure S2**). Peptide fraction would be directly injected for LC-MS/MS analysis after desalting, and the acquired data would be

subject to search engine Byonic for comprehensive PTMs analysis.

In CSF samples from healthy individual, a total of 1492 endogenous peptides were identified (**Figure 6**), with an over 2-fold increase compared to previous reports.[21, 22] The large increase could possibly result from multiple factors including the use of instrument with higher sensitivity, optimized 10kDa MWCO protocol to increase the recovery rate, as well as more comprehensive PTMs included. Benefiting from the hybrid fragmentation method EThcD utilized in our study, the labile PTMs including N-/O-glycosylation and phosphorylation could be well preserved, which otherwise could be lost in CID used by previous studies. In total, 370 endogenous peptides with PTMs were identified **(Figure 6**), indicating that endogenous peptides in CSF went through extensive modifications. Among them, 95 and 35 peptides were O-glycosylated and phosphorylated, respectively. It is worth noting that the present study is the first report that CSF endogenous peptides were extensively O-glycosylated, and in fact the number of O-glycosylated endogenous peptides exceeded all other PTMs modified peptides. These O-glycopeptides were derivatized from 32 protein precursors, and 28 O-glycan compositions were identified. Interestingly, only 1 N-glycosylated peptide T<u>N</u>STFVQALVEHVK from precursor protein prosaposin was found with N-glycosite at previously reported 215th Asn residue.

Of all these identified endogenous peptides, neuropeptides as a subset is of special interest due to their large variety of functions as inter-cellular signaling molecules.[66] By referring to the human neuropeptide database NeuroPep,[67] 282 were identified as neuropeptides **(Figure 6)**. More than half of them (182, 65%) belonged to chromogranin/secretogranin family, with ProSAAS (28, 10%), NPY (19, 7%), VGF (18, 6%), opioid (11, 4%) and 7B2 (7, 2%) following behind (**Figure**

S3). As mentioned previously, neuropeptides could undergo further post-translational modifications after initial peptidase cleavages. In the present study, 88 out of the 282 identified neuropeptides were modified by PTMs, including acetylation, amidation, O-glycosylation, phosphorylation, Gln to pyro-Glu conversion. Among them, 15 neuropeptides were O-glycosylated, with 14 originating from ProSAAS and 1 from secretogranin-1. Broadly expressed in brain and neuroendocrine tissues, ProSAAS is the precursor of a number of neuropeptides.[68] The proSAAS-derived neuropeptides SAAS, PEN, and LEN, which are among the most abundant peptides present in mouse hypothalamus,[69-71] have been implicated in the regulation of food intake and body weight.[72, 73] In the present study, the two O-glycosylation sites (53Asn, 247Asn) detected were in the LEN and SAAS region, and 11 O-glycosylated LEN peptides and 3 O-glycosylated SAAS peptides were identified (**Supplemental Table S4**). Previous study of intact glycoprotein ProSAAS also identified these two sites, but the detailed O-glycoforms information were missing. At site 53Asn, two O-glycoforms with sialylated T-antigen were identified, whereas at site 247Asn three O-glycoforms with T-antigen in both sialylalted form and non-sialylated forms were identified, along with another O-glycoform with non-sialylated Tn-antigen. O-glycosylated neuropeptide EEKGHPQEESEESNV<u>S</u>MA derived from precursor protein secretogranin-1 was also identified with fucosylated O-glycan attached ($HexNAc_{(3)}Hex_{(3)}Fuc_{(1)}$).

CSF O-glycoproteome and endogenous peptidome mapping in MCI and AD

There have been a few studies that reported the N-glycosylation pattern alterations in some of the well-known AD-related proteins, including APP, tau, acetylcholinesterase, and transferrin.[74-77]

Some other studies described global glycosylation-related alterations in the brain and CSF of AD patients.[78-83] Most of these global studies relied on lectin-based staining techniques by recognizing specific glycan motif. Even fewer studies exist for O-glycosylation alteration study due to its larger degree of structure complexity and diversity. As a special case of O-glycosylation, O-GlcNAcylation has been shown to protect tau against aberrant phosphorylation and subsequent aggregation, suggesting O-GlcNAcase could be used as a potential therapeutic target in AD intervention.[77, 84, 85] However, the landscape O-glycosylation mapping in states along the AD progression hasn't been conducted yet. Here, we applied the developed site-specific O-glycoproteomic strategies into CSF samples from MCI and AD to get an idea of the overall picture of O-glycosylation in these disease states. Venn diagram analysis of O-glycoproteins, O-glycosites and O-glycopeptides from control, MCI and AD was shown in **Figure 7a (Supplemental Table S5)**. In order to highlight the more specific features of O-glycosylation pattern belonging to each stage, the percentage of different categories of glycoforms were compared. As discussed previously, the core 1 O-glycoform accounts for the most percentage of all kinds of O-glycoforms, so we compared the core 1 percentage across the three states, along with their sialylated and fucosylated forms. The percentage of global sialylated and fucosylated O-glycoforms were also compared, because altered sialylation and fucosylation were often implicated in various diseases.[86] However, quite comparable percentage of core 1, fucosylated core 1, sialylatd core 1 and global sialylated was obtained between these three states except for the fucosylated O-glycoforms, which showed a decrease trend in MCI and AD (**Figure 7b, Supplemental Table S5**). This is quite interesting because we also found alteration in fucosylation was the most dominant in another

separate N-glycosylation study (in a separate manuscript). Unfortunately, there is few study we can correlate, and further studies are needed to reveal the role of fucosylation in AD mechanism. Endogenous peptides analysis showed that there is a decrease in the number in MCI and AD, and a closer look into the percentage constitute showed shat the decrease is mainly the unmodified peptides (**Figure 8a, Supplemental Table S6**). In terms of the different kind of PTMs detected in control, MCI and AD, there is an gradual increase for O-glycosylation, Gln->pyro-Glu, acetylation, phosphorylation and a decrease for amidation, deamination. (**Figure 8b, Supplemental Table S6**). Such observations above represents a global landscape of the O-glycoproteome and endogenous peptidome from control, MCI and AD, revealing the dominant differences in different states. Further qualitative and quantitative analysis based on individual subject is needed to account for individual variation, pinpointing more specific changes of interesting glycoprotein/glycosite/glycoform.

Conclusion

In summary, a workflow that combined site-specific O-glycoproteome study and comprehensive PTMs analysis of endogenous peptides in CSF was developed. For O-glycoproteome study, an optimized boronic acid enrichment method combined with HpH fractionation was developed for enhanced site-specific O-glycopeptide analysis enabled by EThcD. After applying the optimized approach to CSF O-glycoproteome study, 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins were identified, representing the largest dataset for CSF site-specific O-glycoproteome study so far. For endogenous peptides study, a peptidomics workflow that enables

comprehensive PTMs analysis, including glycosylation, phosphorylation etc, of CSF endogenous peptides was developed, which was the first report that incorporates both N-glycosylation and O-glycosylation in the workflow. It was found that 95 CSF endogenous peptides were O-glycosylated, whereas few peptides were N-glycosylated. Among them, 15 O-glycosylated peptides were neuropeptides originating from ProSAAS. Besides, site-specific O-glycoproteome and endogenous peptides analysis were also conducted in CSF samples from MCI and AD patients, which yielded a glycosylation landscape picture to guide any following-up studies.

Acknowledgements

This research was supported in part by the National Institutes of Health (NIH) grants R21AG055377, R01 DK071801, and R56 MH110215. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges a Vilas Distinguished Achievement Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. We thank Dr. Marshall Bern from Protein Metrics for providing access to Byonic software package.

References:

1. McComb, J. G., Recent research into the nature of cerebrospinal fluid formation and absorption. *J. Neurosurg.* **1983,** 59, (3), 369-383.
2. Sakka, L.; Coll, G.; Chazal, J., Anatomy and physiology of cerebrospinal fluid. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **2011,** 128, (6), 309-316.
3. Stopa, E. G.; Berzin, T. M.; Kim, S.; Song, P.; Kuo-LeBlanc, V.; Rodriguez-Wolf, M.; Baird, A.; Johanson, C. E., Human choroid plexus growth factors: What are the implications for CSF dynamics in Alzheimer's disease? *Exp. Neurol.* **2001,** 167, (1), 40-47.
4. Chodobski, A.; Wojcik, B. E.; Loh, Y. P.; Dodd, K. A.; Szmydynger-Chodobska, J.; Johanson, C. E.; Demers, D. M.; Chun, Z. G.; Limthong, N. P., Vasopressin gene expression in rat choroid plexus. In *Vasopressin and Oxytocin*, Springer: 1998; pp 59-65.
5. Zhang, J., Proteomics of human cerebrospinal fluid–the good, the bad, and the ugly. *PROTEOMICS-Clinical Applications* **2007,** 1, (8), 805-819.
6. Fonteh, A. N.; Harrington, R. J.; Huhmer, A. F.; Biringer, R. G.; Riggins, J. N.; Harrington, M. G., Identification of disease markers in human cerebrospinal fluid using lipidomic and proteomic methods. *Dis. Markers* **2006,** 22, (1-2), 39-64.
7. Yuan, X.; Desiderio, D. M., Proteomics analysis of human cerebrospinal fluid. *J. Chromatogr. B* **2005,** 815, (1), 179-189.
8. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.
9. Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J., Essentials of glycobiology. 1999. *Cold Spring Harber Laboratory Press, New York* **1998**.
10. Tabak, L. A. In *The role of mucin-type O-glycans in eukaryotic development*, Semin. Cell Dev. Biol., 2010; Elsevier: 2010; pp 616-621.
11. Ten Hagen, K. G.; Fritz, T. A.; Tabak, L. A., All in the family: the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* **2003,** 13, (1), 1R-16R.
12. Tabak, L. A., In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. *Annu. Rev. Physiol.* **1995,** 57, (1), 547-564.
13. Tran, D. T.; Ten Hagen, K. G., Mucin-type O-glycosylation during development. *J. Biol. Chem.* **2013,** 288, (10), 6921-6929.
14. Yang, X.; Qian, K., Protein O-GlcNAcylation: emerging mechanisms and functions. *Nature Reviews Molecular Cell Biology* **2017**.
15. Bennett, H., Glycosylation, phosphorylation, and sulfation of peptide hormones and their precursors. *Peptide Biosynthesis and Processing* **1991**, 111-140.
16. Bradbury, A.; Smyth, D., Modification of the N-and C-termini of bioactive peptides: amidation and acetylation. *Peptide Biosynthesis and Processing* **1991**, 231-250.
17. Thotakura, N. R.; Blithe, D. L., Glycoprotein hormones: glycobiology of gonadotrophins, thyrotrophin and free α subunit. *Glycobiology* **1995,** 5, (1), 3-10.
18. Yu, Q.; Canales, A.; Glover, M. S.; Das, R.; Shi, X.; Liu, Y.; Keller, M. P.; Attie, A. D.; Li, L., Targeted Mass Spectrometry Approach Enabled Discovery of O-Glycosylated Insulin and Related

Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal. Chem.* **2017,** 89, (17), 9184-9191.
19. Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G., Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **2009,** 6, (11), 809-811.
20. Halim, A.; Rüetschi, U.; Larson, G. r.; Nilsson, J., LC–MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* **2013,** 12, (2), 573-584.
21. Höttä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. *PLoS One* **2012,** 7, (8), e42555.
22. Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kumar, C.; Mann, M., Integrated analysis of the cerebrospinal fluid peptidome and proteome. *J. Proteome Res.* **2007,** 7, (01), 386-399.
23. Zielinska, D. F.; Gnad, F.; Schropp, K.; Wiśniewski, 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* **2012,** 46, (4), 542-548.
24. Wang, X.; Yuan, Z.-F.; Fan, J.; Karch, K. R.; Ball, L. E.; Denu, J. M.; Garcia, B. A., A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol. Cell. Proteomics* **2016,** 15, (7), 2462-2475.
25. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1999,** 1473, (1), 4-8.
26. Zhang, Y.; Zhang, C.; Jiang, H.; Yang, P.; Lu, H., Fishing the PTM proteome with chemical approaches using functional solid phases. *Chem. Soc. Rev.* **2015,** 44, (22), 8260-8287.
27. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nature methods* **2009,** 6, (5), 359-362.
28. Cunningham, R.; Wang, J.; Wellner, D.; Li, L., Investigation and reduction of sub‐microgram peptide loss using molecular weight cut‐off fractionation prior to mass spectrometric analysis. *J. Mass Spectrom.* **2012,** 47, (10), 1327-1332.
29. Wang, J.; Cunningham, R.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L., Label‐free quantitative comparison of cerebrospinal fluid glycoproteins and endogenous peptides in subjects with Alzheimer's disease, mild cognitive impairment, and healthy individuals. *PROTEOMICS-Clinical Applications* **2016,** 10, (12), 1225-1241.
30. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res.* **2004,** 14, (6), 1188-1190.
31. Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y., Advances in mass spectrometry driven O-glycoproteomics. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2015,** 1850, (1), 33-42.
32. Sparbier, K.; Koch, S.; Kessler, I.; Wenzel, T.; Kostrzewa, M., Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. *Journal of biomolecular techniques: JBT* **2005,** 16, (4), 407.
33. Xu, Y.; Wu, Z.; Zhang, L.; Lu, H.; Yang, P.; Webley, P. A.; Zhao, D., Highly specific

enrichment of glycopeptides using boronic acid-functionalized mesoporous silica. *Anal. Chem.* **2008,** 81, (1), 503-508.
34. Chen, W.; Smeekens, J. M.; Wu, R., A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol. Cell. Proteomics* **2014,** 13, (6), 1563-1572.
35. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **2014,** 13, (12), 6176-6186.
36. Liu, H.; Sadygov, R. G.; Yates, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004,** 76, (14), 4193-4201.
37. Frese, C. K.; Altelaar, A. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012,** 84, (22), 9668-9673.
38. Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. M.; Mechtler, K.; Heck, A. J.; Mohammed, S., Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **2013,** 12, (3), 1520-1525.
39. Marino, F.; Bern, M.; Mommen, G. P.; Leney, A. C.; van Gaans-van den Brink, J. A.; Bonvin, A. M.; Becker, C.; van Els, C. c. A.; Heck, A. J., Extended O-GlcNAc on HLA class-I-bound peptides. *J. Am. Chem. Soc.* **2015,** 137, (34), 10922-10925.
40. Zhang, Y.; Xie, X.; Zhao, X.; Tian, F.; Lv, J.; Ying, W.; Qian, X., Systems analysis of singly and multiply O-glycosylated peptides in the human serum glycoproteome via EThcD and HCD mass spectrometry. *J. Proteomics* **2018,** 170, 14-27.
41. Glover, M. S.; Yu, Q.; Chen, Z.; Shi, X.; Kent, K. C.; Li, L., Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass spectrom.* **2017**.
42. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.-W.; Kent, K. C.; Li, L., Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J. Am. Soc. Mass Spectrom.* **2017,** 28, (9), 1751-1764.
43. Mitoma, J.; Petryniak, B.; Hiraoka, N.; Yeh, J.-C.; Lowe, J. B.; Fukuda, M., Extended core 1 and core 2 branched O-glycans differentially modulate sialyl Lewis X-type L-selectin ligand activity. *J. Biol. Chem.* **2003,** 278, (11), 9953-9961.
44. Qin, H.; Cheng, K.; Zhu, J.; Mao, J.; Wang, F.; Dong, M.; Chen, R.; Guo, Z.; Liang, X.; Ye, M., Proteomics Analysis of O-GalNAc Glycosylation in Human Serum by an Integrated Strategy. *Anal. Chem.* **2017,** 89, (3), 1469-1476.
45. Darula, Z.; Sherman, J.; Medzihradszky, K. F., How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol. Cell. Proteomics* **2012,** 11, (7), O111. 016774.
46. Hansen, J. E.; Lund, O.; Engelbrecht, J.; Bohr, H.; Nielsen, J. O.; Hansen, J. E.; Brunak, S., Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase. *Biochem. J.* **1995,** 308, (3), 801-813.

