Development and Application of Mass Spectrometry Methods for Proteomic and Post-translational Modification Analysis

By

Danqing Wang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

University of Wisconsin-Madison

2023

Date of Final Oral Examination: 5/8/2023

This dissertation is approved by the following members of the Final Oral Exam Committee:

Lingjun Li, Professor, Department of Chemistry and School of Pharmacy

Ying Ge, Professor, Department of Chemistry and Department of Cell and Regenerative Biology

Manish Patankar, Professor, Department of Obstetrics and Gynecology

Luigi Puglielli, Professor, Department of Medicine

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my PhD advisor, Dr. Lingjun Li. Without her unwavering support, expert guidance, and profound knowledge, none of my work would have been possible. I still remember when I first met her during my rotation in 2018, I was warmly welcomed and impressed by her outstanding research. Since I joined the lab, Dr. Li has been incredibly supportive to any of my research interest and project. She always tries her best to provide the academic resources to us and encourage us to pursue all the possibilities in science. The road towards the PhD is not easy, but I am glad to have Dr.Li as my advisor, always there to support and inspire me. She is an excellent role model in all aspects and has taught me to be an independent research scientist. I will always be grateful for her patience and hard work, which will continue to motivate me throughout my future career. I feel genuinely thankful for the opportunity to join her lab, work with her, and grow under her mentorship.

I would also like to express my gratitude to my committee members, Dr. Ying Ge, Dr. Manish Patankar, and Dr. Luigi Puglielli, for their valuable time, helpful suggestions, and inspirations during my thesis background examination, original research proposal, 4th year committee meeting, and finally, this dissertation. Dr. Puglielli has been a fantastic collaborator, and I am grateful for his patience and guidance throughout of our collaboration.

I am also indebted to my collaborators who lent their expertise and allowed me to work on various exciting bioanalytical projects: Dr. Andrew Alpert, Dr. Henrik Zetterberg, Dr. Ezazul Haque, Dr. Timothy Rhoads, Dr. Rozalyn Anderson, Dr. Paul Ahlquist, Dr. Masaki Nishikiori, and Dr. Xudong Shi. I consider myself fortunate to have worked with such talented individuals and hope for more fruitful collaborations between our labs in the future.

I have benefited a lot from my mentors and have been fortunate to work with many excellent colleagues in the lab. I would like to acknowledge Dr. Junfeng Huang, who became my mentor since my first rotation, introduced me into phosphoproteomics and glycoproteomics research, and trained me in many essential skills, including instrument set-up, sample preparation, data analysis. I am grateful for his willingness to share his expertise and provide valuable guidance throughout my research. I would also like to thank Dr. Yusi Cui for her excellent guidance on my glycoproteomic-related projects, and Dr. Miyang (Mike) Li and Zicong Wang for their collaboration on the isobaric tag-related project and tag synthesis. I extend my gratitude to Dr. Dustin Frost for his advice on DiLeu labeling and my peer lab mate, Ting-Jia Gu, for her tremendous help and inspiration to many of my research projects. I am also thankful for the help and contributions of my mentees, Feixuan Wu and Peng-Kai Liu. I would also like to address my appreciation to Dr. Zhengwei (Tony) Chen, Dr. Xiaofang Zhong, Dr. Haiyan Lu, Dr. Hua Zhang, Dr. Shuling Xu, Dr. Yatao Shi, Dr. Zihui Li, Dr. Qinying Yu, Bin Wang, Yuan Liu, Dylan Tabang, Jericha Mill, Zhijun (Andrew) Zhu, Haoran Zhang, and many other Li lab members. Working in such a large family and meeting so many wonderful colleagues and friends has been an incredible experience.

Finally, I am grateful to all my family and friends who have supported me through this challenging yet rewarding journey. My heartfelt thanks go to my parents, Bo Wang and Ye Xu, for their unreserved love throughout my entire life. I would like to thank my grandparents, who are/would be very proud of my achievements. Last and most importantly, I want to extend my deepest appreciation to my boyfriend, Zichong Li, for his unwavering love, companionship, and encouragement that have sustained me throughout the ups and downs of this journey. I consider myself incredibly fortunate to have him by my side.

To all of those who have supported me, thank you again from the bottom of my heart.

Table of Contents

Development and Application of Mass Spectrometry Methods for Proteomic and Post-translational Modification Analysis

Danqing Wang

Under the supervision of Professor Lingjun Li

University of Wisconsin-Madison

Abstract

Proteins are essential biomolecules that perform a wide range of biological functions. Post-translational modifications (PTMs) substantially impact protein structure and function, making their characterization essential for understanding complex biological systems. This dissertation focuses on developing and applying novel mass spectrometry (MS)-based methodologies to address challenges in studying two common and important PTMs: phosphorylation and glycosylation. To this end, new enrichment materials and their corresponding workflows, including Cotton Ti-IMAC, epoxy-ATP-Ti^{4+} IMAC, and Very Weak Anion Exchange (VWAX) have been introduced for efficient phosphopeptide and glycopeptide enrichment. A strategy combining boronic acid enrichment, high-pH fractionation, and EThcD has been developed for comprehensive O-glycosylation profiling. Additionally, the Boost-DiLeu quantitative approach has been introduced to enhance glycopeptide quantification in size-limited samples, while a periodate oxidation-based SUGAR tag labeling method has been established for high-throughput, intact sialylated glycopeptide-specific quantification. These methods have been applied to study human diseases, such as Alzheimer's Disease, providing insights into dysregulated glycosylation patterns and their potential implications in disease pathogenesis. Overall, this work contributes to advancing MS-based proteomics strategies and broadening our understanding of the roles of PTMs in biological systems and is anticipated to inspire future research endeavors in related fields.

Chapter 1

Introduction and Research Summary

Introduction

Proteins are important biomolecules that play critical roles in nearly all biological processes, including enzymatic catalysis, signal transmission, and structural support.[1] A systematic study of proteins is essential for a comprehensive understanding of biological networks and systems, both structurally and functionally. The proteome refers to the total number of proteins present within either a cell, tissue, or organism at a given time.[2] However, it should be noted that the proteome is not static, as the protein population can differ from cell to cell and change over time. Proteomics, as a postgenomic discipline, aims to identify and quantify all proteins in a proteome, including their expression, cellular localization, interactions, post-translational modifications (PTMs), and turnover across time, space, and cell type.[3] Exploring the proteome poses more significant challenges than sequencing the genome, making it an exciting and challenging field of study.

Mass spectrometry (MS) is a powerful tool for large-scale proteome analysis, with soft ionization techniques such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) commonly used to analyze biomolecules.[4,5] ESI involves applying a high voltage to a flow of liquid at atmospheric pressure, generating gas phase ions through the desolvation of highly charged droplets. Due to its ability to produce multiply charged ions, ESI is ideal for high-molecular-weight molecule analysis such as peptides and proteins. Tandem mass spectrometry can isolate different analyte ions in the analyzer, generate fragment ions from the selected ions and analyze the fragmented ion spectra, providing more structural information of the original parent ions. When coupled with liquid chromatography for separation, LC-ESI-MS/MS is a suitable platform for large-scale proteomics analysis.

While the MS-based bottom-up workflow has been well-established for global proteomic analysis in the past decades,[6] an additional layer of complexity arises from PTMs of proteins. Besides the post-transcriptional regulation, the proteins can be further post-translationally modified by covalently adding some chemical functional groups or altering the structure of the amino acids, which can significantly impact the function and properties of proteins.[3] Mapping PTMs is therefore a critical aspect of characterizing the whole proteome.

This dissertation focuses on two of the most common protein PTMs: phosphorylation and glycosylation. Approximately one-third of all proteins in mammalian cells are phosphorylated at any given time, with this reversible modification mainly happening on serine, threonine, and tyrosine residues.[7] Phosphorylation is involved in numerous cellular functions, including protein synthesis, cell signaling, apoptosis, and enzymatic regulation.[8–10] On the other hand, glycosylation is critical for physiological and pathological cellular functions, with almost half of all proteins being glycosylated.[11,12] Aberrant glycosylation has been correlated with altered cancer cell adhesion, invasion, and proliferation.[13] N-linked and O-linked glycosylation are the two main types, with a consensus amino acid sequence (Asn-X-Thr/Ser, X can be any amino acid except proline) containing the glycosylation site of N-linked glycoproteins, while no consensus motif has been found for O-linked glycoproteins yet.[14]

Despite many technological improvements in MS instruments, the MS analysis of these two PTMs still faces numerous challenges, particularly arising from the low abundance and poor ionization efficiency of modified peptides.[15] Additionally, the heterogeneity of glycosylation raises the complexity of performing site-specific MS analysis. Furthermore, development of high-

throughput quantification methods is still in need to reveal protein and PTM expression patterns in different biological systems and disease models. This dissertation aims to address some of these challenges by developing novel and improved MS-based analytical strategies, as well as to apply these new methodologies to the studies of human diseases.