47. Elhammer, A. P.; Poorman, R. A.; Brown, E.; Maggiora, L. L.; Hoogerheide, J.; Kezdy, F., The specificity of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides. *J. Biol. Chem.* **1993,** 268, (14), 10029-10038.
48. O'Connell, B.; Hagen, F.; Tabak, L., The influence of flanking sequence on the O-glycosylation of threonine in vitro. *J. Biol. Chem.* **1992,** 267, (35), 25010-25018.
49. Elliott, S.; Bartley, T.; Delorme, E.; Derby, P.; Hunt, R.; Lorenzini, T.; Parker, V.; Rohde, M. F.; Stoney, K., Structural requirements for addition of O-linked carbohydrate to recombinant erythropoietin. *Biochemistry* **1994,** 33, (37), 11237-11245.
50. Wilson, I.; Gavel, Y.; Von Heijne, G., Amino acid distributions around O-linked glycosylation sites. *Biochem. J.* **1991,** 275, (2), 529-534.
51. Nehrke, K.; Hagen, F. K.; Tabak, L. A., Charge distribution of flanking amino acids influences O-glycan acquisition in vivo. *J. Biol. Chem.* **1996,** 271, (12), 7061-7065.
52. Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R., Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J. Proteome Res.* **2012,** 11, (3), 1949-1957.
53. Ramachandran, P.; Boontheung, P.; Xie, Y.; Sondej, M.; Wong, D. T.; Loo, J. A., Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* **2006,** 5, (6), 1493-1503.
54. Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D., Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **2005,** 4, (6), 2070-2080.
55. Remaley, A.; Wong, A.; Schumacher, U.; Meng, M.; Brewer, H.; Hoeg, J., O-linked glycosylation modifies the association of apolipoprotein A-II to high density lipoproteins. *J. Biol. Chem.* **1993,** 268, (9), 6785-6790.
56. Beisiegel, U.; Weber, W.; Havinga, J. R.; Ihrke, G.; Hui, D. Y.; Wernette-Hammond, M. E.; Turck, C. W.; Innerarity, T. L.; Mahley, R. W., Apolipoprotein E-binding proteins isolated from dog and human liver. *Atertio. Thromb. Vasc. Biol.* **1988,** 8, (3), 288-297.
57. Wada, Y.; Kadoya, M.; Okamoto, N., Mass spectrometry of apolipoprotein C-III, a simple analytical method for mucin-type O-glycosylation and its application to an autosomal recessive cutis laxa type-2 (ARCL2) patient. *Glycobiology* **2012,** 22, (8), 1140-1144.
58. Cubedo, J.; Padró, T.; Badimon, L., Glycoproteome of human apolipoprotein AI: N-and O-glycosylated forms are increased in patients with acute myocardial infarction. *Translational Research* **2014,** 164, (3), 209-222.
59. Woody, S. K.; Zhao, L., Clusterin (APOJ) in Alzheimer's Disease: An Old Molecule with a New Role. In *Update on Dementia*, InTech: 2016.
60. Burkey, B.; Harmony, J., Intracellular processing of apolipoprotein J precursor to the mature heterodimer. *J. Lipid Res.* **1991,** 32, (6), 1039-1048.
61. Liang, H.-C.; Russell, C.; Mitra, V.; Chung, R.; Hye, A.; Bazenet, C.; Lovestone, S.; Pike, I.; Ward, M., Glycosylation of human plasma clusterin yields a novel candidate biomarker of Alzheimer's disease. *J. Proteome Res.* **2015,** 14, (12), 5063-5076.

62. Kapron, J. T.; Hilliard, G. M.; Lakins, J. N.; Tenniswood, M. P.; West, K. A.; Carr, S. A.; Crabb, J. W., Identification and characterization of glycosylation sites in human serum clusterin. *Protein Sci.* **1997,** 6, (10), 2120-2133.
63. Dallas, D. C.; Guerrero, A.; Parker, E. A.; Robinson, R. C.; Gan, J.; German, J. B.; Barile, D.; Lebrilla, C. B., Current peptidomics: applications, purification, identification, quantification, and functional analysis. *Proteomics* **2015,** 15, (5-6), 1026-1038.
64. Bern, M.; Kil, Y. J.; Becker, C., Byonic: advanced peptide and protein identification software. *Current protocols in bioinformatics* **2012**, 13.20. 1-13.20. 14.
65. Secher, A.; Kelstrup, C. D.; Conde-Frieboes, K. W.; Pyke, C.; Raun, K.; Wulff, B. S.; Olsen, J. V., Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nature communications* **2016,** 7.
66. Sossin, W. S.; Fisher, J. M.; Scheller, R. H., Cellular and molecular biology of neuropeptide processing and packaging. *Neuron* **1989,** 2, (5), 1407-1417.
67. Wang, Y.; Wang, M.; Yin, S.; Jang, R.; Wang, J.; Xue, Z.; Xu, T., NeuroPep: a comprehensive resource of neuropeptides. *Database* **2015,** 2015.
68. Wardman, J. H.; Berezniuk, I.; Di, S.; Tasker, J. G.; Fricker, L. D., ProSAAS-derived peptides are colocalized with neuropeptide Y and function as neuropeptides in the regulation of food intake. *PLoS One* **2011,** 6, (12), e28152.
69. Fricker, L. D.; McKinzie, A. A.; Sun, J.; Curran, E.; Qian, Y.; Yan, L.; Patterson, S. D.; Courchesne, P. L.; Richards, B.; Levin, N., Identification and characterization of proSAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing. *J. Neurosci.* **2000,** 20, (2), 639-648.
70. Fricker, L. D. In *Neuropeptides and other bioactive peptides: from discovery to function*, Colloquium Series on Neuropeptides, 2012; Morgan & Claypool Life Sciences: 2012; pp 1-122.
71. Mzhavia, N.; Berman, Y.; Che, F.-Y.; Fricker, L. D.; Devi, L. A., ProSAAS processing in mouse brain and pituitary. *J. Biol. Chem.* **2001,** 276, (9), 6207-6213.
72. Gomes, I.; Bobeck, E. N.; Margolis, E. B.; Gupta, A.; Sierra, S.; Fakira, A. K.; Fujita, W.; Müller, T. D.; Müller, A.; Tschöp, M. H., Identification of GPR83 as the receptor for the neuroendocrine peptide PEN. *Science signaling* **2016,** 9, (425), ra43.
73. Gomes, I.; Aryal, D. K.; Wardman, J. H.; Gupta, A.; Gagnidze, K.; Rodriguiz, R. M.; Kumar, S.; Wetsel, W. C.; Pintar, J. E.; Fricker, L. D., GPR171 is a hypothalamic G protein-coupled receptor for BigLEN, a neuropeptide involved in feeding. *Proceedings of the National Academy of Sciences* **2013,** 110, (40), 16211-16216.
74. Van Rensburg, S.; Berman, P.; Potocnik, F.; Taljaard, J., Glycosylation of transferrin in Alzheimer's disease and alcohol-induced dementia. *Metab. Brain Dis.* **2000,** 15, (4), 243-247.
75. Sáez‐Valero, J.; Fodero, L.; Sjögren, M.; Andreasen, N.; Amici, S.; Gallai, V.; Vanderstichele, H.; Vanmechelen, E.; Parnetti, L.; Blennow, K., Glycosylation of acetylcholinesterase and butyrylcholinesterase changes as a function of the duration of Alzheimer's disease. *J. Neurosci. Res.* **2003,** 72, (4), 520-526.
76. Jacobsen, K. T.; Iverfeldt, K., O-GlcNAcylation increases non-amyloidogenic processing of the amyloid-β precursor protein (APP). *Biochem. Biophys. Res. Commun.* **2011,** 404, (3), 882-886.

77. Liu, F.; Iqbal, K.; Grundke-Iqbal, I.; Hart, G. W.; Gong, C.-X., O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, (29), 10804-10809.
78. Sihlbom, C.; Davidsson, P.; Sjögren, M.; Wahlund, L.-O.; Nilsson, C. L., Structural and quantitative comparison of cerebrospinal fluid glycoproteins in Alzheimer's disease patients and healthy individuals. *Neurochem. Res.* **2008,** 33, (7), 1332-1340.
79. Palmigiano, A.; Barone, R.; Sturiale, L.; Sanfilippo, C.; Bua, R. O.; Romeo, D. A.; Messina, A.; Capuana, M. L.; Maci, T.; Le Pira, F., CSF N-glycoproteomics for early diagnosis in Alzheimer's disease. *J. Proteomics* **2016,** 131, 29-37.
80. Zhu, Y.; Shan, X.; Yuzwa, S. A.; Vocadlo, D. J., The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **2014,** 289, (50), 34472-34481.
81. Schedin‐Weiss, S.; Winblad, B.; Tjernberg, L. O., The role of protein glycosylation in Alzheimer disease. *The FEBS journal* **2014,** 281, (1), 46-62.
82. Butterfield, D. A.; Owen, J. B., Lectin‐affinity chromatography brain glycoproteomics and Alzheimer disease: Insights into protein alterations consistent with the pathology and progression of this dementing disorder. *PROTEOMICS-Clinical Applications* **2011,** 5, (1‐2), 50-56.
83. Kanninen, K.; Goldsteins, G.; Auriola, S.; Alafuzoff, I.; Koistinaho, J., Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach. *Neurosci. Lett.* **2004,** 367, (2), 235-240.
84. Yuzwa, S. A.; Shan, X.; Macauley, M. S.; Clark, T.; Skorobogatko, Y.; Vosseller, K.; Vocadlo, D. J., Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **2012,** 8, (4), 393-399.
85. Li, X.; Lu, F.; Wang, J. Z.; Gong, C. X., Concurrent alterations of O‐GlcNAcylation and phosphorylation of tau in mouse brains during fasting. *Eur. J. Neurosci.* **2006,** 23, (8), 2078-2086.
86. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015,** 15, (9), 540-555.

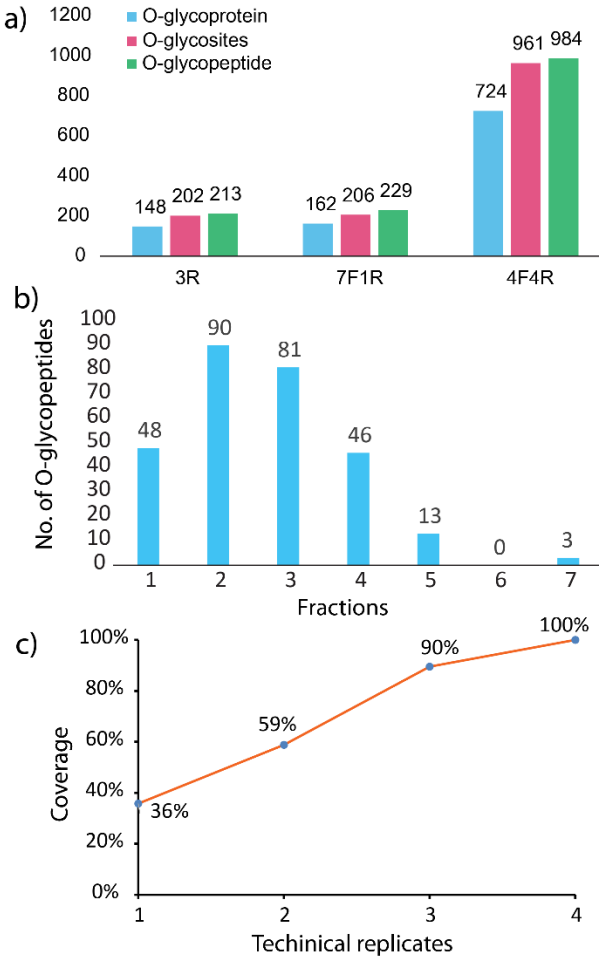


Figure 1. Boronic acid enrichment optimization. a) Optimizing the number of fractions and technical replicates used. b) The effects of technical replicates on the overall coverage. c) The distribution of O-glycopeptides identified in different HpH fractions. (F: fractions, R: technical replicates)

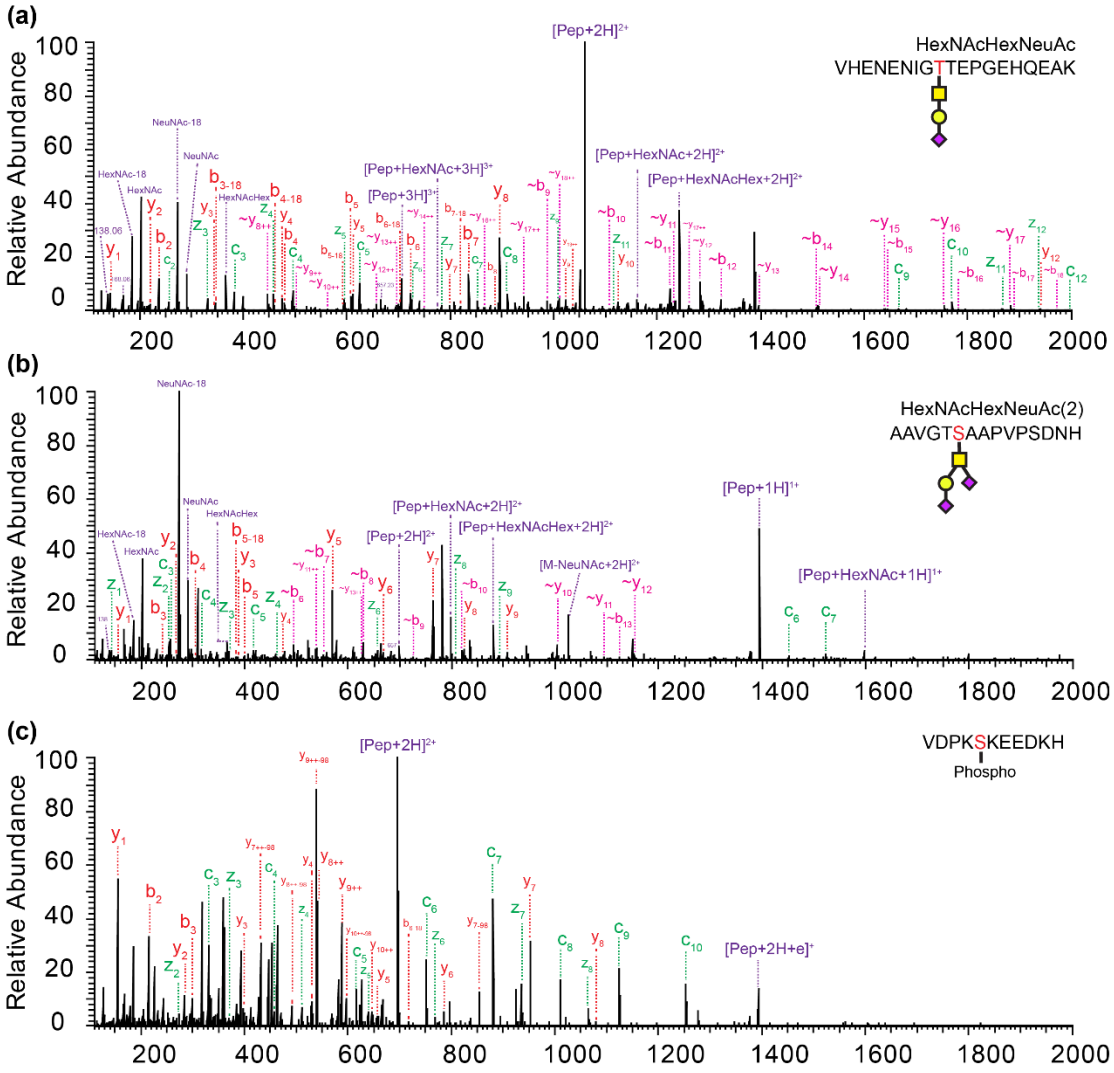


Figure 2. EThcD spectra of O-glycopeptide from O-glycoproteome analysis (a), endogenous O-glycopeptide (b), and endogenous phosphorylated peptide (c).

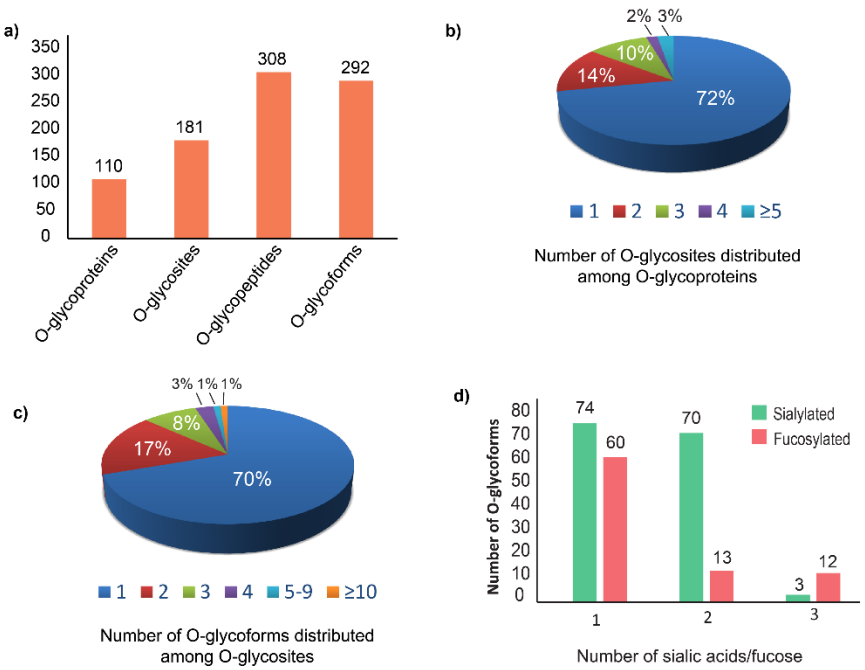


Figure 3. a) The O-glycoproteome mapping in CSF proteins. b) The number of O-glycoforms with 1-3 sialic acids/fucoses. c) The distribution of the number of O-glycosites among O-glycoproteins. d) The distribution of the number of O-glycoforms among O-glycosites.

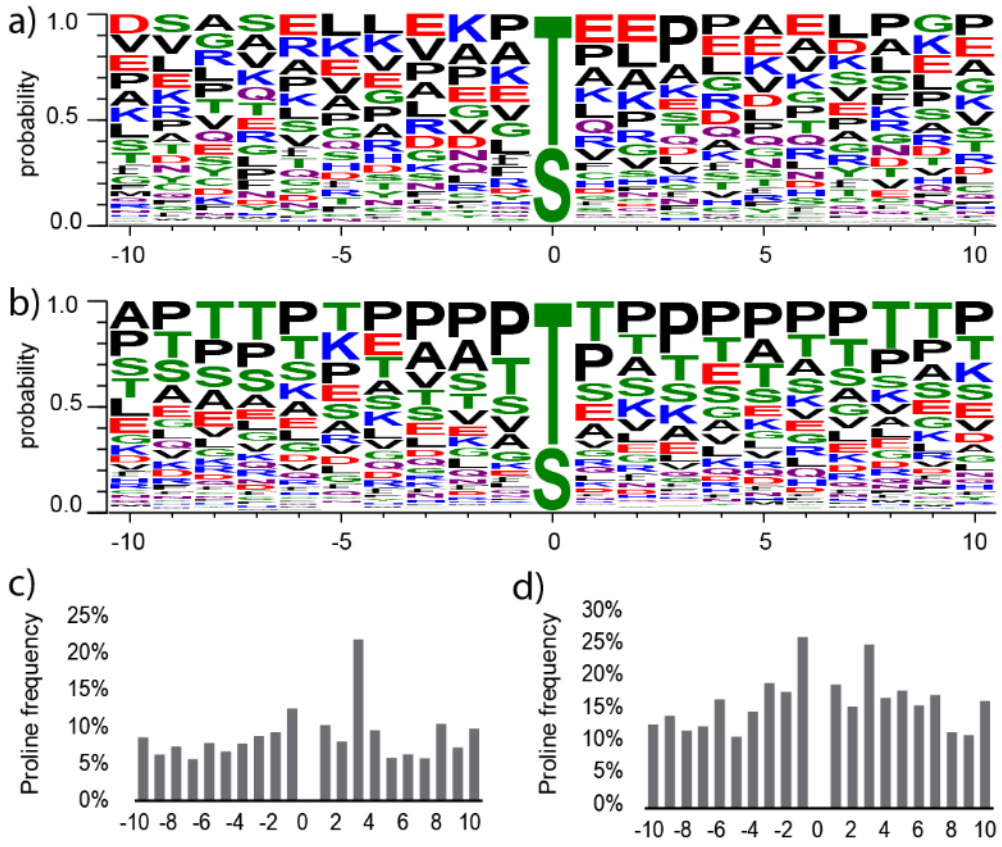


Figure 4. Weblogo probability plots and proline frequency analysis. a) Weblogo plot for 181 identified O-glycosites in the current study. b) Weblogo plot for 435 O-glycosites from Uniprot. c) Proline frequency analysis of 181 identified O-glycosites in the current study. d) Proline frequency analysis of 435 O-glycosites from Uniprot. ±10 amino acid residues surrounding the O-glycosite were analyzed.

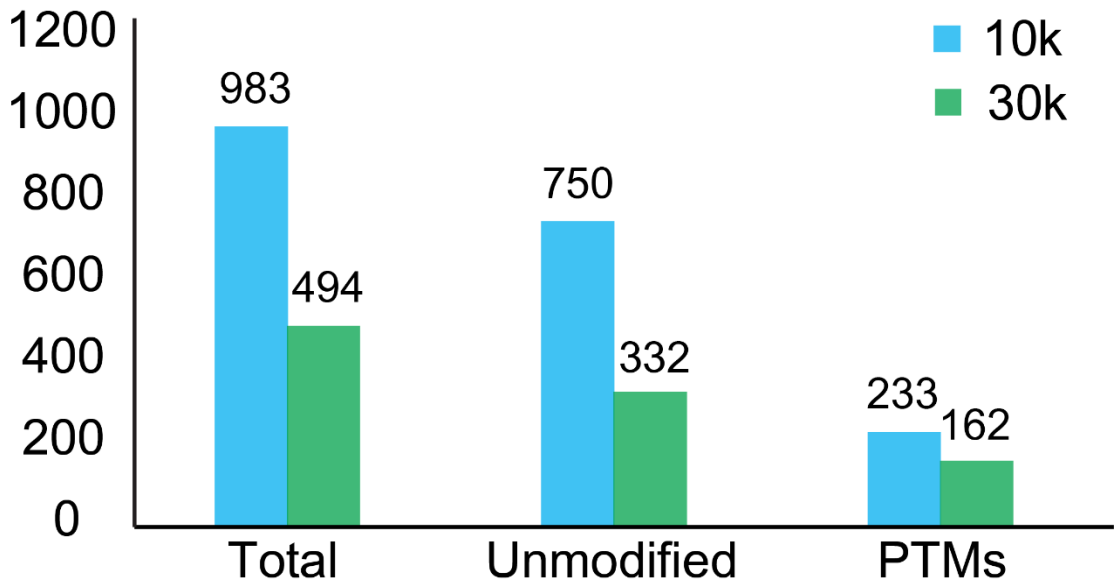


Figure 5. The performance comparison between 10k-based protocol and 30k-based protocol.

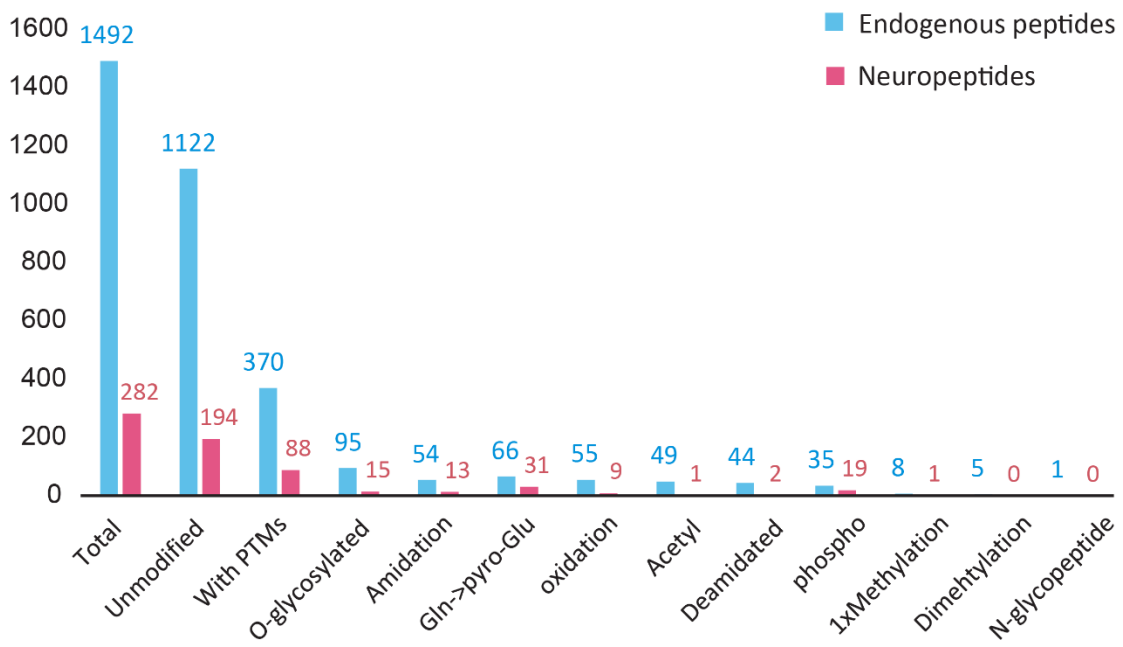


Figure 6. The number of endogenous peptides and neuropeptides identified with/without different PTMs.

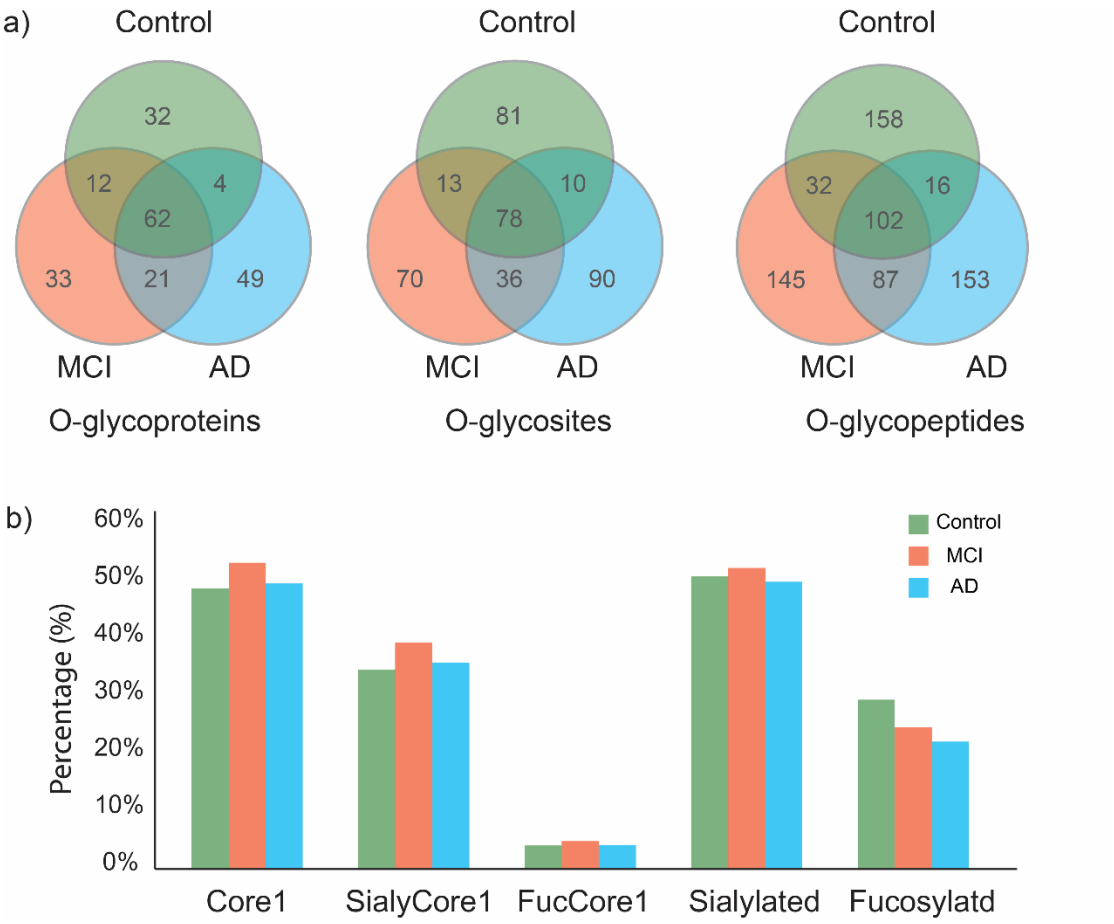


Figure 7. O-glycoproteome comparison between control, MCI and AD. a) Venn diagram analysis of total O-glycoproteome in control, MCI and AD. b) The comparison of different O-glycoforms in control, MCI and AD.