Research Summary

Chapter 1 serves as a general introduction to the work in this dissertation and briefly introduces the background and major findings in each project.

Due to the low abundance and poor ionization efficiency, direct analysis of phosphopeptides in complex biological samples by MS is challenging and a reliable and efficient enrichment method is essential. **Chapter 2** introduces a phosphorylated cotton fiber-based Ti(IV)-IMAC material (termed as: Cotton Ti-IMAC) as a novel platform for phosphopeptide enrichment. The material can be easily prepared using cost-effective reagents within 4 hours and can be packed as a spin-tip for convenient enrichment. Experimental results demonstrate its high selectivity and sensitivity in phosphopeptide enrichment from both protein standards and cell digests. With its flexibility and reproducibility, Cotton Ti-IMAC is ready to serve as a widely applicable and robust platform for achieving large-scale phosphopeptide enrichment and expanding our knowledge of phosphoproteomics in complex biological systems.

Chapter 3 describes a method that extends the enrichment approach from a single PTM to multiple PTMs. Previously, we explored the electrostatic and hydrophilic properties of commercial centrifuge-assisted-extraction Titanium (IV) IMAC (CAE-Ti-IMAC) material and its application in simultaneously separating common N-glycopeptides, phosphopeptides, and M6P glycopeptides

with a dual-mode enrichment.[16,17] Yet we found the hydrophilic property of the CAE-Ti-IMAC microspheres was not satisfactory, since the material was not specifically developed for hydrophilic interaction chromatography (HILIC). To address this issue, this chapter presents the development of a hydrophilicity-enhanced bifunctional Ti-IMAC material with grafted adenosine triphosphate (denoted as: epoxy-ATP-Ti^{4+}). The epoxy-ATP-Ti^{4+} IMAC material can be prepared from epoxy-functionalized silica particles via a convenient two-step process. The ATP molecule not only provides strong and active phosphate sites for binding phosphopeptides in conventional IMAC mode,[18–21] but also contributes significantly to the hydrophilicity, which permits the enrichment of glycopeptides via HILIC. Both modes can be implemented simultaneously, allowing for the sequential collection of glycopeptides and phosphopeptides in a single experiment from the same sample. This material and associated fractionation method enable simple and effective enrichment and separation of glycopeptides and phosphopeptides from HeLa cell digests and mouse lung tissue samples, providing a useful tool to study potential crosstalk between these two important PTMs in biological systems.

Apart from IMAC-based enrichment strategy, **Chapter 4** introduces the application of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) in the field of glycopeptide enrichment. While HILIC has been widely used for glycopeptide enrichment,[14,15] signals of glycopeptides carrying a negatively charged glycan can still be suppressed by the more abundant neutral glycopeptides in positive ion mode LC-MS/MS analysis, thus impairing their detection and identification.[17] ERLIC overcomes this limitation by using charged stationary phases with HILIC mobile phases (MPs), where electrostatic and hydrophilic interactions between

analytes and stationary phase are superimposed.[22] The charge states of analytes can be adjusted via altering the pH of the MPs to control the electrostatic interaction between the analytes of interest and the stationary phase. However, the MPs used for conventional ERLIC involve high concentrations of salt, which are not compatible with MS analysis without extra desalting steps.[23–25] In this chapter, we introduce experimental materials for Very Weak Anion Exchange (VWAX) that become uncharged more readily than conventional anion-exchange materials. Well-retained peptides are eluted with significantly less salt, which enables a desalt-free workflow. With optimized eluting gradient, VWAX can effectively separate glycopeptides with negatively charged glycans from neutral glycopeptides, allowing for enhanced detection of sialylated, M6P, and sulfated glycopeptides. We anticipate VWAX become a new tool to profile the glycoproteome and provides new insights into those less-studied glycosylation types.

Chapter 5 represents an extension of the study to investigate O-glycosylation in addition to N-glycosylation. Although O-glycosylation is also involved in different biological interactions and disease progression and development, its structural complexity and lack of robust enrichment methods from complex biological systems have impeded its study.[26] In this chapter, we present an integrated strategy that combines universal boronic acid enrichment, high-pH fractionation, and electron-transfer and higher-energy collision dissociation (EThcD) for enhanced intact O-glycopeptide analysis to achieve an unbiased O-glycosylation profiling. This strategy is applied to analyze the O-glycoproteome in cerebrospinal fluid (CSF), which is a direct source reflecting the ongoing physiological or pathological state of the central nervous system (CNS).[27,28] The results provide the largest dataset of O-glycoproteins and O-glycosites reported for human CSF to date.

Additionally, we develop a peptidomics workflow that utilizes EThcD and a three-step database searching strategy for comprehensive PTM analysis of endogenous peptides, including N-glycosylation, O-glycosylation, and other common peptide PTMs. Subsequently, we profile CSF samples from healthy individuals, mild cognitive impairment (MCI), and Alzheimer's Disease (AD) patients, and revealed a landscape of glycosylation patterns in different disease states.

Chapter 2-5 of this thesis focus on the development of new enrichment techniques and qualitative analysis of PTMs, whereas **Chapter 6-7** shift the focus to the development of quantitative strategies for investigating changes in PTM expression in biological samples.

Chapter 6 introduces the Boost-DiLeu strategy for intact glycopeptide quantification, which utilizes an additional boosting channel to enhance *N*,*N*-dimethyl leucine (DiLeu) tagging-based quantitative glycoproteomic analysis from quantity-limited samples. By integrating a one-tube sample processing workflow and high-pH fractionation, over 3500 quantifiable N-glycopeptides are identified from only 30 µg HeLa cell tryptic digests with reliable quantification performance. Furthermore, this strategy is applied to human CSF samples to differentiate N-glycosylation profiles between AD patients and non-AD donors. The results reveal processes and pathways affected by dysregulated *N*-glycosylation in AD, including platelet degranulation, cell adhesion, and extracellular matrix, which highlights the involvement of N-glycosylation aberrations in AD pathogenesis. Moreover, weighted gene co-expression network analysis (WGCNA) shows 9 modules of glycopeptides, two of which are associated with the AD phenotype. Collectively, our findings demonstrate the feasibility of using this strategy for comprehensive glycoproteomic

analysis of size-limited clinical samples, as well as its great potential in identifying novel therapeutic targets or biomarkers in other biological systems with limited sample quantity.

Sialylation is one of the most common glycosylation types and plays a crucial role in cellular recognition, cell adhesion, and cell signaling. The high acidity and numerous hydroxyl groups in sialic acids strongly influence the surrounding species.[29] Altered expression of sialylation has been reported in many diseases.[11] To gain further insights into its role with site-specific information, there is a need for developing efficient quantification strategies that target intact sialylated glycopeptides (SGPs). In **Chapter 7**, we describe a novel high-throughput quantification strategy for intact SGPs by combining mild periodate oxidation and our in-house developed 12-plex isobaric multiplex labeling reagents for carbonyl-containing compound (SUGAR) tag labeling. The oxidation reaction exclusively introduces an aldehyde group to the sialic acid, which can directly react with the SUGAR tag. We first demonstrate the efficacy of this pipeline using an SGP standard and further extend its utility to provide accurate and precise quantitation when analyzing complex samples. The labeled sample achieves comparable sialyl-glycoproteome coverage compared to label-free samples in bovine fetuin and mouse heart homogenates while maintaining good quantification accuracy. The method is then applied to quantify SGP in two brain regions between wild-type and AD mouse models. Quantitative results reveal dysregulated sialylated glycoforms in both cortex and hippocampus regions, with more distinct differences observed for hippocampus region. We performed the gene ontology (GO) analysis to illustrate the network of enriched biological processes based on the dysregulated glycoforms. Our results demonstrate the

feasibility of our method for high-throughput and site-specific quantification of sialylation from different disease models in a single experiment.

Chapter 8 provides a comprehensive overview of the significant results and findings obtained from the preceding chapters. Furthermore, it deliberates on the future possibilities and potential future directions of my dissertation research.

Reference

(1) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J. Jr.; Stryer, L. Exploring Proteins and Proteomes. In *Biochemistry 8th edition*; W. H. Freeman and Company: New York, NY., **2015**.

(2) Chandramouli, K.; Qian, P.-Y. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Hum. Genom. Proteom.* **2009**, *1* (1), 239204.

(3) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394.

(4) Singhal, N.; Kumar, M.; Kanaujia, P. K.; Virdi, J. S. MALDI-TOF Mass Spectrometry: An Emerging Technology for Microbial Identification and Diagnosis. *Front Microbiol.* **2015**, *6*, 791.

(5) Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **2003**, *24* (1), 3–12.

(6) Gillet, L. C.; Leitner, A.; Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **2015**, *9* (1), 1–24.

(7) Zolnierowicz, S.; Bollen, M. Protein Phosphorylation and Protein Phosphatases. *EMBO J.* **2000**, *19* (4), 483–488.

(8) Gjertsen, B. T.; Døskeland, S. O. Protein Phosphorylation in Apoptosis. *Biochim. Biophys. Acta* **1995**, *1269* (2), 187–199.

(9) Pawson, T.; Scott, J. D. Protein Phosphorylation in Signaling – 50 Years and Counting. *Trends Biochem. Sci.* **2005**, *30* (6), 286–290.

(10) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. L. The Crucial Role of Protein Phosphorylation in Cell Signaling and Its Use as Targeted Therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271–280.

(11) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.

(12) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1* (1), 90.

(13) Frost, D. C.; Li, L. Recent Advances in Mass Spectrometry-Based Glycoproteomics. *Adv. Protein Chem. Struct. Biol* **2014**, *95*, 71–123.

(14) Chen, Z.; Huang, J.; Li, L. Recent Advances in Mass Spectrometry (MS)-Based Glycoproteomics in Complex Biological Samples. *TrAC - Trends Anal. Chem.* **2019**, *118* (Cell 126 2006), 880–892.

(15) Huang, J.; Wang, F.; Ye, M.; Zou, H. Enrichment and Separation Techniques for Large-Scale Proteomics Analysis of the Protein Post-Translational Modifications. *J. Chromatogr. A* **2014**, *1372*, 1–17.

(16) Huang, J.; Dong, J.; Shi, X.; Chen, Z.; Cui, Y.; Liu, X.; Ye, M.; Li, L. Dual-Functional Titanium(IV) Immobilized Metal Affinity Chromatography Approach for Enabling Large-Scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. *Anal. Chem.* **2019**, *91* (18), 11589–11597.

(17) Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

(18) Zhang, L.; Zhao, Q.; Liang, Z.; Yang, K.; Sun, L.; Zhang, L.; Zhang, Y. Synthesis of Adenosine Functionalized Metal Immobilized Magnetic Nanoparticles for Highly Selective and Sensitive Enrichment of Phosphopeptides. *Chem. Commun.* **2012**, *48* (50), 6274–6276.

(19) Zhang, L.; Liang, Z.; Zhang, L.; Zhang, Y.; Shao, S. Facile Synthesis of Gallium Ions Immobilized and Adenosine Functionalized Magnetic Nanoparticles with High Selectivity for Multi-Phosphopeptides. *Anal. Chim. Acta* **2015**, *900* (C), 46–55.

(20) Su, J.; He, X.; Chen, L.; Zhang, Y. Adenosine Phosphate Functionalized Magnetic Mesoporous Graphene Oxide Nanocomposite for Highly Selective Enrichment of Phosphopeptides. *ACS Sustain. Chem. Eng.* **2018**, *6* (2), 2188–2196.

(21) Yan, Y.; Lu, Y.; Chen, M.; Liang, H. A Novel IMAC Platform – Adenosine Coupled Functional Magnetic Microspheres for Phosphoproteome Research. *Anal. Methods* **2018**, *10* (10), 1190–1195.

(22) Alpert, A. J. Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Anal. Chem.* **2008**, *80* (1), 62–76.

(23) Alpert, A. J.; Hudecz, O.; Mechtler, K. Anion-Exchange Chromatography of Phosphopeptides: Weak Anion Exchange versus Strong Anion Exchange and Anion-Exchange Chromatography

versus Electrostatic Repulsion–Hydrophilic Interaction Chromatography. *Anal. Chem.* **2015**, *87* (9), 4704–4711.

(24) Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L. Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2019**, 1–11.

(25) Cui, Y.; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L. Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Anal. Chem.* **2021**, *93* (22), 7908–7916.

(26) Hounsell, E. F.; Davies, M. J.; Renouf, D. V. O-Linked Protein Glycosylation Structure and Function. *Glycoconjugate J.* **1996**, *13* (1), 19–26.

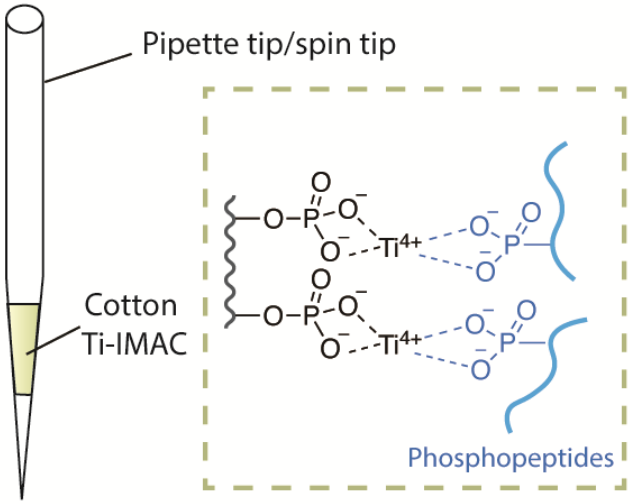
(27) Zhang, J. Proteomics of Human Cerebrospinal Fluid – the Good, the Bad, and the Ugly. *Proteom. - Clin. Appl* **2007**, *1* (8), 805–819.

(28) 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*, 39–64.

(29) Puigdellívol, M.; Allendorf, D. H.; Brown, G. C. Sialylation and Galectin-3 in Microglia-Mediated Neuroinflammation and Neurodegeneration. *Front. Cell Neurosci.* **2020**, *14*, 162.

Chapter 2

Cotton Ti-IMAC: Developing Phosphorylated Cotton as a Novel Platform for Phosphopeptide Enrichment



Adapted from: **Wang, D.** [#]; Huang, J. [#]; Zhang, H.; Gu, T.-J.; Li, L. Cotton Ti-IMAC: Developing Phosphorylated Cotton as a Novel Platform for Phosphopeptide Enrichment. (To be submitted; [#]co-first authors) Author contribution: study was designed by D.W and J.H under the supervision of L.L.; experiment was performed by D.W. and H.Z.; data was analyzed by D.W.; manuscript was written by D.W. and J.H., and edited by H.Z., T.G., and L.L.

Abstract

Protein phosphorylation is one of the most common post-translational modifications (PTMs), which is involved in many important cellular functions. Understanding protein phosphorylation at the molecular level is critical to decipher its relevant biological processes and signaling networks. Mass spectrometry (MS) has been proven to be a powerful tool for comprehensive characterization of protein phosphorylation. Yet the low abundance and poor ionization efficiency of phosphopeptides in complex biological samples make its MS analysis challenging; an enrichment strategy with high efficiency and selectivity is always necessary before MS analysis. In this study, we developed a phosphorylated cotton fiber-based Ti(IV)-IMAC material (termed as: Cotton Ti-IMAC) that can serve as a novel platform for phosphopeptide enrichment. The cotton fiber can be effectively grafted with phosphate groups covalently in a single step, where the titanium ions can then be immobilized to enable capturing phosphopeptides. The material can be prepared using cost-effective reagents within only 4 hours. Benefiting from the flexibility and filterability of cotton fibers, the material can be easily packed as a spin-tip and make the enrichment process convenient. Cotton Ti-IMAC successfully enriched phosphopeptides from protein standard digests and exhibited a high selectivity (BSA/β-casein = 1000:1) and excellent sensitivity (0.1 fmol/μL). Moreover, 2354 phosphopeptides were identified in a single LC-MS/MS injection after enriching from only 100 μg HeLa cell digests with an enrichment specificity of up to 97.51%. Taken together, we believe that Cotton Ti-IMAC can serve as a widely applicable and robust platform for achieving large-scale phosphopeptide enrichment and expanding our knowledge of phosphoproteomics in complex biological systems.