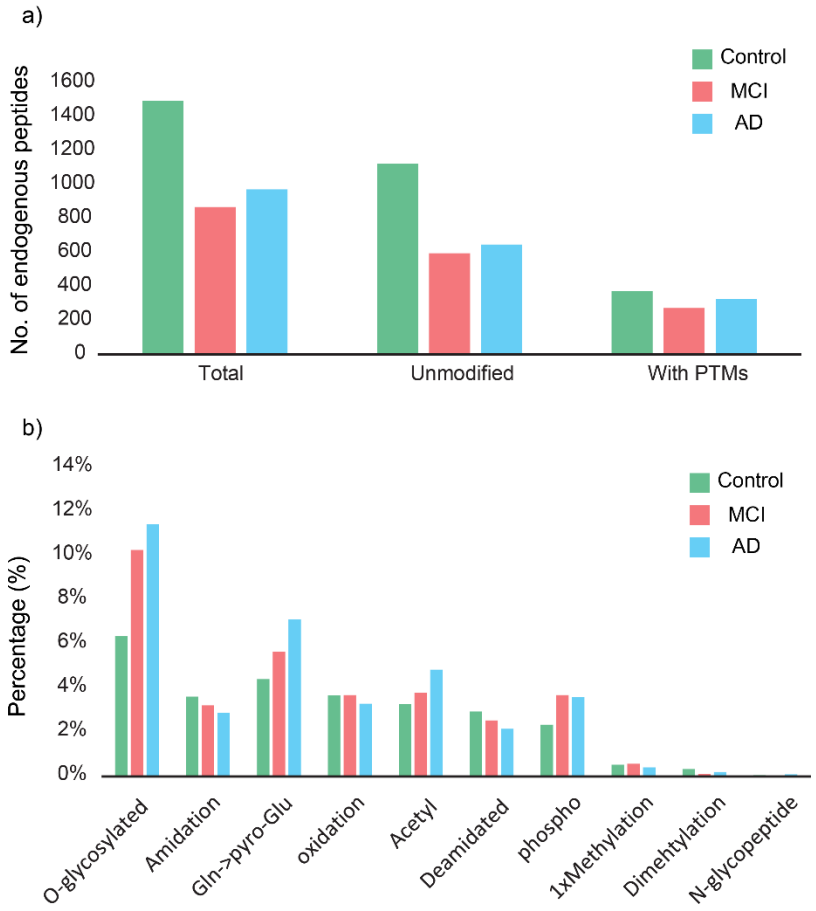


Figure 8. Endogenous peptidome comparison between control, MCI and AD. a) The number of endogenous peptides identified with/without PTMs. B) The comparison of the percentage of endogenous peptides with different PTMs identified in control, MCI and AD.

Table 1 Representative O-glycopeptides identified from human CSF glycoproteins.

O-glycoprotein ID	O-glycosite	O-glycopeptide sequence	O-glycan compositions
Prosaposin receptor GPR37L1	107	[R].GNLTGAPGQR.[L]	HexNAc(2)Hex(2)
	107	[R].GNLTGAPGQR.[L]	HexNAc(2)Hex(2)Fuc(1)NeuAc(1)
Protein YIPF3	339	[R].LPTTVLNATAK.[A]	HexNAc(1)Hex(1)NeuAc(1)
	339	[R].LPTTVLNATAK.[A]	HexNAc(1)Hex(1)NeuAc(2)
Apoliprotein D	86	[K].VLNQELRADGTVNQIEGEATPVNLTEPAK.[L]	HexNAc(2)Hex(2)Fuc(1)
	86	[K].VLNQELRADGTVNQIEGEATPVNLTEPAK.[L]	HexNAc(2)Hex(2)Fuc(1)NeuAc(2)
	86	[R].ADGTVNQIEGEATPVNLTEPAK.[L]	HexNAc(4)Hex(4)Fuc(3)
Apoliprotein E	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(1)Hex(1)NeuAc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(1)Hex(1)NeuAc(2)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)Fuc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)Fuc(1)NeuAc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)NeuAc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)NeuAc(2)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(3)Hex(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(3)Hex(1)Fuc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(3)Hex(1)Fuc(1)NeuAc(1)
	36	[R].QQTEWQSGQR.[W]	HexNAc(1)Hex(1)NeuAc(1)
	36	[R].QQTEWQSGQR.[W]	HexNAc(1)Hex(1)NeuAc(2)
Apoliprotein A-I	92	[R].EQLGPVTQEFWDNLEK.[E]	HexNAc(3)Hex(2)Fuc(1)
Apoliprotein J	105	[K].ELPGVCNETMMALWEECK.[P]	HexNAc(3)Fuc(1)
	210	[R].EPQDTYHYLPFSLPHR.[R]	HexNAc(1)Hex(1)
	210	[R].EPQDTYHYLPFSLPHR.[R]	HexNAc(1)Hex(1)NeuAc(2)
	376	[R].LANLTQGEDQYYLR.[V]	HexNAc(5)Hex(5)Fuc(3)
	376	[R].LANLTQGEDQYYLR.[V]	HexNAc(6)Hex(5)Fuc(3)

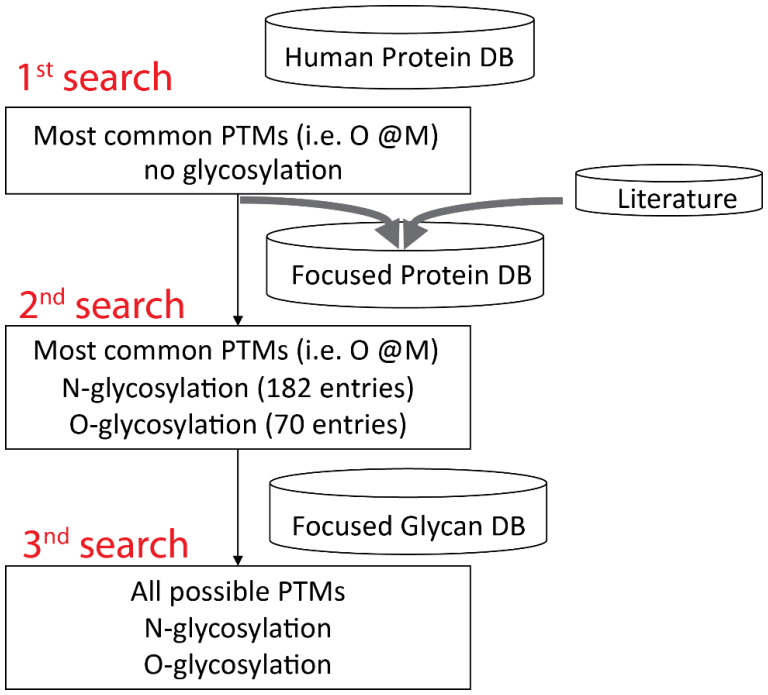


Figure S1. The three-step searching strategy for comprehensive PTMs analysis of CSF endogenous peptides.

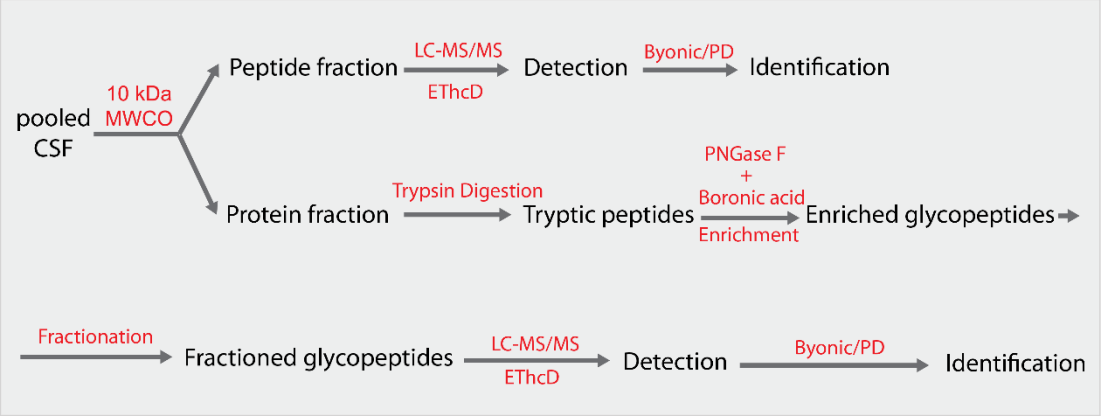


Figure S2. The overall workflow for O-glycoproteome and endogenous peptides analysis.

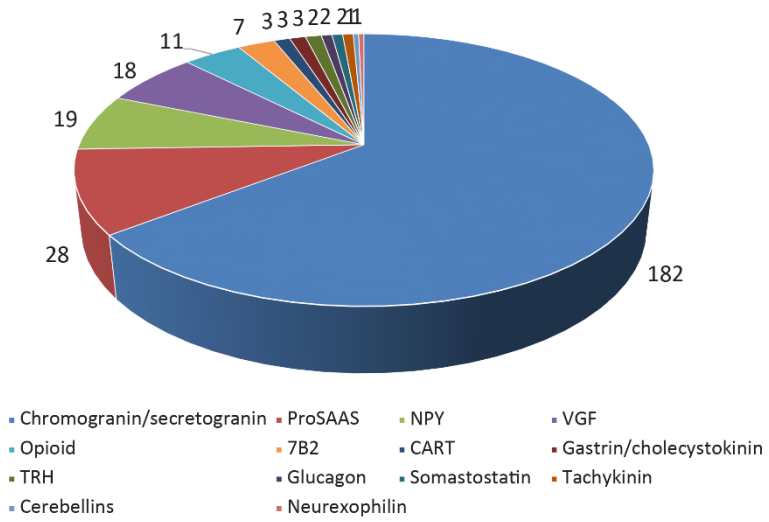


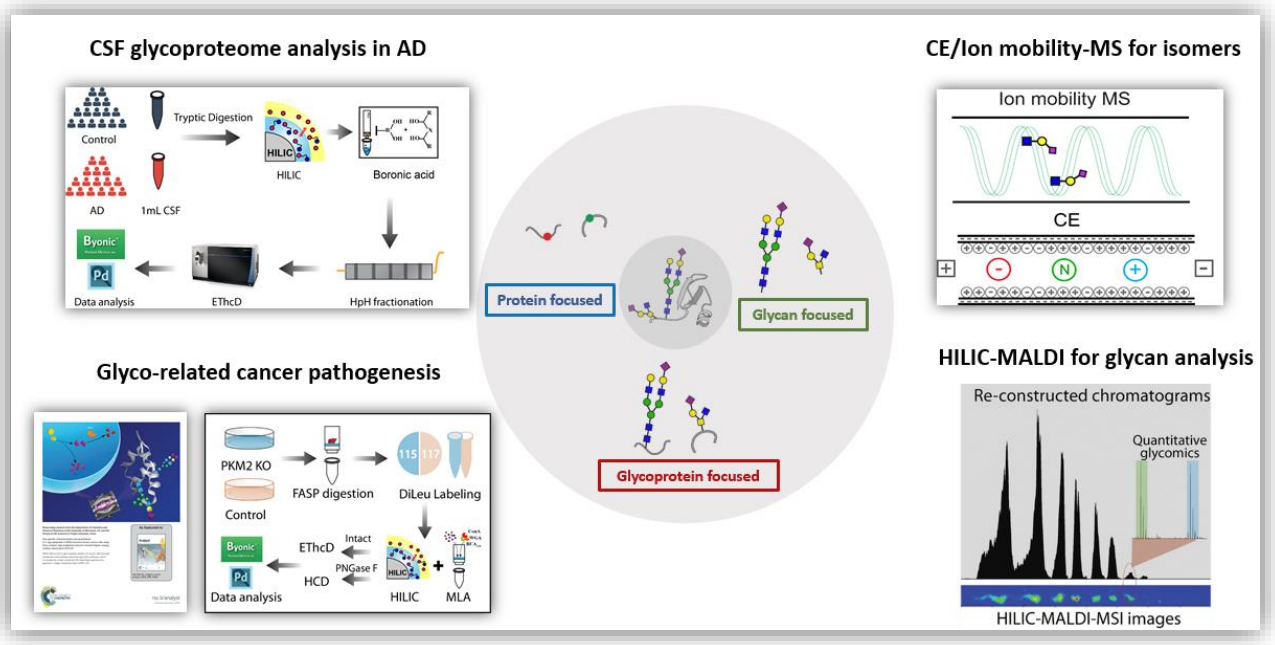
Figure S3. Different families of neuropeptides in CSF.

Supplemental Table S1 Subjects information from healthy control, MCI and AD group.

Groups	No. of Subjects	Male, No. (%)	Female, No. (%)	Age, mean (SD)	APOE4 positive, No. (%)
Normal Control	16	8 (50%)	8 (50%)	73.0 (5.3)	3 (19%)
MCI	16	12 (75%)	4 (25%)	75.6 (7.7)	9 (56%)
AD	16	8 (50%)	8 (50%)	73.2 (8.6)	10 (63%)

Chapter 9

Conclusions and future directions



Conclusions

In this dissertation, novel methodologies including separation techniques such as RPLC, HILIC, CE and IM, sample preparation techniques such as glycoprotein extraction and glycopeptide enrichment, and also MALDI and ESI-based quantitative MS strategies were developed to characterize and quantify protein glycosylation both at released glycans and intact glycopeptide level. Such novel methodologies and workflow were successfully applied to analyze complex biological samples to answer glycosylation-related biological questions in disease progression. This work not only provides powerful MS-based glycomics and glycoproteomics tools for protein glycosylation study for the glycoscience community, but also generates a wealth

of site-specific glycosylation data to better understand the role of glycosylation in the pathological processes of disease.

Traditional MS-based strategies struggle to distinguish the large number of coexisting isomeric glycans that are indistinguishable by mass alone. Ion mobility spectrometry coupled to MS (IM–MS) has received considerable attention as an analytical tool for improving glycan characterization due to the capability of IM to resolve isomeric glycans before MS measurements. In Chapter 2, I reviewed recent applications of IM–MS that illustrate the enormous potential of this technology in a variety of research areas, including glycomics, glycoproteomics, and glycobiology.

Mass spectrometry (MS)-based glycoproteomics is a powerful approach that provides the ability for large-scale analysis protein glycosylation in complex bioloigcal sample. Significant advances have been achieved including novel glycopeptide enrichment, hybrid fragmentation techniques, maturing softwares, and effective quantitation strategies. In Chapter 3, I reviewed the recent advances in MS-based glycoproteomics and their application to glycosylatio pattern changes in various diseases. This review not only serves as an introduction to the basic knowledge of MS-based glycoproteomics, but also offers perspectives and suggestions for the future direction of this field.

As a complementary ionization method to ESI, MALDI generates only singly charged ions which simplifies the spectrum and the signal of a specific analyte is “focused” and “enhanced” on these singly charged ions. As MALDI is more tolerant to salts and other contaminants when proper matrix or matrix additives is applied, less sample cleanup is needed, thus increasing sample recovery rate and decreasing sample preparation time. However, MALDI-MS analyses of complex

glycan samples are often hampered by the complexity of the sample due to ion suppression effect, leading to low glycan coverage, poor MS/MS fragmentation quality, and inaccurate quantitation. To this end, Chapter 4 focused on solving these problems by introducing HILIC as a front-end separation and mass spectrometric imaging (MSI) to thoroughly sample all the analytes on the MALDI plate. I demonstrated the capability of the newly developed HILIC-MALDI-MSI platform by analyzing the N-glycans released from human serum, providing a 10-fold intensity improvement, increased N-glycan quantitation accuracy and coverage compared to conventional MALDI-MS. With various effective approaches being developed such as MALDI-MS, LC-MALDI-MS, and LC-ESI-MS for quantitative glycomics study, our current developed approach enriches the toolbox by pushing MALDI-based quantitative glycomics one step forward.

Many of the monosaccharides that compose larger glycans are structural isomers, and they can be connected via either α- or β-stereochemistry at multiple linkage positions, resulting in many glycan isomers. In Chapter 5, by taking advantage of the TWIM capability of the Waters Synapt G2 Q-TOF instrument, we were able to further improve the resolution of aminoxyTMT-labeled isomeric HMO standards by integrating a second separation dimension, the ion mobility separation in the gas phase, with the first dimension CE separation. In the IM drift time profile, aminoxyTMT labeled LSTc could be differentiated from those of LSTa and LSTb, while LSTa and LSTb exhibited the same drift time. In the CE electropherogram, LSTa and LSTb were successfully separated, but LSTb was co-eluted with LSTc. After on-line coupling CE with IM, we showed that aminoxyTMT labeled LSTa, LSTb and LSTc could be base-line separated, demonstrating the separation complementary capability between CE and IM. Benefiting from the successful separation of these three isomers, we were able to achieve accurate quantitation of each isomer across different samples.

Besides the glycomics approach, another strategy for MS-based glycosylation study is glycoproteomics approach. Previous glycoproteomics studies were somewhat biased towards glycosylation site mapping or released glycan analysis mainly due to a multitude of structural complexity encompassing attached glycans and a lack of enabling analytical technology. In order to achieve quantitative glycosylation analysis in a site-specific manner, Chapter 6 focused on developing a powerful workflow including improved specific extraction of membrane-bound glycoproteins using the filter aided sample preparation (FASP) method, enhanced enrichment of N-glycopeptides using sequential hydrophilic interaction liquid chromatography (HILIC) and multi-lectin affinity (MLA) enrichment, site-specific N-glycopeptide characterization enabled by EThcD, relative quantitation utilizing isobaric N,N-dimethyl leucine (DiLeu) tags and automated FDR-based large-scale data analysis by Byonic. To demonstrate the effectiveness of this workflow, we applied the workflow to study the glycosylation alteration in PKM2 knockout breast cancer cells vs. parental cells. Upon loss of PKM2, the abundance ratios of different glycoforms on the same glycosylation site vary differently and increased fucosylation was observed in several of the examined glycosylation sites. Further deglycoproteomics studies revealed that the 10 glycoproteins in the PI3K/Akt signaling pathway were altered, which supported the previous finding that PKM2 knockdown cancer cells rely on the activation of Akt for their survival.

As the only body fluid that directly interchanges with the extracellular fluid of CNS, CSF reflects the ongoing pathological changes in the CNS most directly. Thus, biochemical analysis of CSF has great potential for CNS-related diseases diagnostics, such as neurological disorders. Initial efforts into site-specific analysis of intact glycopeptides in CSF have been made, but the depth reached is not satisfactory, with one study showing the identification of 36 N-glycosylation sites from 23 N-glycoproteins [1], and another study showing the identification of 55 N-

glycosylation sites from 36 N-glycoproteins [2]. With such limited site-specific N-glycoproteome information, the process of uncovering potential roles of different glycoproteins in the CNS is hampered, and it will also hinder the design of studies to explore disease-related glycosylation alterations. Therefore, Chapter 7 focused on developing an enhanced dedicated large-scale site-specific glycoproteomics approach for in-depth CSF N-glycoproteome analysis, including sequential HILIC and boronic acid enrichment for improved N-glycopeptide coverage, intact N-glycopeptide characterization enabled by EThcD and automated FDR-based large-scale data analysis by Byonic. In total, 3596 intact N-glycopeptides from 676 N-glycosylation sites and 358 N-glycoproteins were identified in CSF, representing the largest reported site-specific CSF N-glycoproteome dataset to date.

Site-specific O-glycoproteome mapping in complex biological system provides molecular basis for understanding the structure-function relationships of glycoproteins and their roles in physiological and pathological processes. Previous O-glycoproteome analysis in CSF focused on sialylated glycoforms, and many CSF glycoproteins have not been characterized comprehensively with respect to their O-glycosylation. In order to provide an unbiased O-glycosylation profiling, Chapter 8 describes the development of an integrated strategy combining universal boronic acid enrichment, high-pH fractionation, and EThcD for improved intact O-glycopeptide analysis. This strategy was applied to analyze O-glycoproteome in CSF, leading to identifications of 308 O-glycopeptides from 110 O-glycoproteins, covering both sialylated and non-sialylated glycoforms. To our knowledge, this is the largest number of O-glycoproteins and O-glycosites reported for CSF so far, including 154 novel O-glycosites.

Alterations in protein glycosylation have been implicated in human neurodegenerative diseases, such as Alzheimer’s disease (AD), Parkinson disease and Creutzfeldt-Jakob disease.

Abnormal glycosylation patterns of amyloid precursor protein (APP), tau and numerous other proteins have been reported in AD. It has also been shown that O-glycosylation protects tau against aberrant phosphorylation and subsequent aggregation. Thus, comparative glycoproteomics were conducted in Chapters 7-8 to examine the glycosylation alterations in AD patients. A similar overall N-glycopeptide/O-glycopeptide coverage in AD patients and controls was obtained but striking differences in certain glycoforms were evident, particularly decreased fucosylation in AD CSF. Altered glycosylation patterns were detected for a number of N-glycoproteins including alpha-1-antichymotrypsin, Ephrin-A3, carnosinase CN1, and voltage-dependent T-type calcium channel subunit alpha-1H etc., which serve as interesting targets for further glycosylation-based AD study and may eventually contribute to improved understanding of the role of glycosylation in AD progression.

For endogenous peptidome study in CSF, one study identified 730 endogenous peptides, including 138 peptides with PTMs such as acetylation, amidation, phosphorylation, Gln to pyro-Glu conversion.[3] But none of the identified peptides is reported to be glycosylated. In another study with 563 endogenous peptides identified,[4] Zougman et al. noted the presence of glycan oxonium ions in some spectra, indicating the existence of glycosylated endogenous peptides in CSF. By lowering the HCD collision energy, labile glycan moiety could be partially preserved, and 28 O-glycopeptides were successfully identified. However, only two O-glycan compositions were found. Considering that the two O-glycan modifications set in the search engine were found by merely incidental observation and the large diversity of O-glycans known in existence, it is highly possible that there are more O-glycosylated endogenous peptides in CSF, and perhaps some N-glycosylated endogenous peptides are present in CSF as well. Therefore, a systemic approach needs to be developed, which features a glycosylation-centered analytical strategy and searching

strategy to take vast categories of N-/O-glycans into account. In Chapter 8, we developed a novel peptidomics workflow that combined CSF endogenous peptide extraction by 10k molecular weight cut-off (MWCO), EThcD fragmentation, and a three-step database searching strategy for comprehensive PTM analysis. This developed workflow enables a comprehensive PTM analysis of endogenous peptides, including both N-/O-glycosylation, phosphorylation, amidation, acetylation, Gln to pyro-Glu conversion. In total, 1492 endogenous peptides were identified, with 370 of them carrying post-translational modifications including O-glycosylation, phosphorylation, acetylation, amidation, Gln to pyro-Glu conversion. Among them, 95 endogenous peptides were O-glycosylated, and 15 of them were O-glycosylated neuropeptides, which serves as the first report on the discovery of endogenous O-glycosylated peptides and neuropeptides in human CSF.

Future directions

With various effective approaches being developed such as MALDI-MS, LC-MALDI-MS, and LC-ESI-MS for quantitative glycomics studies, our current developed approach enriches the toolbox by pushing MALDI-based quantitative glycomics one step forward via the incorporation of mass spectrometric imaging to thoroughly sample the analytes on a typical MALDI plate. Benefiting from the almost 10-fold intensity improvement, increased N-glycan quantitation accuracy and coverage using the HILIC-MALDI-MSI platform has been demonstrated compared to conventional MALDI-MS in the present study, but still there are many facets of this platform worth being further explored. For example, how will it perform for the characterization and quantitation of glycan isomers when a porous graphitic carbon (PGC) column is used? To what extent can the glycan separation, especially for glycan isomer, benefit from using high concentration of volatile salts or even non-volatile salts such as ion pairing agent? Moreover, a systematic performance comparison with LC-ESI will be quite interesting to see how they

complement with each other, in terms of glycan separation efficiency, glycan coverage, glycan isomer characterization, and required efforts in sample preparation. Future work include additional in-depth evaluations of these aspects to help further promote the HILIC-MALDI-MSI platform as a viable alternative that can be included in the rapidly expanding quantitative glycomics toolbox.

The current exploratory glycosylation-based biomarker study in AD utilizing MS-based glycoproteomics approach focuses on in-depth mapping the representative and averaged glycosylation landscape of CSF proteins in healthy control and AD. Such a comparison will shed light upon the overall average, dominant glycosylation difference and similarities, and also some of the interesting glycoprotein candidates with specific glycosylation pattern alterations in AD. Still, such list is merely a preliminary exploration and further investigations are needed to narrow down or provide a more complete list of potentially interesting glycosylation-based biomarker candidates in AD. Future studies include conducting high-throughput quantitative studies using 12-plex DiLeu isobaric tags developed in our lab, which could allow us to take the individual patient-to-patient variation into account and validate the results described here. Besides, we found a gender difference for the glycoproteome analysis in the previous study, so subjects from healthy control and AD group will be divided into groups based on their gender.

Future work on developing a more robust spinnable and automated hydrophilic interaction chromatography-stationary phase extraction (HILIC-SPE) tips based enrichment method for N-glycoproteomics study is also ongoing. The current data we acquired shows that such method provides a highly-efficient, fast and cost-effective method for N-glycopeptide enrichment. Different HILIC materials including SAX, neutral and ZIC will be evaluated in terms of their ability to enrich DiLeu labeled N-glycopeptides. And also we will continue to develop more robust O-glycopeptide enrichment method based on boronic acid or SAX HILIC enrichment for

quantitative analysis of O-glycopeptides. The glycoform changes captured by using the pooled samples from 16 subjects from each stage will further be validated in the quantitative studies based on the individual samples, which would also provide abundance changes information. After summarizing the interesting biomarker candidates from this study and correlating these candidates with previous reports on their association with AD or neurodegeneration, a compiled list could be generated as potential targets for the following-up validation study. The following targeted studies exploring the glycosylation function of these markers both *in vitro* and *in vivo* will be conducted to help us better understand their role in AD progression.

References

1. Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larson, G., Enrichment of glycopeptides for glycan structure and attachment site identification. *Nature methods* **2009,** 6, (11), 809.
2. Goyallon, A.; Cholet, S.; Chapelle, M.; Junot, C.; Fenaille, F., Evaluation of a combined glycomics and glycoproteomics approach for studying the major glycoproteins present in biofluids: Application to cerebrospinal fluid. *Rapid Communications in Mass Spectrometry* **2015,** 29, (6), 461-473.
3. Höltä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. *PloS one* **2012,** 7, (8), e42555.
4. Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kumar, C.; Mann, M., Integrated analysis of the cerebrospinal fluid peptidome and proteome. *Journal of proteome research* **2007,** 7, (01), 386-399.

Appendix I

List of publications and presentations

Publications

1. **Chen, Z.**, Yu, Q., Johnson, J., Shipman, R., Zhong X., Asthana, S., Carlsson, C., Okonkwo, O., Li, L, **2018.** In-depth site-specific O-glycosylation analysis of glycoproteins and endogenous peptides in cerebrospinal fluid (CSF) from healthy individual, mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients. *Journal of Proteomics Research.* (To be submitted).
2. **Chen, Z.**, Yu, Q., Johnson, J., Shipman, R., Zhong X., Asthana, S., Carlsson, C., Okonkwo, O., Li, L, **2018.** In-depth site-specific analysis of N-glycoproteome in human cerebrospinal fluid (CSF) and glycosylation landscape changes in Alzheimer's disease (AD). *Molecular Cellular Proteomics.* (To be submitted).
3. **Chen, Z.**, * Huang, J., * Li, L. Recent advances in mass spectrometry (MS)-based glycoproteomics in complex biological samples. *Trends in Anal. Chem.*(Submitted)
4. **Chen, Z.**,* Yu, Q.,* Hao, L., Liu, F, Johnson, J., Tian, Z., Kao, J., Xu, W., Li, L, **2017.** Site-specific characterization and quantitation of N-glycoproteins in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *Analyst.* DOI: 10.1039/C8AN00216A
5. **Chen, Z.**,* Glover, M.,* Li, L, **2017.** Recent Advances in Ion Mobility-Mass Spectrometry for Improved Structural Characterization of Glycans and Glycoconjugates. *Current Opinion in Chemical Biology. 42, pp.1-8.*
6. **Chen, Z.**, Zhong, X., Tie, C., Chen, B., Zhang, X. and Li, L., **2017**. Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents. *Analytical and Bioanalytical Chemistry*, 409(18), pp. 4437–4447.
7. **Chen, Z**., Shi, Y., Huang, J., Johnson, J., Miesbauer, E., Zhogn, X., Yu, Q., Li, L., **2018**. A spinnable and automatable hydrophilic interaction chromatography-stationary phase extraction (HILIC-SPE) tips based enrichment method for N-glycoproteomics study. (In preparation)
8. Glover, M., Yu, Q., **Chen, Z.**, Shi X., Kent, C., and Li, L, **2017.** Characterization of Intact Sialylated Glycopeptides and Phosphorylated Glycopeptides from IMAC enriched samples by EThcD Fragmentation: Toward Combining Phosphoproteomics and Glycoproteomics. *The International Journal of Mass Spectrometry.* 427, pp. 35-42
9. Yu, Q., Wang, B., **Chen, Z.**, Urabe, G., Glover, M.S., Shi, X., Guo, L.W., Kent, K.C. and Li, L., **2017.** Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *Journal of The American Society for Mass*

Spectrometry, pp.1-14.