Introduction

Protein phosphorylation is among the most prevalent protein post-translational modifications (PTM), as one-third of all proteins in eukaryotic cells are estimated to be phosphorylated at any time.[1] This reversible modification is involved in many cellular functions, such as protein synthesis, cell signaling, apoptosis, and enzymatic regulation.[2–4] Consequently, detailed analyses of protein phosphorylation at a molecular level are critical to decipher its relevant biological processes and signaling networks. Mass spectrometry (MS)-based methods have been widely applied to study phosphoproteome owing to the high sensitivity and high throughput capability.[5] However, the low abundance and ionization efficiency of phosphopeptides pose challenges for their direct analysis, especially in complex biological samples. To address this issue, an efficient enrichment method is necessary to separate the phosphopeptides from the bulk of unmodified peptides prior to MS analysis.

Various techniques have been developed for phosphopeptide enrichment so far.[6] Among them, immobilized metal affinity chromatography (IMAC) has become the most effective and widely used one. IMAC was first introduced in 1975 by Porath et al. to fractionate protein solutions based on the affinity of transition metal ions.[7] Since then the technique has been successfully adapted to phosphopeptide enrichment, where strong, specific interaction between metal ions and phosphate groups on phosphopeptides separate the phosphopeptides from other peptides.[8] Commonly immobilized metal ions include Fe^{3+}, Ga^{3+} and Ni^{2+}, which are bound to chelating ligands including iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA).[8–11] However, the

enrichment still suffer from non-specific binding of acidic peptides as the metal ion also binds to the carboxylic acid groups.[12] To enhance the enrichment selectivity, Zhou et al. discovered that using phosphate groups as binding ligands and Ti^{4+} as metal ions provides stronger metal-phosphate affinity, as in the MO_6 octahedra layer structure of metal(IV) ions, each metal ion coordinates to multiple phosphate molecules and *vice versa*.[13] Many Ti(IV)-IMAC materials have been reported since then.[14–19] Despite their excellent enrichment specificity, the synthesis and phosphate group functionalization of these nanomaterials or mesoporous polymeric materials are often expensive or require specialized expertise in material science, potentially limiting their applicability in a MS lab.

Cellulose, the most abundant natural polymer and a sustainable green resource, has been utilized for various materials and applications.[20] Examples include phosphorylated cellulose in flame retardant materials,[21,22] drug delivery matrices,[23] and metal ion adsorbents for wastewater purification.[24,25] Phosphorylation of cellulose material can be achieved through both chemical modification[22] and enzymatic phosphorylation[26]. Although both methods can effectively substitute hydroxyl groups on the cellulose surface, chemical methods are generally more cost-effective and faster.[27,28] The phosphorylation reaction of cellulose can be realized with various phosphorylation reagents, e.g., phosphoric acid, phosphate, polyphosphate, phosphorus pentoxide, and phosphoryl chloride with the assistance of urea in one step.[29–36] Inspired by the abundance of phosphate groups on the material surface, phosphorylated cellulose has also been used as IMAC material to enrich

phosphopeptides.[37,38] However, the preparation process of cellulose itself is still complicated, which hinders its application for large-scale MS-based analysis.

On the other hand, cotton fiber is the most common source of cellulose with good flexibility, strength, chemical resistance, and hydrophilicity. It has been used as the plug material in pipette tip-based solid-phase extractions,[39,40] and as hydrophilic stationary phase for glycans and glycopeptides enrichment through hydrophilic interaction liquid chromatography (HILIC).[41–43] Cotton fiber has also been reported to be grafted with carboxyl groups or phosphate groups to function as Ti-IMAC material for phosphopeptide enrichment.[44,45] However, the reported chemical modification steps in these cases are still relatively complicated.

To address the above limitations and develop a more accessible IMAC material, we have developed a phosphorylated cotton fiber-based Ti(IV)-IMAC material (termed as: Cotton Ti-IMAC) as a novel platform for phosphopeptide enrichment. Taking advantage of the convenient phosphorylation modification of cellulose materials with phosphoric acid and urea, the cotton fiber can be effectively grafted with phosphate groups in a single step. Titanium ions can then be immobilized onto the surface of the cotton fiber to capture phosphopeptides. In addition to the monophosphate groups, polyphosphate groups can also be introduced to the phosphorylated cotton, which facilitate to chelate more Ti^{4+} ions.[46,47] Moreover, these modifications maintain the cotton's properties, such as flexibility and filterability, enabling the material to be easily packed into a spin-tip and making the phosphopeptide enrichment procedures more convenient. Overall, the material can be prepared in just two steps within 4 hours, which do not require any specialized procedures

or equipment. Through the evaluation with standard protein digests, Cotton Ti-IMAC has been proven to have excellent selectivity and high sensitivity. Our method has also been applied to enrich phosphopeptides from HeLa cells, achieving an enrichment specificity as high as 97.51%. Taken together, we have developed a widely applicable, cost-effective, and robust platform for phosphopeptide enrichment that is suitable for large-scale LC-MS/MS analysis of complex biological samples.

Experimental Section

Chemicals. Dithiothreitol (DTT) and sequencing grade trypsin were from Promega (Madison, WI). Optima LC/MS grade solvents, formic acid (FA), urea, sodium chloride, ammonium hydroxide, 2,5-Dihydroxybenzoic acid (DHB), phosphate buffered saline (PBS), and fetal bovine serum were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoric acid, trifluoroacetic acid (TFA), iodoacetamide (IAA), triethylammonium bicarbonate (TEAB) sodium dodecyl sulfate (SDS), penicillin-streptomycin, bovine serum albumin (BSA), and α-casein and β-casein from bovine milk were purchased from Sigma-Aldrich (St Louis, MO). Dulbecco's Modified Eagle Medium (DMEM) was from Cytiva (Marlborough, MA). Titanium sulfate ($Ti(SO_4)_2$) was purchased from Sinopharm Chemical Reagent Co., Ltd (Shanghai, China). Sterile cotton was from First Aid Only (Shelton, CT). Empty 200 µL TopTips were from Glygen Corp (Columbia, MD). Protease inhibitor cocktail tablets and phosphatase inhibitor cocktail tablets were from Roche (Mannheim, Germany). Centrifuge-assisted extraction Ti-IMAC (CAE-Ti-IMAC) microspheres were obtained from J&K Scientific Ltd. (Beijing, China).

Preparation of Cotton Ti-IMAC. The phosphorylation of cotton fiber was performed based on previous literature with slight modification.[30] A total of 200 mg cotton fiber was wetted with a solution containing 49.6% urea (w/v), 18.4% phosphoric acid, and 32% water. Excess solution was squeezed out of the cotton. The cotton fiber was placed in a flask and heated at 150°C for 60 minutes, followed by thorough washing with distilled water. After squeezing out the extra water, the cotton fiber was incubated in 100 mM $Ti(SO_4)_2$ in 0.1% TFA for 2 h. The cotton fiber was then washed with 0.1% TFA solution for three times to remove excess Ti^{4+} and stored at 4°C for future use.

Characterization. Attenuated total reflectance Fourier-transform infrared spectroscopy (ATR FT-IR) spectra were measured with an Equinox 55/S FT-IR spectrophotometer (Bruker, Germany). Scanning electron microscope (SEM) images were obtained with a Zeiss Gemini 450 field emission scanning electron microscope (Carl Zeiss, Germany). Energy-dispersive spectroscopy (EDS) images were acquired by Thermo Noran Energy dispersive X-ray microanalysis system (Thermo Fisher Scientific, San Jose, CA).

Cell Lysate. The samples were prepared per previous work.[39] HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin, and incubated at 37 °C in a humidified chamber with 5% CO_2. Cells were harvested by washing with PBS buffer and the pellets were lysed in buffer consisting of 50 mM Tris base (pH 7.4), 4% SDS, 65 mM DTT, 175 mM NaCl. 1 phosphatase inhibitor tablet and 1 protease inhibitor tablet were added per every 10 mL of the lysis buffer. The lysate was sonicated with a probe sonicator in ice water bath at 50%

power with a 5s on/5s off pulse for 12 cycles. Cell lysates were centrifuged, and supernatant was collected and mixed with ice-cold precipitation buffer (acetone: ethanol: acetic acid=50: 50: 0.1, v/v) at a 1:5 ratio. The mixture was allowed to precipitate overnight at -20^{o}C. Protein pellets were collected by centrifugation at 18000 g and washed twice with the ice-cold precipitation buffer. The pellets were dried in the fume hood for 10 min and then redissolved in 8M urea/50 mM TEAB buffer (pH 8.0). Protein concentrations were determined using a BCA assay kit (Thermo Fisher Scientific, San Jose, CA).