10. Pang, X., Jia, C., **Chen, Z.** and Li, L., **2017.** Structural Characterization of Monomers and Oligomers of D-Amino Acid-Containing Peptides Using T-Wave Ion Mobility Mass Spectrometry. *Journal of The American Society for Mass Spectrometry*, *28*(1), pp.110-118.
11. Lietz, C.B., **Chen, Z.**, Son, C.Y., Pang, X., Cui, Q. and Li, L., **2016.** Multiple gas-phase conformations of proline-containing peptides: is it always cis/trans isomerization?. *Analyst*, 141(16), pp.4863-4869.
12. Zhong, X., **Chen, Z.**, Snovida, S., Liu, Y., Rogers, J.C. and Li, L., **2015.** Capillary electrophoresis-electrospray ionization-mass spectrometry for quantitative analysis of glycans labeled with multiplex carbonyl-reactive tandem mass tags. *Analytical chemistry*, 87(13), pp.6527-6534.
13. Zhong, X., Yu, Q., Ma, F., Frost, D., Lu, L., **Chen, Z**, Zetterberg, H., Carlsson, C., Okonkwo, O., and Li, L, **2018**. A multiplexed absolute quantification strategy for biomarker candidates verification in preclinical Alzheimer's disease. (Under review)
14. Zhu, J, **Chen**, **Z**, Zhang, J, An, M, Wu, J, Bern, M, and Sen, I, Weatherly, B, Skilton, J, and Lubman, **2018.** D. Quantitative LC-EThcD-MS/MS determination of intact *N*-glycopeptides in serum haptoglobin between heptocellular carcinoma and liver cirrhosis. (Under review)
15. Huang, J., Dong, J., Shi, X., **Chen, Z**, Cui Y., Liu X., Ye, M., and Li, L, **2018.** Enrichment and separation of phosphopeptides and mannose-6-phosphate glycopeptides by Ti(IV)-IMAC in a typical HILIC-mode elution. (To be submitted)
16. Cao, Q., Yu, Q, Liu, Y, **Chen, Z**, and Li, L, **2018**. Discovery of N-linked and O-linked Glycosylation in Crustacean Nervous System. (In preparation)

Presentations:

1. **Chen, Z.**, Yu, Q., Johnson, J., Shipman, R., Zhong X., Asthana, S., Carlsson, C., Okonkwo, O., Li, L, **2018.** Glycosylation-based cerebrospinal fluid (CSF) biomarker discovery in Alzheimer's disease (AD). Alzheimer's Disease & Related Disorders Research Day, March 23th, 2018, Madison, WI. (poster presentation)
2. **Chen, Z.**, Yu, Q., Johnson, J., Shipman, R., Zhong X., Asthana, S., Carlsson, C., Okonkwo, O., Li, L, **2018.** In-depth site-specific analysis of glycoproteome in human cerebrospinal fluid (CSF) and glycosylation alterations in Alzheimer's disease (AD). 66th American Society for Mass Spectrometry (ASMS) Annual Conference, June 3th, 2018, San Diego, CA. (poster presentation)
3. **Chen, Z.**,* Yu, Q.,* Hao, L., Liu, F., Johnson, J., Tian, Z., Kao, J., Xu, W., Li, L, **2017.** "Site-specific characterization and quantitation of N-glycoproteins using DiLeu isobaric tags enabled by Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)" *13th Annual Midwest Carbohydrate and Glycobiology Symposium,* Septemper 22th, 2017, Madison, WI. (poster presentation, outstanding presentation award 100 $)
4. **Chen, Z.**,* Yu, Q.,* Hao, L., Liu, F., Johnson, J., Tian, Z., Kao, J., Xu, W., Li, L, **2017.** "Site-specific characterization and quantitation of N-glycoproteins using DiLeu isobaric tags enabled by Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)" *65th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 5th, 2017, Indianapolis, IN. (poster presentation)
5. **Chen, Z.**,* Yu, Q.,* Hao, L., Liu, F., Johnson, J., Tian, Z., Kao, J., Xu, W., Li, L, **2017.** "Site-specific characterization and quantitation of N-glycoproteins using DiLeu isobaric tags enabled by Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)" *Pharmaceutical Sciences POSTER SESSION,* Feb. 9th, 2018, Madison, WI.
6. **Chen, Z.**, Zhong, X., Tie, C., Shi, Y., Zhang, X., Li, L., **2016.** "Hydrophilic interaction liquid chromatography-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents." *64th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 7th, 2016, San Antonio, TX. (poster presentation)
7. **Chen, Z.**, Zhong, X., Tie, C., Shi, Y., Zhang, X., Li, L., **2016.** "Hydrophilic interaction liquid chromatography-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents." *PharmSci Division Retreat Poster Session,* Aug. 18th, 2016, Madison, WI. (poster presentation)
8. **Chen, Z.**, Lietz, C., Son, C., Cui, Q., Li, L, **2015.** "Evidence for Differential Structural Preferences of the Leu7Pro Mutant Neuropeptide Y Signal Peptide Probed by Ion Mobility-

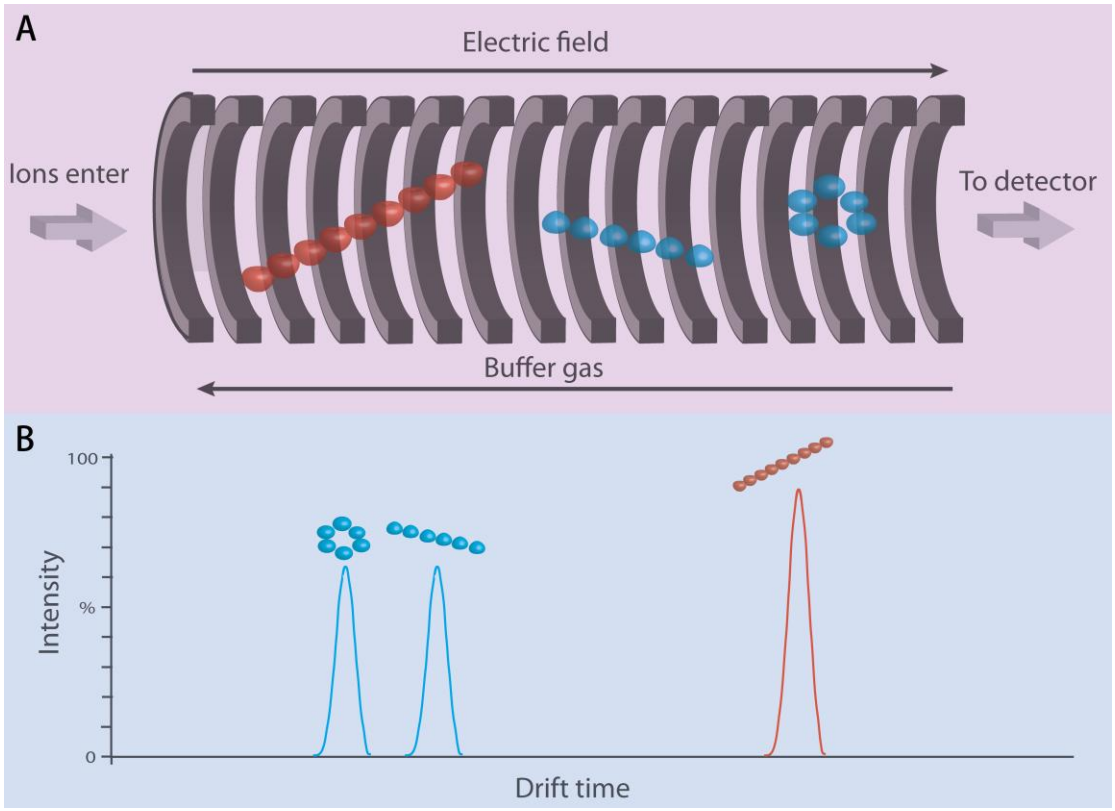
Mass Spectrometry." *63rd American Society for Mass Spectrometry (ASMS) Annual Conference, June 1st, 2015, St. Louis, MO*. (poster presentation)

9. Cui, Y., Dieterich , I., Rhoads, T., **Chen, Z.**, Anderson, R., Puglielli, L., Li, L. **2018.** Isobaric Labeling-based Quantitative Studies of Protein Expression and N-glycosylation of AT-1 sTg Mouse Model Reveal Molecular Basis of Aging. 66th American Society for Mass Spectrometry (ASMS) Annual Conference, June 3th, 2018, San Diego, CA. (poster presentation)
10. Zhu, J, **Chen**, **Z**, Zhang, J, An, M, Wu, J, Bern, M, and Sen, I, Weatherly, B, Skilton, J, and Lubman, **2018.** D. Quantitative LC-EThcD-MS/MS determination of intact *N*-glycopeptides in serum haptoglobin between heptocellular carcinoma and liver cirrhosis. *66th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 3th, 2018, San Diego, CA. (poster presentation)
11. Cao, Q., Yu, Q, Liu, Y, **Chen, Z**, and Li, L, **2018**. Discovery of N-linked and O-linked Glycosylation in Crustacean Nervous System, *66th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 3th, 2018, San Diego, CA. (poster presentation)
12. Zhong, X., Yu, Q., Ma, F., Frost, D., Lu, L., **Chen, Z**, Zetterberg, H., Carlsson, C., Okonkwo, O., and Li, L, **2018**. A multiplexed absolute quantification strategy for biomarker candidates verification in preclinical Alzheimer's disease, *66th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 3th, 2018, San Diego, CA. (poster presentation)
13. Huang, J., Dong, J., Shi, X., **Chen, Z**, Cui Y., Liu X., Ye, M., and Li, L, **2018.** Enrichment and separation of phosphopeptides and mannose-6-phosphate glycopeptides by Ti(IV)-IMAC in a typical HILIC-mode elution. *66th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 3th, 2018, San Diego, CA. (poster presentation)
14. Glover, M., Yu, Q., **Chen, Z.**, Shi X., Kent, C., and Li, L, **2017.** Development of combined phosphoproteomic and glycoproteomic workflows using EThcD fragmentation. *65th American Society for Mass Spectrometry (ASMS) Annual Conference*, June 5th, 2017, Indianapolis, IN. (Oral presentation)
15. Cao, Q., Yu, Q., Liu, Y., **Chen, Z.**, Li, L., **2017.** "Discovery of N-linked and O-linked Glycosylation in Crustacean Nervous System" 13th Annual Midwest Carbohydrate and Glycobiology Symposium, Septemper 22th, 2017, Madison, WI. (poster presentation)
16. Shi, Y., **Chen, Z.**, Yu, Q., Wang, B., Glover, M., Shi, X., Guo, L., Kent, C., Li, L., **2017.** Large-scale characterization and quantitation of citrullinated proteins involved in restenosis by hcd product ion triggered EThcD mass spectrometry. *65th American Society for Mass Spectrometry (ASMS) Annual Conference*, June 5th, 2017, Indianapolis, IN. (poster presentation)

17. Tian, Z., **Chen, Z.**, Glover, M., Cao, Q., Li, L., **2017.** Large scale collision cross section (CCS) profiling of endogenous neuropeptides by ion mobility mass spectrometry. *65th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 5th, 2017, Indianapolis, IN. (poster presentation)
18. Pang, X., Jia, C., Baird, M.; **Chen, Z.**, Shvartsburg, A., Li, L., **2016.** Structural characterization of monomers and oligomers of D-amino acid containing peptides using linear and nonlinear ion mobility separations. *64th American Society for Mass Spectrometry (ASMS) Annual Conference,* June 7th, 2016, San Antonio, TX. (Oral presentation)
19. Jia, C., Pang, X., **Chen, Z.**, Li, L., **2015.** Probing the D-amino acid-containing peptide oligomers by ion mobility spectrometry and native mass spectrometry. *31th Asilomar Conference on Mass Spectrometry.* October 16th,2016, Pacific Grove, CA. (poster presentation)
20. Zhong X., **Chen, Z.**, Snovida, S., Rogers J., Li, L., **2014.** NanoHILIC-ESI-MS and CE-ESI-MS for quantitative glycan analysis using multiplexed carbonyl-reactive tandem mass tags, *30th Asilomar Conference on Mass Spectrometry.* October 10th, 2015, Pacific Grove, CA. (poster presentation)

Appendix II

Review on structure analysis of neuropeptides with Ion Mobility Mass Spectrometry



When combined with molecular dynamics (MD) simulations, ion mobility-mass spectrometry (IM-MS) is a powerful tool to provide gas-phase peptide ion structural insights at atomic level. Analyte structure is determined from temperature-dependent rotationally averaged experimental collision cross sections (CCS) values, which reflects the gas-phase ion conformations originating from solution-phase after desolvation. [1] Compared to other biophysical techniques such as X-ray crystallography or NMR spectroscopy, IM-MS can be used to ascertain structural information using much lower purity and trace amount of sample due to its ability to select a specific ion of interest and high sensitivity. [2] Moreover, unlike other biophysical techniques that provide an averaged structure, IM-MS can be used to interrogate dynamic heterogeneity, which allows snapshots of short-lived intermediates and conformational transitional states to be obtained. [3, 4] In fact, quite a few studies [5-9] have been conducted to explore analyte ion gas phase structure and conformational dynamics, which provides important insights into conformational dynamics that occur in solution.

Owning to their biological significance, neuropeptides have attracted much attention and are extensively studied on their conformation dynamic nature which is most biologically relevant aspect in defining their biological function. Among them, bradykinin, a nine residue neuropeptide, stands out as a model peptide both for its conformational dynamics studies [10-15] and development of IM-MS assisted by MD approach for structures study in general. Previous studies [16-18] revealed that the region Ser^{6}-Pro^{7}-Phe^{8} -Arg^{9} favors a β-turn motif, while Arg^{1}-Pro^{2}-Pro^{3}-Gly^{4}-Phe^{5} is a highly flexible random coil that results in a large degree of conformational flexibility and doesn't adopt a single preferred stable conformation in solution. Clemmer group extensively utilized IM-

MS assisted by MD approach to characterize bradykinin conformational dynamics, revealing there exists 10 independent populations of structures in solution and 3 gas-phase quasi-equilibrium conformations. [10-12, 15] Investigations into these multiple conformations revealed that they are associated with different combinations of cis and trans forms of the three proline residues. [12] Inspired by the fact that penultimate proline residues are frequently found in neuropeptides, Glover and co-workers [19] utilized IM-MS to probe the effect of penultimate proline on neuropeptide conformations. Their results showed that, besides penultimate proline's well-known role in protecting peptides from enzymatic degradation, it also plays a key role in increasing the conformational heterogeneity of neuropeptides, which may have different functions or binding affinities for their receptors.

In some cases, neuropeptides perform their biological activity through self-assembling into oligomers [20, 21] and IM-MS has proved extremely useful in advancing our understanding in self-assembly mechanism and oligomerization process. As an example, various amyloid-forming peptides self-assembly mechanism was deduced, providing important insights into fundamental mechanism for the amyloid firbils formation that is central implication in amyloid diseases such as Alzheimer's or Parkinson's diseases. [22, 23] Also, the study revealed that the largest oligomer for neuropeptide Leu-enkephalin identified is sixfold protonated nonadecamer and the the self-assembly mechanism is proposed, which is the same as granular, isotropic, insoluble macroscopic aggregates. [23] Subsequently, IM-MS studies of various Leu-enkephalin mutants highlighted the importance of characterizing dimer and higher oligomers in determining possible protofibril structures that a peptide system can access (i.e., single β-sheet or doublesheet steric zipper). [24, 25]

Utilizing IM-MS approach, Soper and colleges [26] screened a panel of neuropeptides for their direct interactions with Aβ monomers and small oligomers, aiming at finding a new class of biotherapeutics that can modulate aggregation and toxicity of the amyloid-β protein (Aβ), demonstrated that Leu-enkephalin binds selectively within a region of Aβ between its N-terminal tail and hydrophobic core.

One of the concern when conducting structure analysis by IM-MS is that how different conformations evolve during the ESI process. Even though a number of studies [27-30] demonstrated that peptide and protein ions in the gas phase can retain a memory of their solution structures upon ESI, but it remains unresolved how exactly the structure in the gas phase mimic the solution phase. Considerable efforts by Russell group [31-34] using cryogenic IM-MS (cryo-IM-MS) has revealed that intramolecular interactions can stabilize the kinetically trapped substance P dehydrated conformer on the time scale of several milliseconds, demonstrating cryo-IM-MS as a powerful approach that can provide a means to experimentally monitor snapshots of the structural evolution of biomolecule ions during the final stages of the evaporative processes of ESI.

Indeed, various factors including peptide inherent secondary structural elements and external environmental factors (solvent composition, temperature etc.) may affect peptide conformational preference in the gas phase. Increasing number of studies have been conducted utilizing the IM-MS tool to investigate the temperature [35, 36], activation voltage, [10] solvent composition, [11] and metal binding [37] dependant peptide conformational dynamics, which offers us important insights into peptide structure-functional relationship.

References:

1. Jurneczko, E.; Barran, P. E., How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **2011,** 136, (1), 20-28.
2. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X‐ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Communications in Mass Spectrometry* **2008,** 22, (20), 3297-3304.
3. Gidden, J.; Bushnell, J. E.; Bowers, M. T., Gas-phase conformations and folding energetics of oligonucleotides: dTG-and dGT. *Journal of the American Chemical Society* **2001,** 123, (23), 5610-5611.
4. Gidden, J.; Bowers, M., Gas-phase conformational and energetic properties of deprotonated dinucleotides. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics* **2002,** 20, (3), 409-419.
5. Wyttenbach, T.; Grabenauer, M.; Thalassinos, K.; Scrivens, J. H.; Bowers, M. T., The effect of calcium ions and peptide ligands on the relative stabilities of the calmodulin dumbbell and compact structures. *The Journal of Physical Chemistry B* **2009,** 114, (1), 437-447.
6. Jenner, M.; Ellis, J.; Huang, W. C.; Lloyd Raven, E.; Roberts, G. C.; Oldham, N. J., Detection of a Protein Conformational Equilibrium by Electrospray Ionisation‐Ion Mobility‐Mass Spectrometry. *Angewandte Chemie International Edition* **2011,** 50, (36), 8291-8294.
7. Bereszczak, J. Z.; Barbu, I. M.; Tan, M.; Xia, M.; Jiang, X.; van Duijn, E.; Heck, A. J., Structure, stability and dynamics of norovirus P domain derived protein complexes studied by native mass spectrometry. *Journal of structural biology* **2012,** 177, (2), 273-282.
8. Shi, H.; Pierson, N. A.; Valentine, S. J.; Clemmer, D. E., Conformation types of ubiquitin [M+ 8H] 8+ ions from water: methanol solutions: evidence for the N and A states in aqueous solution. *The Journal of Physical Chemistry B* **2012,** 116, (10), 3344-3352.
9. Shi, L.; Holliday, A. E.; Shi, H.; Zhu, F.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E., Characterizing intermediates along the transition from polyproline I to polyproline II using ion mobility spectrometry-mass spectrometry. *Journal of the American Chemical Society* **2014,** 136, (36), 12702-12711.
10. Pierson, N. A.; Valentine, S. J.; Clemmer, D. E., Evidence for a quasi-equilibrium distribution of states for bradykinin [M+ 3H] 3+ ions in the gas phase. *The Journal of Physical Chemistry B* **2010,** 114, (23), 7777-7783.
11. Pierson, N. A.; Chen, L.; Valentine, S. J.; Russell, D. H.; Clemmer, D. E., Number of solution states of bradykinin from ion mobility and mass spectrometry measurements. *Journal of the American Chemical Society* **2011,** 133, (35), 13810-13813.
12. Pierson, N. A.; Chen, L.; Russell, D. H.; Clemmer, D. E., Cis–Trans Isomerizations of Proline Residues Are Key to Bradykinin Conformations. *Journal of the American Chemical Society* **2013,** 135, (8), 3186-3192.
13. Papadopoulos, G.; Svendsen, A.; Boyarkin, O. V.; Rizzo, T. R., Conformational Distribution

of Bradykinin [bk+ 2 H] 2+ Revealed by Cold Ion Spectroscopy Coupled with FAIMS. *Journal of the American Society for Mass Spectrometry* **2012,** 23, (7), 1173-1181.
14. Voronina, L.; Rizzo, T. R., Spectroscopic studies of kinetically trapped conformations in the gas phase: the case of triply protonated bradykinin. *Physical Chemistry Chemical Physics* **2015**.
15. Pierson, N. A.; Clemmer, D. E., An IMS–IMS threshold method for semi-quantitative determination of activation barriers: Interconversion of proline cis↔ trans forms in triply protonated bradykinin. *International journal of mass spectrometry* **2015,** 377, 646-654.
16. Denys, L.; Bothner-By, A.; Fisher, G., Conformational diversity of bradykinin in aqueous solution. *Biochemistry* **1982,** 21, (25), 6531-6536.
17. Ottleben, H.; Haasemann, M.; Ramachandran, R.; Görlach, M.; Müller‐Esterl, W.; Brown, L. R., An NMR Study of the Interaction of 15N‐Labelled Bradykinin with an Antibody Mimic of the Bradykinin B2 Receptor. *European Journal of Biochemistry* **1997,** 244, (2), 471-478.
18. Chatterjee, C.; Mukhopadhyay, C., Conformational alteration of bradykinin in presence of GM1 micelle. *Biochemical and biophysical research communications* **2004,** 315, (4), 866-871.
19. Glover, M. S.; Bellinger, E. P.; Radivojac, P.; Clemmer, D. E., Penultimate Proline in Neuropeptides. *Analytical chemistry* **2015,** 87, (16), 8466-8472.
20. COWLEY, D. J.; HOFLACK, J. M.; PELTON, J. T.; SAUDEK, V., Structure of neuropeptide Y dimer in solution. *European Journal of Biochemistry* **1992,** 205, (3), 1099-1106.
21. Smith, D.; Griffin, J. F., Conformation of [Leu5] enkephalin from X-ray diffraction: features important for recognition at opiate receptor. *Science* **1978,** 199, (4334), 1214-1216.
22. Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V., Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature chemistry* **2009,** 1, (4), 326-331.
23. Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., Ion mobility–mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. *Nature chemistry* **2011,** 3, (2), 172-177.
24. Bleiholder, C.; Dupuis, N. F.; Bowers, M. T., Dimerization of chirally mutated enkephalin neurotransmitters: implications for peptide and protein aggregation mechanisms. *The Journal of Physical Chemistry B* **2013,** 117, (6), 1770-1779.
25. Do, T. D.; LaPointe, N. E.; Sangwan, S.; Teplow, D. B.; Feinstein, S. C.; Sawaya, M. R.; Eisenberg, D. S.; Bowers, M. T., Factors that drive peptide assembly from native to amyloid structures: experimental and theoretical analysis of [Leu-5]-enkephalin mutants. *The Journal of Physical Chemistry B* **2014,** 118, (26), 7247-7256.
26. Soper, M. T.; DeToma, A. S.; Hyung, S.-J.; Lim, M. H.; Ruotolo, B. T., Amyloid-β–neuropeptide interactions assessed by ion mobility-mass spectrometry. *Physical Chemistry Chemical Physics* **2013,** 15, (23), 8952-8961.
27. Kaddis, C. S.; Loo, J. A., Native protein MS and ion mobility: large flying proteins with ESI. *Analytical chemistry* **2007,** 79, (5), 1778-1784.
28. Heck, A. J., Native mass spectrometry: a bridge between interactomics and structural biology. *Nature methods* **2008,** 5, (11), 927-933.

29. Kondrat, F. D.; Kowald, G. R.; Scarff, C. A.; Scrivens, J. H.; Blindauer, C. A., Resolution of a paradox by native mass spectrometry: facile occupation of all four metal binding sites in the dimeric zinc sensor SmtB. *Chemical Communications* **2013,** 49, (8), 813-815.
30. Konijnenberg, A.; Butterer, A.; Sobott, F., Native ion mobility-mass spectrometry and related methods in structural biology. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2013,** 1834, (6), 1239-1256.
31. Silveira, J. A.; Fort, K. L.; Kim, D.; Servage, K. A.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H., From solution to the gas phase: stepwise dehydration and kinetic trapping of Substance P reveals the origin of peptide conformations. *Journal of the American Chemical Society* **2013,** 135, (51), 19147-19153.
32. Fort, K. L.; Silveira, J. A.; Pierson, N. A.; Servage, K. A.; Clemmer, D. E.; Russell, D. H., From Solution to the Gas Phase: Factors That Influence Kinetic Trapping of Substance P in the Gas Phase. *The Journal of Physical Chemistry B* **2014,** 118, (49), 14336-14344.
33. Servage, K. A.; Silveira, J. A.; Fort, K. L.; Russell, D. H., From Solution to Gas Phase: The Implications of Intramolecular Interactions on the Evaporative Dynamics of Substance P During Electrospray Ionization. *The Journal of Physical Chemistry B* **2015,** 119, (13), 4693-4698.
34. Silveira, J. A.; Servage, K. A.; Gamage, C. M.; Russell, D. H., Cryogenic ion mobility-mass spectrometry captures hydrated ions produced during electrospray ionization. *The Journal of Physical Chemistry A* **2013,** 117, (5), 953-961.
35. Zilch, L. W.; Kaleta, D. T.; Kohtani, M.; Krishnan, R.; Jarrold, M. F., Folding and unfolding of helix-turn-helix motifs in the gas phase. *Journal of the American Society for Mass Spectrometry* **2007,** 18, (7), 1239-1248.
36. Berezovskaya, Y.; Porrini, M.; Barran, P. E., The effect of salt on the conformations of three model proteins is revealed by variable temperature ion mobility mass spectrometry. *International Journal of Mass Spectrometry* **2013,** 345, 8-18.
37. Chen, L.; Gao, Y. Q.; Russell, D. H., How alkali metal ion binding alters the conformation preferences of gramicidin A: a molecular dynamics and ion mobility study. *The Journal of Physical Chemistry A* **2011,** 116, (1), 689-696.

Appendix III

Protocols for protein extraction, enzymatic digestion, glycopeptide enrichment and DiLeu labeling *etc*.

1. Preparation of cell lysate: (~0.5 hours)

Prepare 4% SDS (Sigma), 100mM Tris/Base in 10 mL of water.
(4 mL 10% SDS, 2.5 times of dilution).
(0.1M x 121.1 mg/mol x 0.01 L=121.1 mg).
Turning on the 95℃ heating block first if it's needed.

1. Add 65 μL 12 M HCl into it to adjust the pH to 8. The solution prepared here is **buffer.**
2. Lyse cells or tissues in buffer using (1:3-1:10, volume ratio) sample to buffer ratio is added and vortex to dissolve the cells. (From experience, below 200 μL buffer works fine.)
   - ✓ 50 μl Hela cell pellet, 150 μl buffer.
   - ✓ Usually 100-200 μl buffer.
3. Use micro sonicator 15s, 30s (interval) to dissolve the cells, 4 cycles in total. I've put the tube in ice water to avoid heating. 50% Amplitude. (FASP website: 95℃ for 3 min)
4. Store it in -80℃ freezer.

For PSC cells: Add 30 μL digestion buffer to PSC cells pellets, 95 ℃ for 3 min, 4 cycles of sonication at 10% amplitude.

2. BCA assay: （~1.5 hours）

1. Use DI water to dilute protein extract solution, 10x or 20x of dilution is good. (Prepare 70 μL,7μL + 63μL, two replicates for proteins)
2. Put the sample in -4 or -80 degree.
3. Calculate how much working reagent is needed. (Standards+1+samples)*replicates*200 μL.
   Standards: 2, 1, 0.5, 0.25, 0.125 μg/μL
   e.x. Standards=5, samples=1, replicates=2
4. (5+1+1)*2*200 μL=2.8 mL
5. Working reagent (A:B=50:1), A=5mL, B=0.1mL. Total volume: 5.1 mL.
6. Serial dilution to prepare standards, use DI water to dilute in a 0.6 ml vial. (75 μL + 75 μL). (Prepare 75 μL for each standard, two replicates for standards).
7. Transfer 25 μL of standards and diluted protein solution into 96 plate.
8. Add 200 μL working reagent into each vial and vortex for 1min.
9. Incubate in 37℃ for 40 min.
10. Turn on the UV 10 minutes beforehand and measure @570 nm.

3. FASP protocol: (~ 5 hours)

Turn on the 95℃ heating block first before preparing solution.

1. Solution preparation, please use the EXCEL. (Calculate how much solution is needed first) 1.5 hours. IAA should be put in darkness.
2. Centrifuge the thawed protein at 16, 000g for 5 min.
3. Combine 3 μL of DTT solution with 27 μL (200 μg protein in total, need to calcuLate how much volume is needed) protein extract. (10 times of dilution for DTT). (From previous experience, protein solution with a 70 μL works fine.)
4. Incubate the sample at 95℃ for 3 min to reduce disulfide bonds.
5. Add 200 μL of UA to the sample, transfer it to the YM-30 filter, Centrifuge at 14, 000g for 15 min at 20℃.
6. Add another 200 μL of UA to the sample, Centrifuge at 14, 000g for 15 min at 20℃.
7. Add 100 μL of IAA solution, gently swirl to mix, then incubate in dark for 20 min.
8. Centrifuge at 14, 000g for 10 min at 20℃.
9. Add 100 μL of UA, Centrifuge at 14, 000g for 15 min at 20℃, repeat 2x times.
10. Add 100 μL of ABC, Centrifuge at 14, 000g for 15 min at 20℃, repeat 2x times.

Trypsin digestion:

11. 200 μg protein needs 4 μg of enzyme (Ratio: 50:1). Add 8 μL 0.5 μg/μL onto the filter and 40 μL of ABC buffer.