Protein Digestion. Standard proteins or protein extracts from cell lysate were dissolved in 8M urea/50 mM TEAB buffer, reduced by adding DTT to a final concentration of 20 mM and incubated at 37 ^{o}C for 2h. Proteins were alkylated with 40 mM IAA at room temperature in the dark for 30 min. Alkylation was quenched by adding 20 mM DTT and incubating for another 10 min. The urea buffer was diluted to 1 M with 50 mM TEAB buffer. Proteins were initially digested with trypsin at a protein-to-enzyme ratio of 100:1 and incubated at 37 ^{o}C for 12 h. The same amount of trypsin was added again and incubated at 37 ^{o}C for another 4 h to achieve a final protein-to-enzyme ratio of 50:1. The digestion was stopped by adding TFA to a final concentration of 1%. The samples were stored at -80^{o}C for future use.

Phosphopeptide Enrichment. Cotton Ti-IMAC material (3 mg) was packed into an empty TopTip, which was placed on a 2 mL microcentrifuge tube using an adapter unit. The Cotton Ti-IMAC tip was equilibrated by adding 300 µL of 0.1% TFA solution and centrifuging at 200 g three times. Peptide samples were dissolved in 100 µL of loading buffer (40% ACN, 3% TFA) and

loaded onto the Cotton Ti-IMAC tip. The tip was centrifuged at 200 g for 2 min, and the flow-through was re-loaded four more times to ensure complete retention. The Cotton Ti-IMAC tip was washed three times each with washing buffer I (50% ACN, 6% TFA and 200 mM NaCl) and washing buffer II (30% ACN, 0.1% TFA) at 200 g for 2 min. Finally, phosphopeptides were eluted with consecutive additions of 150 µL of 1% NH_4OH (v/v), 150 µL of 5% NH_4OH (v/v), and 150 µL of 10% NH_4OH (v/v). The phosphopeptides enrichment procedures of CAE-Ti-IMAC followed the manufacturer's instructions. The eluent was pooled together and dried down *in vacuo* before MS analysis.

MALDI-TOF and NanoLC-MS/MS analysis. Standard protein samples were analyzed on a Bruker Rapiflex MALDI-TOF/TOF instrument (Bruker Daltonik, Bremen, Germany) with DHB matrix (25 mg/mL in 50%ACN/1%H_3PO_4 (v/v)). Ion source parameters were set as the following: laser energy at 80%, frequency at 5000 Hz, and each data collection comprising 2000 shots. HeLa cell samples were analyzed by LC–MS/MS with an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Samples were reconstituted in 20 mM citric acid/1% FA solution prior to MS analysis. Peptides were separated on a 15 cm length, 75 µm i.d. in-house packed BEH C18 (1.7 µm, 130 Å, Waters) capillary column with a 96 min gradient from 0 to 30% ACN (0.1% FA). The flow rate was set at 0.3 µL/min. MS1 scans of peptides were acquired from 300 to 1800 *m/z* at a resolution of 120 K, with the AGC target of 2E5 and maximum injection time of 50 ms. MS/MS data was collected in a top 20 data-dependent acquisition (DDA) mode, where

spectra with a charge state between 2–8 were selected for stepped higher-energy collision dissociation (HCD) fragmentation. The normalized collision energy (NCE) was set as 30 ± 8%. Other MS/MS parameters included resolution of 30 K, an AGC target of 5E4, and a maximum injection time of 150 ms.

Data Analysis. Raw data files were searched using MaxQuant software (Version 1.6.7.0) against UniProt *Mus musculus* reviewed database (January 2021, 17033 sequences).[48] Precursor ion mass tolerance was set as 4.5 ppm and fragment ion mass tolerance was set as 20 ppm. Carbamidomethylation on cysteine was set as a fixed modification, while oxidation on methionine and phosphorylation on serine, threonine, and tyrosine were set as variable modifications. Trypsin was set as the specific proteolytic enzyme, with up to two missed cleavages allowed. Peptides and proteins were filtered at FDR < 0.01, and a minimum score of 40 was accepted as confident peptide identification. The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD037549.[49] Gene ontology (GO) annotations of identified phosphoproteins were performed on the DAVID Bioinformatics Resources following the website instructions.[50]

Results and Discussion

Synthesis and Characterization of Cotton Ti-IMAC material. The Cotton Ti-IMAC material was prepared as illustrated in **Figure 1A**. In the presence of urea and phosphoric acid, the cellulose of the cotton fiber underwent phosphorylation through a dehydration condensation reaction between the hydroxy groups and phosphoric acid, which primarily occurred at the C2 and C6

positions of glucose.[46] Consequently, the reaction product may contain the following functional groups: PO_3^{2-}, P-O-P, and HPO_2^-.[47]

To confirm the phosphorylation of cotton fiber, both unmodified cotton fiber and phosphorylated cotton fiber was characterized using ATR-FTIR. As shown in **Figure 2**, the spectrum of the phosphorylated cotton exhibited a peak at 2366cm^{-1} which was attributed to P-H stretching in a phosphite group.[31] The peak at 1220 cm^{-1} was assigned to P-O-P vibrations. Peaks observed at 970 cm^{-1} and 1003 cm^{-1}, as well as the broader shoulder from 900cm^{-1} to 942 cm^{-1} on the main 1030 cm^{-1} peak, were evidence of P-O stretching vibrations.[34,51] In the range of 3200–3600 cm^{-1}, the OH stretching vibration band became more asymmetric due to the introduction of more acidic OH groups of phosphoric acids into the polymer.[35] The presence of these characteristic bands indicate the successful reaction between phosphoric acid and cotton fiber, with phosphate groups chemically bonding to the cellulose structure. Owing to the abundant phosphate groups in the modified cotton fiber, Ti^{4+} ions were easily immobilized onto the surface of the material via ionic interaction at room temperature.

Furthermore, SEM analysis was carried out to investigate the effect of these modifications on the cotton morphology. The SEM images (**Figure S1**) revealed that the fibrous morphology of the cotton fiber was not significantly affected during each step of the reaction, and the diameter of the cotton fiber remained around 15 µm. To examine the elemental composition, the Cotton Ti-IMAC fiber was characterized by EDS spectroscopy. As depicted in **Figure 3**, the Cotton Ti-IMAC fiber contained evenly distributed O, P, and Ti elements. The detailed composition of each element and the corresponding EDS spectrum are included as **Figure S2**. It should be noted that the atom

percentages of P and Ti were estimated to be 12.76% and 35.03%, respectively, indicating that the Ti^{4+} ions were successfully chelated to the cotton fiber, and the Cotton Ti-IMAC material had high Ti content that would contribute to a great binding capacity for phosphopeptides.

Evaluating the Phosphopeptide Enrichment Performance of Cotton Ti-IMAC. After the successful fabrication of Cotton Ti-IMAC material, it was first tested to enrich phosphopeptides from standard protein digests, including α-casein and β-casein. The protein digests were loaded onto the Cotton Ti-IMAC spin-tip with an acidic loading buffer (40% ACN, 3% TFA), and washed several times by washing buffer I (50% ACN, 6% TFA and 200 mM NaCl) and washing buffer II (30% ACN, 0.1% TFA) to remove non-specific bindings based on previous reported workflow.[17] The retained phosphopeptides were eventually eluted in an ammonia solution. As shown in **Figure 4A,** the signals on the MALDI-TOF spectrum were dominated by unmodified peptides, where only two singly phosphorylated peptides were detected. After the Cotton Ti-IMAC enrichment, most unmodified peptides were removed, and phosphopeptides could be detected with enhanced signals (**Figure 4B**). In total, 15 phosphopeptides were detected from the α-casein digests with greatly improved signal-to-noise (S/N) ratios, including 5 singly phosphorylated and 10 multiply phosphorylated peptides. Our identification result covered all phosphorylation sites reported in α-casein that were enriched by other IMAC or commercial TiO_2 materials.[16,19,52] The detailed information of identified phosphopeptides is listed in **Table S1**. The results demonstrate the ability of the modified cotton fibers to enrich phosphopeptides as an IMAC material.

As phosphopeptides are often of very low abundance in real samples, it is essential to prepare a Ti-IMAC material that possesses high enrichment selectivity and sensitivity. To examine the selectivity of Cotton Ti-IMAC material, a mixture of BSA and β-casein tryptic digests were used for the enrichment. BSA is a protein without any phosphorylation modification, so it was chosen as the background interference sample when enriching phosphopeptides from β-casein tryptic digests. When BSA and β-casein tryptic digests were mixed at a molar ratio of 100:1, it was hard to find any signals of phosphopeptides in the MALDI MS spectrum which were all suppressed by the high abundance unmodified peptides (**Figure 5A**). After enrichment, the signals from phosphopeptides were significantly improved with practically no interference from other peptides (**Figure 5B**). Three phosphopeptides including 2 multiply phosphorylated phosphopeptides were successfully separated from the mixture of peptides. The same experiment was also performed at higher molar ratios; the material exhibited remarkable selectivity even at molar ratios of 500:1 and 1000:1 (**Figure 5C, 5D**), where signals from both singly and multiply phosphorylated peptides could still be clearly observed.

To further investigate the sensitivity of the Cotton Ti-IMAC enrichment, this approach was applied to enrich phosphopeptides from diluted β-casein tryptic digests, with sample concentrations starting to decrease from 10 fmol/μL. The MALDI-MS spectra showed that signals from the phosphopeptides were still clearly observed with S/N over 3 when the concentration was as low as 0.1 fmol/μL (**Figure 6**). The results demonstrated that the abundant phosphate groups on the modified cotton fiber provided strong chelation to Ti^{4+} ions, which in turn contributed to

the strong affinity for phosphopeptides, thus greatly enhancing the sensitivity of phosphopeptide enrichment and subsequent detection. Taken together, the remarkable selectivity and sensitivity suggest exciting the potential of this material in enriching phosphopeptides from more complex biological samples.

Phosphopeptide Enrichment from HeLa cell lysate. In view of the outstanding performance of phosphopeptide enrichment in standard samples, the Cotton Ti-IMAC spin-tip approach was applied to enrich phosphopeptides from HeLa cell protein digests. The enrichment procedures were the same as the one performed on the standard samples. As a comparison, we also performed phosphopeptide enrichment from the same aliquots of HeLa cell protein digests with conventional CAE-Ti-IMAC material following the manufacturer's instructions. The enriched phosphopeptides were analyzed by LC-MS/MS, and the identification results were listed in **Table S2**. Generally, Cotton Ti-IMAC and CAE-Ti-IMAC identified 2354, and 2029 phosphopeptides, respectively, as an average of three technical replicates (**Figure 7A**). This demonstrated that the Cotton Ti-IMAC method could identify 300 more phosphopeptides in a single injection compared to the conventional enrichment approach. It is also worth noting that the numbers of multi-phosphopeptides identified by the two methods were comparable, which means that the Cotton-Ti-IMAC method identified more mono-phosphopeptides. Mono-phosphopeptides generally have a lower affinity for IMAC enrichment than multi-phosphopeptides. With more mono-phosphopeptide identified, Cotton Ti-IMAC demonstrated greater affinity for phosphopeptides than the conventional CAE-Ti-IMAC. Moreover, although both methods exhibited very high

enrichment specificity, the Cotton Ti-IMAC method (97.51%) still slightly outperformed the CAE-Ti-IMAC method (96.56%). The Cotton Ti-IMAC also showed a better reproducibility due to its simpler enrichment procedure. The complementarity of the phosphopeptides and phosphoproteins identified by the two approaches was also investigated. As shown in **Figure 7B** and **7C,** the combined results of three replicates demonstrated that the two methods enabled profiling of 5356 phosphopeptides in total corresponding to 1759 phosphoproteins, with an overlap of 2780 phosphopeptides (51.90%), and 1182 phosphoproteins (67.20%), respectively. According to **Figure 7B**, 1570 phosphopeptides were uniquely identified in the Cotton Ti-IMAC method while only 1006 unique phosphopeptides were presented in the CAE-Ti-IMAC dataset. The distribution of the phosphoprotein closely followed the pattern of the phosphopeptides, with 398 phosphoproteins uniquely identified in the Cotton Ti-IMAC dataset and 179 identified in the CAE-Ti-IMAC dataset, repectively (**Figure 7C**). These data clearly showed that the majority of phosphopeptides found in the CAE-Ti-IMAC dataset could be identified using the Cotton Ti-IMAC, and more importantly, the Cotton Ti-IMAC approach enabled profiling of more phosphopeptides and phosphoproteins.

To functionally categorize the profiled phosphoproteins, GO annotation of phosphoproteins identified in the Cotton Ti-IMAC method was performed based on three categories: biological process (**Figure 7D**), cellular component (**Figure 7E**), and molecular function (**Figure 7F**). The top 20 most significant categories were plotted. The results of biological processes showed that the most significant terms were associated with translation and transcription, representing major

biological processes of cancer cells. The top 5 significant cellular component annotation showed that the phosphoproteins were evenly distributed in the cell nucleus, cytosol, and membrane, which indicated that the phosphoprotein enrichment was unbiased for all cellular components. Molecular functional analysis of the identified phosphoproteins showed that the majority of the phosphoproteins were binding associated proteins, especially for protein and RNA binding, which was consistent with the fact that phosphorylation was an important signaling modification in cells.

Conclusions

In this study, a novel Cotton Ti-IMAC platform for phosphopeptide enrichment was developed based on the phosphorylation modification of cotton fibers. The synthesis of Cotton Ti-IMAC was facile, rapid, and cost-effective, as only sterile cotton, urea, phosphoric acid, and titanium sulfate were involved during the synthesis. The whole synthesis only requires two reaction steps which can be completed within a short time without any requirement for specialized procedures or equipment, indicating its broad applicability to any analytical or MS lab. The Cotton Ti-IMAC has been demonstrated to have excellent phosphopeptide enrichment performance on both standard protein digests and cell lysate digests, including extremely high selectivity and sensitivity. Owing to its good biocompatibility and high titanium content, the enrichment specificity could achieve 97.51% when analyzing complex samples. In addition, benefiting from the physical property of cotton fiber, the enrichment can be performed in various forms. Depending on the sample type or amount, the enrichment can be achieved either in a packed spin-tip, a packed pipette tip, or even just in-solution mode, which offers great flexibility to bioanalysis. Given the

numerous merits of the material, it is anticipated to serve as a widely applicable and robust platform for large-scale phosphopeptide enrichment, which can provide deeper understanding of phosphoproteome in any complex biological or clinical samples.

Acknowledgements

The authors would like to thank Gary Girdaukas from Analytical Instrumentation Center (AIC) at School of Pharmacy, University of Wisconsin-Madison for providing help in FT-IR analysis. This work was supported, in part, by the National Institutes of Health Grants U01CA231081, RF1AG052324, R01AG078794, and R01 DK071801 (to L.L.). Some of the mass spectrometers were acquired using NIH shared instrument grants S10 OD028473, S10 RR029531, and S10 OD025084 (to L.L.), and the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Reference

(1) Zolnierowicz, S.; Bollen, M. Protein Phosphorylation and Protein Phosphatases. *EMBO J.* **2000**, *19* (4), 483–488.

(2) Gjertsen, B. T.; Døskeland, S. O. Protein Phosphorylation in Apoptosis. *Biochim. Biophys. Acta* **1995**, *1269* (2), 187–199.

(3) Pawson, T.; Scott, J. D. Protein Phosphorylation in Signaling – 50 Years and Counting. *Trends Biochem. Sci.* **2005**, *30* (6), 286–290.

(4) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. L. The Crucial Role of Protein Phosphorylation in Cell Signaling and Its Use as Targeted Therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271–280.

(5) Yang, C.; Zhong, X.; Li, L. Recent Advances in Enrichment and Separation Strategies for Mass Spectrometry-based Phosphoproteomics. *Electrophoresis* **2014**, *35* (24), 3418–3429.

(6) Li, X.-S.; Yuan, B.-F.; Feng, Y.-Q. Recent Advances in Phosphopeptide Enrichment: Strategies and Techniques. *TrAC - Trends Anal. Chem.* **2016**, *78*, 70–83.

(7) Porath, J.; Carlsson, J.; OLSSON, I.; Belfrage, G. Metal Chelate Affinity Chromatography, a New Approach to Protein Fractionation. *Nature* **1975**, *258* (5536), 598–599.

(8) Han, G.; Ye, M.; Zou, H. Development of Phosphopeptide Enrichment Techniques for Phosphoproteome Analysis. *Analyst* **2008**, *133* (9), 1128–1138.

(9) Neville, D. C. A.; Townsend, R. R.; Rozanas, C. R.; Verkman, A. S.; Price, E. M.; Gruis, D. B. Evidence for Phosphorylation of Serine 753 in CFTR Using a Novel Metal-ion Affinity Resin and Matrix-assisted Laser Desorption Mass Spectrometry. *Protein Sci.* **1997**, *6* (11), 2436–2445.