PNGase-F: (Promega, 8 μL PNGase F is added for 200 μg of protein, 72 μL of ABC buffer is added.)

12. Incubate in 37℃ for 18h.
13. Transfer the filter unit to new collection tube and centrifuge at 14, 000g for 10min.
14. Add 50 μL 0.5 M NaCl and Centrifuge the filter units at 14, 000g for 10min. Repeat for 1x time.

    Preparation of 10% TFA

15. Add 350 μL DI water and 12.5 μL (10% TFA) to the vial. (Bigger volume to decrease loss, finally 0.25% TFA, pH 2~3, use pH paper to test the pH, finally 500 μL)
16. Use Sep-Pak to desalt. (Use Ziptip if it's less than 100 μg)

Calculate how much solution is needed, Preparation of 0.1% TFA (4.5 mL/vial), 50% ACN 0.2% FA (1 mL/vial), 70% ACN 0.2% FA (0.7 mL/vial)

1. Conditioning solution: 100% ACN, 1mL.
2. Equilibration solution: 0.1% TFA in water, 3 mL.
3. Loading sample: 500 μL of sample.
4. Use another 500 μL 0.1% TFA to wash the sample.
5. Desalting solution: 0.1% TFA in water, 1 mL.
6. Elution solution:
   50% ACN 0.2% FA 1 mL.
   70% ACN 0.2% FA 0.7 mL.
7. Drying down.

Note: After SCX, if ammonium formate wants to be removed. Then add 1 mL 0.1% TFA to it and

add 10 uL TFA to the sample to adjust the pH to 2-3. Use 2 mL 0.1% TFA to wash to get rid of salts.

4. Dileu labeling: (3.5 hours, drying down takes another 1 hour)

Calculate how much activation is needed: 50 μL x number of samples
Preparation of activation solution, see EXCEL on P107 on notebook. (1mL activation solution: 12.0651 mg DMTMM (in -20℃ freezer), 4.0434 μL NMM (in the hood), the solvent is DMF)
Note: 115a and 117b were used. (0.5 mg dileu, 50 uL activation solution)

1. Dissolve 1 mg of Dileu in 100 μL MeOH and take (10xsample weight) in new vials.
2. Dry vials down.
3. Weigh out DMTMM and NMM and dissolve them in DMF to make activation solution. Dissolve labels in 50 μL activation solution and vortex for 1h.
4. Dissolve samples in 10 μL (50 ug peptides) 0.5 M TEAB buffer and combine activated labels with samples in each vial, vortex for another 2h.
5. Quench with 5% NH2OH to make final NH2OH 0.25%, vortex. (60 μL reaction solution, 3.16 μL of 5% NH2OH is needed.)
6. Drying down.

5. Getting rid of Dileu (Use SCX column, for proteomics column): (~3.5 hours and drying down takes another 2 hours) PolySULFOETHYL A™

Preparation of mobile phase A and B, (Mobile phase A: 25% ACN, 10 mM ammonium formate, mobile phase, formic acid to adjust it to **pH=3** B: 25% ACN, 500 mM ammonium formate, **pH=6.8**)
200 mL A needs 126 mg ammonium formate, and 710 uL FA to adjust pH TO 3.
200 mL B needs 6.3 g ammonium formate, and don't need NH3 to adjust pH.
Proteomic SCX column less pressure 1500~2000 psi.

1. Preparing the LC system following the instructions. LC method used: SCX_QY_CombineAll.
2. Install the column, set the column temperature at 30℃, let it condition for 15 min.
3. Pressure should be ~1000 psi at 0.2 μL/min, 100% A.
4. Run a blank 85 min.
5. Dissolve two samples in 54 μL mobile phase A and combine two channels, total volume 108 μL, transfer it to an injection vial.
6. Run sample, and discard the first 20 min (Dileu labeling reagent).
7. Use 8 vials to collect 40 min runs, at 0.2 mL/min, 5 min/vial, namely 1 mL/vial. Use 2 mL vial to collect samples.
8. Drying down and combine all the samples in ONE vial.

6. High pH fractionation on C18 column,

Method used: 2015_Xiaofan_HpH_C18_94min
Normal Pressure ~1500 psi. Prime again if pressure not stable.
Mobile phase A: H2O, 10 mM ammonium formate, pH=10,
(126mg is needed to make 200 mL, 920 uL ammonium hydroxide to adjust the pH)
Mobile phase B: 90% ACN, 10 mM ammonium formate, pH=10
(126mg is needed to make 200 mL, 4.4mL ammonium hydroxide to adjust the pH)

7. Buffers to prepare:

1. Binding buffer: 0.1% TFA, 19.9% H_2O, 80% ACN
2. Elution buffer: 0.1% TFA, 99.9% H_2O

Calculate how much solvent is needed:

1) 1.5 mL binding buffer per experiment for 200μg tryptic peptides.
2) 0.6 mL elution buffer per experiment for 200μg tryptic peptides.
3)

4)	Total volume	ACN	H2O	TFA
Biding buffer	10 mL	8 mL	1.99 mL	10 uL
Biding buffer	5 mL	4	995 uL	5 uL
Elution buffer	5 mL	0	4995 uL	5 uL

	Activation	washing	Dissolving peptide	Wash/Elute
50 ug peptide	50 uL	50 uL	125 uL	25 uL
100 ug peptide	100 uL	100 uL	250 uL	50 uL
200 ug peptide	200 uL	200 uL	500 uL	150 uL

The concentration of beads in elution buffer in the activation step is 0.05 mg/uL, which means 100 ul, there is 5 mg, 200 uL there is 10 mg. Beads concentration: 0.05 mg/uL, weighing 16.1 mg, volume needed is 16.1/0.05=322 uL.

<u>For 100 μg tryptic peptides</u>

1. The ZIC-HILIC beads were activated with 100 uL of elution buffer for 30 min, vortex @ rt.
2. Beads were washed with 100 uL binding buffer (2X).
3. 100 ug tryptic peptides were dissolved in 250 uL of binding buffer and mixed with beads at a 1:50 peptide-to-material mass ratio.
4. The tube was shaken over a vortex mixer for 1 h.
5. Supernatant was collected by centrifugation.
6. The beads were washed with 50 uL binding buffer (6X).
7. The beads were eluted with 50 uL elution buffer (5x) and supernatant was dried down.

8. Glycopeptides enrichment by lectin method: （~3.5 hours）

Preparation of lectin mixtures solution (90 μg ConA, 90 μg WGA and 90 μg RCA120 in 36 μL 2xbiniding buffer).

Preparation of sugar mixtures solution (300 mM *N*-acetyl-D-glucosamine, D-lactose, methyl α-D-mannopyranoside 1xbinding buffer). Vortex for 30 min to allow it fully dissolve.

Preparation of 2xbiniding buffer, 1xbinding buffer (1 mM $CaCl_2$, 1 mM $MnCl_2$, 0.5 M NaCl in 20 mM TrisBase, pH 7.3). Note: Use HCl to adjust the pH, ~30 μL 12 M HCl.

1. Dissolve the tryptic peptides in 80 uL binding solution and transfer it to a new filter.
2. Adding 36 uL lectin mixtures solution onto the filter.
3. Incubate at room temperature for 1 hour.
4. Centrifuge at 14, 000 for 10 min at 18 degree.
5. Wash with binding solution for 4 times, each time 200 uL.
6. Change to a new collection vial. Add 100 uL sugar mixtures solution to the filter and vortex for 1min, incubate in room temperature for 30min.
7. Add another 100 uL sugar mixtures to the filter and vortex for 1min, incubate in room temperature for 30min.
8. Acidify with 10% TFA to reach 0.25% TFA (pH 2-3), and desalt with Omix-Tip. Namely, as total volume is 200 uL, add 5 uL 10% TFA.

9. O-glycopeptide enrichment according to Benjamin Garcia Protocol

1. O-HexNAc peptide enrichment was done by solid state extraction with PBA cartridges and non-aqueous DMSO.
2. After washing the cartridge with 3 mL **anhydrous DMSO**, digested peptides dissolved in 35 uL DMSO were loaded and incubated with PBA resin at 37 ^{o}C for 2 hours. Let one and a half drop flow and then start to seal the SPE column with parafilm and plug.
3. Then non-bound peptides were washed away with **anhydrous ACN** 3mL.
4. Lastly, bound peptides were eluted with 0.1% TFA solution after incubation at room temperature for 1 hour. 600 uL one time, incubate, 600 uL another time, incubate.
5. Dry down

10. OMIX tip to get rid of salts: (80 μg/tip protein amount, small scale of removing salts) (Takes ~30min)

Preparation of washing solution (0.1% TFA, 600 μL/sample), Samples were dissolved in 100 μL 0.1% TFA

1. 100 μL ACN to condition tips, 2x times.
2. 100 μL washing solution and discard, 3x times.
3. Load samples onto the tip, aspirate and dispense for 10 cycles.
4. Use 100 μL of washing solution to wash for 3x times.
5. Elute with 100 μL 0.2% FA 50% ACN and 100 μL 0.2% FA 75% ACN, each 10x times in two vials and combine.
6. Drying down.

11. Injection on Fusion Lumos: (Takes 1 day)

1. Dissolve the sample with 0.1% FA water to a concentration of 1 μg/μL. For glycopeptides, dissolve in 10 μL solvent.
2. Injection volume 1-2 μL.
3. Search QC sample with Compass, 16,000-18000 peptides.

12. Glycosylation mapping:

1mL ABC buffer, 3.95 mg (O18 water is in 5234)

1. Following **step 3** in lectin enrichment protocol, the captured peptides were washed four times with 200 μl binding buffer and twice with 50 μl 50 mM NH4HCO3 in H218O.
2. Finally 8 μl PNGase F (1 U/μl H218O) (Roche) in 72 μl 50 mM NH4HCO3 in H218O were added to the filter units and the samples were incubated for 3 hr at 37 ℃ .
3. Centrifuge and collect flow-through at 14, 000 for 10 min at 18℃。
4. The deglycosylated peptides were eluted with 2 × 50 μl 50 mM NH4HCO3.
5. Acidify with 10% TFA to reach 0.25% TFA (pH 2~3) and desalt with Omix-tip. Namely, as total volume is 200 μL, add 5 μL 10% TFA.
6. OMIX tip to desalt.

1. Glycopeptides in 50mM Ammonium Bicarbonate H2O(18) (pH 7.8) to a final volume of 50 uL.
2. Add 2μl of recombinant PNGase F.
3. Incubation 2.5 hours for 37 degree.
4. Zip-tip do desalt.

13. Conditioned media collection protocol:

~ 7 hours (10 am - 5:00 pm)

1. Approximately 1×107 cells were cultured at 37 ℃ in 5 % CO2 in DMEM (Hyclone, USA) supplemented with 10 % fetal bovine serum until reaching 80 % confluence.
2. Cells were washed gently 4 times with Dulbecco's phosphate buffered saline with calcium and magnesium (DPBS) and 4 times with serum-free DMEM (Conditioned media, CM), each time 3 mL.
3. Cells were then incubated in the serum-free DMEM at 37 ℃ for 24 h for pancreatic cancer cells, 48 h for PSC cells.
4. Centrifuge at 3, 500g for 10min at 4℃ to get rid of dead cells and cell debris.
5. Filter with 0.45 um filter to get rid of smaller debris.

After centrifugation take the tube down to the 5 th floor.

6. Formic acid (FA, final concentration of 0.1 %) was immediately added to the final supernatant, which was then stored at -80 ℃. The addition of FA lowered the pH (pH < 4) of the culture supernatants, thus reducing the activity of many proteases[11,22]. 20 uL FA for 20 mL conditioned media.

7. The proteins in the culture supernatants were extracted by Trichloroacetic acid (TCA) (Fridge 2, 5234) precipitation. **(Keep supernatant)**
8. 66.7 uL (Cell media 20 mL) of 30% NLS was added to the solution to reach a concentration of 0.1%。
9. 2.22 mL 100% TCA was added (Cell media 20 mL) to the CM solution to a final concentration of 10 % (w/v, mg/mL). ((100g/100mL, w/v), 1g/mL solution is 100% TCA)
10. After mixing, the solution was precipitated on the ice for 2 hours.
11. Followed by centrifugation (15,000 ×g, 10 min, 4 ℃).
12. Add 1.7 mL precooled (5236 hood, precooled on ice) THF and transfer it to 2 mL vials and the protein pellet was carefully washed twice with 1.7 mL tetrahydrofuran (THF) and each wash was followed by centrifugation (15,000 ×g, 10 min, 4 ℃). (Total washes: 2x times!)
13. The final pellet was re-dissolved in 35 uL lysis buffer (4 % SDS, 0.1 M Tris-HCl, pH 7.6) with a sonicator bath (30 min extraction). (Prepare 1 mL lysis solution, 400 uL 10% SDS + 100 uL 1 M Tris/Base (121 mg in 1 mL H2O)+ 500 uL H2O, add 35 uL 10% HCl.)
14. The concentrations of the extracted proteins were measured by BCA assay.

14. PNGase F digestion:

1. 800 ug tryptic peptides from CSF protein, 16 uL PNGase F from Promega then add 144 uL ABC buffer, 1 promega unit equals to 65 NEB unit.
2. 10 mL of 50 mM ABC buffer, weigh 39.5 mg ABC. pH = 7.8
3. Sep-pak, 1 column (50 mg packing material) for 500 ug tryptic peptides most.
4. While for PBA boronic acid catridge (100 mg packing material), 1 column for 1 mg tryptic peptides.

After PNGase F digestion to get rid of N-glycans, use Sep-pak to seprate released N-glycans and peptides. Dry down!

15. IMAC Protocol (for 1 mg protein)

1. Pipette 500 uL magnetic IMAC beads suspension into 2 mL tube (change beads amt according to protein amt)
2. Wash beads with 1 mL H_2O (3x)
3. Add 1 mL 40 mM EDTA (prepare 10 mL), pH = 8 (need at least 300 uL 2M NaOH to adjust pH, sonicate to dissolve), shake for 30 min @ vortex vigorously
4. Wash beads with 1 mL H_2O (5x)
5. Add 1 mL 100 mM $FeCl_3$ (prepare before this step), shake for 30 min @ vortex vigorously
6. Wash beads with 1 mL 80% ACN 0.15% TFA (4x).
7. Resuspend sample in 500 uL 80% ACN 0.15% TFA. Quantitatively transfer to beads, shake for 30 min
8. Centrifuge 15 s, save supernatant. (phosphor-depleted peptides)
9. Wash beads with 1 mL 80% ACN 0.15% TFA (3x), save first two washes
10. Rinse 0.6 mL tube with ACN, fill with 50 uL 4% FA

11. Elute with 100 uL 50% ACN 0.7% NH_4OH (2x), vortex 1 min, elute into 4% FA tube (2x)

Wash beads:

12. H_2O 3 x 1 mL
13. 1mL 40 mM EDTA, shake for 5 min
14. H_2O 3 x 1 mL

Store beads:

15. 30% EtOH