(10) Posewitz, M. C.; Tempst, P. Immobilized Gallium(III) Affinity Chromatography of Phosphopeptides. *Anal. Chem.* **1999**, *71* (14), 2883–2892.

(11) Trojer, L.; Stecher, G.; Feuerstein, I.; Lubbad, S.; Bonn, G. K. Characterisation and Evaluation of Metal-Loaded Iminodiacetic Acid–Silica of Different Porosity for the Selective Enrichment of Phosphopeptides. *J. Chromatogr. A* **2005**, *1079* (1–2), 197–207.

(12) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. Phosphoproteome Analysis by Mass Spectrometry and Its Application to Saccharomyces Cerevisiae. *Nat. Biotechnol.* **2002**, *20* (3), 301–305.

(13) Zhou, H.; Ye, M.; Dong, J.; Han, G.; Jiang, X.; Wu, R.; Zou, H. Specific Phosphopeptide Enrichment with Immobilized Titanium Ion Affinity Chromatography Adsorbent for Phosphoproteome Analysis. *J. Proteome Res.* **2008**, *7* (9), 3957–3967.

(14) Yu, Z.; Han, G.; Sun, S.; Jiang, X.; Chen, R.; Wang, F.; Wu, R.; Ye, M.; Zou, H. Preparation of Monodisperse Immobilized Ti4+ Affinity Chromatography Microspheres for Specific Enrichment of Phosphopeptides. *Anal. Chim. Acta* **2009**, *636* (1), 34–41.

(15) Zhang, L.; Zhao, Q.; Liang, Z.; Yang, K.; Sun, L.; Zhang, L.; Zhang, Y. Synthesis of Adenosine Functionalized Metal Immobilized Magnetic Nanoparticles for Highly Selective and Sensitive Enrichment of Phosphopeptides. *Chem. Commun.* **2012**, *48* (50), 6274–6276.

(16) Zhao, L.; Qin, H.; Hu, Z.; Zhang, Y.; Wu, R.; Zou, H. A Poly(Ethylene Glycol)-Brush Decorated Magnetic Polymer for Highly Specific Enrichment of Phosphopeptides. *Chem. Sci.* **2012**, *3* (9), 2828–2838.

(17) Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J. R.; Zou, H.; Mohammed, S. Robust Phosphoproteome Enrichment Using Monodisperse Microsphere–Based Immobilized Titanium (IV) Ion Affinity Chromatography. *Nat. Protoc.* **2013**, *8* (3), 461–480.

(18) Su, J.; He, X.; Chen, L.; Zhang, Y. Adenosine Phosphate Functionalized Magnetic Mesoporous Graphene Oxide Nanocomposite for Highly Selective Enrichment of Phosphopeptides. *ACS Sustain. Chem. Eng.* **2018**, *6* (2), 2188–2196.

(19) Hong, Y.; Zhao, H.; Pu, C.; Zhan, Q.; Sheng, Q.; Lan, M. Hydrophilic Phytic Acid-Coated Magnetic Graphene for Titanium(IV) Immobilization as a Novel Hydrophilic Interaction Liquid Chromatography–Immobilized Metal Affinity Chromatography Platform for Glyco- and Phosphopeptide Enrichment with Controllable Selectivity. *Anal. Chem.* **2018**, *90* (18), 11008–11015.

(20) Wang, X.; Yao, C.; Wang, F.; Li, Z. Cellulose-Based Nanomaterials for Energy Applications. *Small* **2017**, *13* (42), 1702240.

(21) Ghanadpour, M.; Carosio, F.; Larsson, P. T.; Wågberg, L. Phosphorylated Cellulose Nanofibrils: A Renewable Nanomaterial for the Preparation of Intrinsically Flame-Retardant Materials. *Biomacromolecules* **2015**, *16* (10), 3399–3410.

(22) Fiss, B. G.; Hatherly, L.; Stein, R. S.; Friščić, T.; Moores, A. Mechanochemical Phosphorylation of Polymers and Synthesis of Flame-Retardant Cellulose Nanocrystals. *ACS Sustain. Chem. Eng.* **2019**, *7* (8), 7951–7959.

(23) Wu, S.; Gong, Y.; Liu, S.; Pei, Y.; Luo, X. Functionalized Phosphorylated Cellulose Microspheres: Design, Characterization and Ciprofloxacin Loading and Releasing Properties. *Carbohydr. Polym.* **2021**, *254*, 117421.

(24) Lehtonen, J.; Hassinen, J.; Kumar, A. A.; Johansson, L.-S.; Mäenpää, R.; Pahimanolis, N.; Pradeep, T.; Ikkala, O.; Rojas, O. J. Phosphorylated Cellulose Nanofibers Exhibit Exceptional Capacity for Uranium Capture. *Cellulose* **2020**, *27* (18), 10719–10732.

(25) Liu, P.; Borrell, P. F.; Božič, M.; Kokol, V.; Oksman, K.; Mathew, A. P. Nanocelluloses and Their Phosphorylated Derivatives for Selective Adsorption of Ag+, Cu2+ and Fe3+ from Industrial Effluents. *J. Hazard Mater.* **2015**, *294*, 177–185.

(26) Božič, M.; Liu, P.; Mathew, A. P.; Kokol, V. Enzymatic Phosphorylation of Cellulose Nanofibers to New Highly-Ions Adsorbing, Flame-Retardant and Hydroxyapatite-Growth Induced Natural Nanoparticles. *Cellulose* **2014**, *21* (4), 2713–2726.

(27) Illy, N.; Fache, M.; Ménard, R.; Negrell, C.; Caillol, S.; David, G. Phosphorylation of Bio-Based Compounds: The State of the Art. *Polym. Chem.* **2015**, *6* (35), 6257–6291.

(28) Noguchi, Y.; Homma, I.; Matsubara, Y. Complete Nanofibrillation of Cellulose Prepared by Phosphorylation. *Cellulose* **2017**, *24* (3), 1295–1305.

(29) Jurgens, J. F.; Reid, J. D.; Guthie, J. D. Phosphorylated Cotton Cellulose as a Cation-Exchange Material. *Text. Res. J.* **1948**, *18* (1), 42–44.

(30) Guthrie, J. D. Ion Exchange Cottons. *Industrial Eng. Chem.* **1952**, *44* (9), 2187–2189.

(31) Mucalo, M. R.; Yokogawa, Y.; Toriyama, M.; Suzuki, T.; Kawamoto, Y.; Nagata, F.; Nishizawa, K. Growth of Calcium Phosphate on Surface-Modified Cotton. *J. Mater. Sci.: Mater. Med.* **1995**, *6* (10), 597–605.

(32) Mucalo, M. R.; Yokogawa, Y.; Suzuki, T.; Kawamoto, Y.; Nagata, F.; Nishizawa, K. Further Studies of Calcium Phosphate Growth on Phosphorylated Cotton Fibres. *J. Mater. Sci.: Mater. Med.* **1995**, *6* (11), 658–669.

(33) Suflet, D. M.; Chitanu, G. C.; Popa, V. I. Phosphorylation of Polysaccharides: New Results on Synthesis and Characterisation of Phosphorylated Cellulose. *React. Funct. Polym.* **2006**, *66* (11), 1240–1249.

(34) Mucalo, M. R.; Kato, K.; Yokogawa, Y. Phosphorylated, Cellulose-Based Substrates as Potential Adsorbents for Bone Morphogenetic Proteins in Biomedical Applications: A Protein

Adsorption Screening Study Using Cytochrome C as a Bone Morphogenetic Protein Mimic. *Colloids Surf. B: Biointerfaces* **2009**, *71* (1), 52–58.

(35) Luneva, N. K.; Ezovitova, T. I. Cellulose Phosphorylation with a Mixture of Orthophosphoric Acid and Ammonium Polyphosphate in Urea Medium. *Russ. J. Appl. Chem.* **2014**, *87* (10), 1558–1565.

(36) Kokol, V.; Božič, M.; Vogrinčič, R.; Mathew, A. P. Characterisation and Properties of Homo- and Heterogenously Phosphorylated Nanocellulose. *Carbohyd. Polym.* **2015**, *125*, 301–313.

(37) Shen, F.; Hu, Y.; Guan, P.; Ren, X. Ti4+-Phosphate Functionalized Cellulose for Phosphopeptides Enrichment and Its Application in Rice Phosphoproteome Analysis. *J. Chromatogr. B* **2012**, *902*, 108–115.

(38) Zhang, L.; Wang, Y.; Pan, L.; Tang, R.; Asoh, T.-A.; Ou, J.; Uyama, H. Fabrication of a Reusable Bifunctional Biomimetic Ti 4+ -Phosphorylated Cellulose Monolith with a Coral-like Structure for Enrichment of Phosphorylated and Glycosylated Peptides. *Green Chem.* **2021**, *23* (19), 7674–7684.

(39) Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

(40) Wang, D.; Ma, M.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, M.; Wang, Z.; Zetterberg, H.; Li, L. Boost-DiLeu: Enhanced Isobaric N,N-Dimethyl Leucine Tagging Strategy for a Comprehensive Quantitative Glycoproteomic Analysis. *Anal. Chem.* **2022**.

(41) Selman, M. H. J.; Hemayatkar, M.; Deelder, A. M.; Wuhrer, M. Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal. Chem.* **2011**, *83* (7), 2492–2499.

(42) Peng, Y.; Lv, J.; Yang, L.; Wang, D.; Zhang, Y.; Lu, H. A Streamlined Strategy for Rapid and Selective Analysis of Serum N-Glycome. *Anal. Chim. Acta.* **2019**, *1050*, 80–87.

(43) Han, J.; Chen, Q.; Jin, W.; Zou, M.; Lu, Y.; Liu, Y.; Wang, C.; Wang, Z.; Huang, L. Purification of N- and O-Glycans and Their Derivatives from Biological Samples by the Absorbent Cotton Hydrophilic Chromatographic Column. *J. Chromatogr. A* **2020**, *1620*, 461001.

(44) He, X.-M.; Chen, X.; Zhu, G.-T.; Wang, Q.; Yuan, B.-F.; Feng, Y.-Q. Hydrophilic Carboxyl Cotton Chelator for Titanium(IV) Immobilization and Its Application as Novel Fibrous Sorbent for Rapid Enrichment of Phosphopeptides. *ACS Appl. Mater. Interfaces* **2015**, *7* (31), 17356–17362.

(45) He, X.-M.; Chen, X.; Yuan, B.-F.; Feng, Y.-Q. Graft Modification of Cotton with Phosphate Group and Its Application to the Enrichment of Phosphopeptides. *J. Chromatogr. A* **2017**, *1484*, 49–57.

(46) Zhao, M.; Fujisawa, S.; Saito, T. Distribution and Quantification of Diverse Functional Groups on Phosphorylated Nanocellulose Surfaces. *Biomacromolecules* **2021**, *22* (12), 5214–5222.

(47) Rol, F.; Sillard, C.; Bardet, M.; Yarava, J. R.; Emsley, L.; Gablin, C.; Léonard, D.; Belgacem, N.; Bras, J. Cellulose Phosphorylation Optimization and Analysis of Phosphorus Position on Cellulose Fibers. *Carbohyd. Polym.* **2019**, *229*, 115294.

(48) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301–2319.

(49) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (Database issue), D442–D450.

(50) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* **2009**, *4* (1), 44–57.

(51) Jastrzębski, W.; Sitarz, M.; Rokita, M.; Bułat, K. Infrared Spectroscopy of Different Phosphates Structures. *Spectrochim. Acta - A: Mol. Biomol.* **2011**, *79* (4), 722–727.

(52) Hong, Y.; Yao, Y.; Zhao, H.; Sheng, Q.; Ye, M.; Yu, C.; Lan, M. Dendritic Mesoporous Silica Nanoparticles with Abundant Ti 4+ for Phosphopeptide Enrichment from Cancer Cells with 96% Specificity. *Anal Chem* **2018**, *90* (12), 7617–7625.

Figures

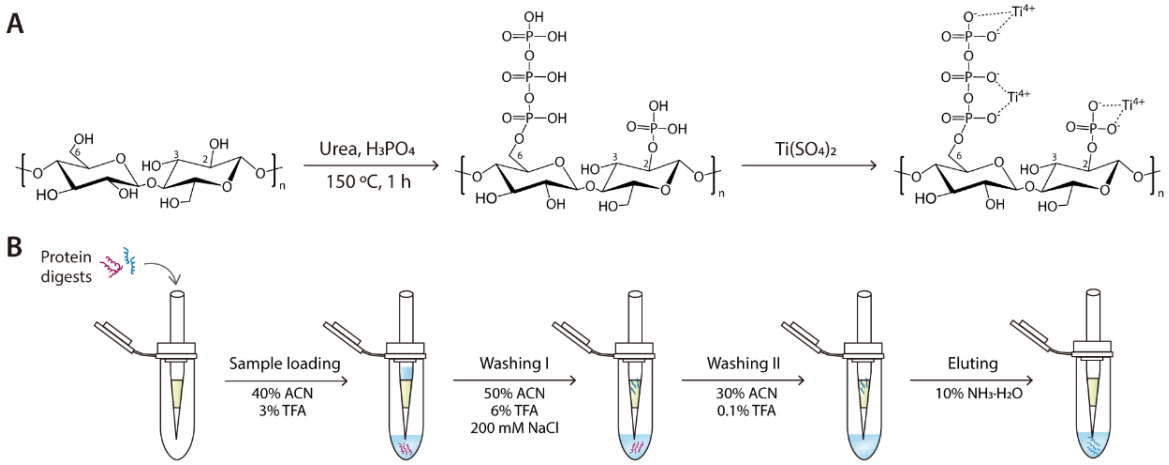


Figure 1. (A) Illustration of procedure for the preparation of Cotton Ti-IMAC; (B) Workflow of phosphopeptide enrichment by Cotton Ti-IMAC on a spin-tip.

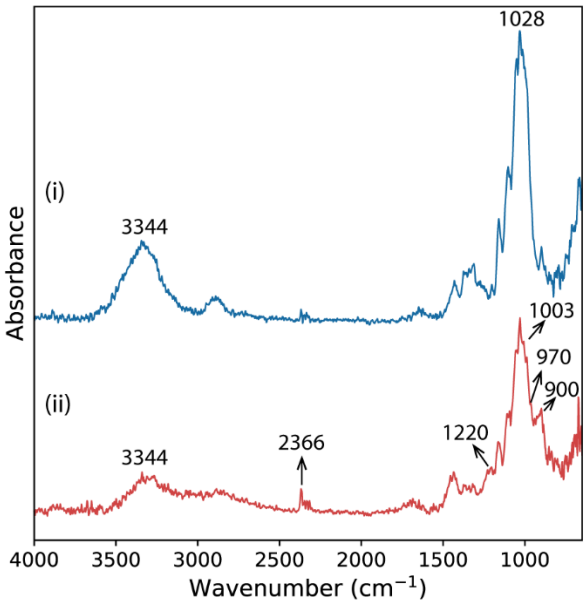


Figure 2. FT-IR characterization of the cotton fiber before (i) and after (ii) the phosphorylation modification.

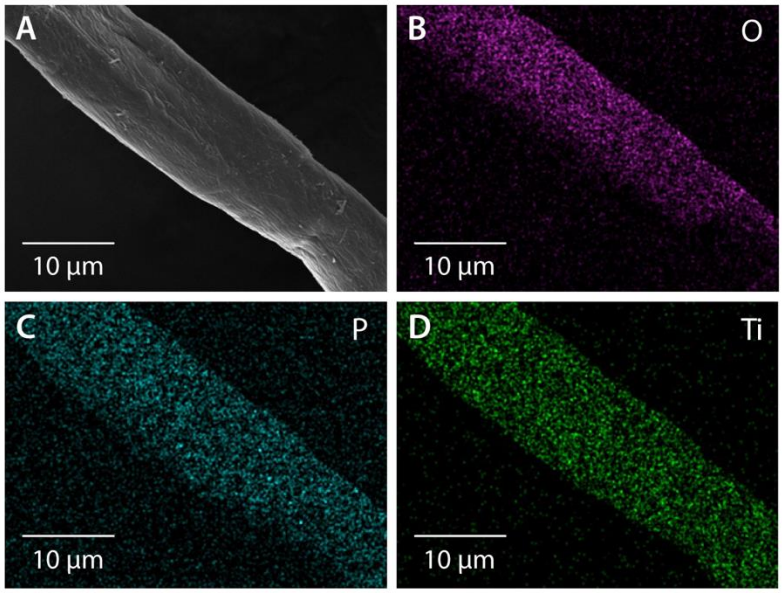


Figure 3. EDS elemental mapping images of Cotton Ti-IMAC. (A) SEM image, (B) O element, (C) P element, and (D) Ti element.

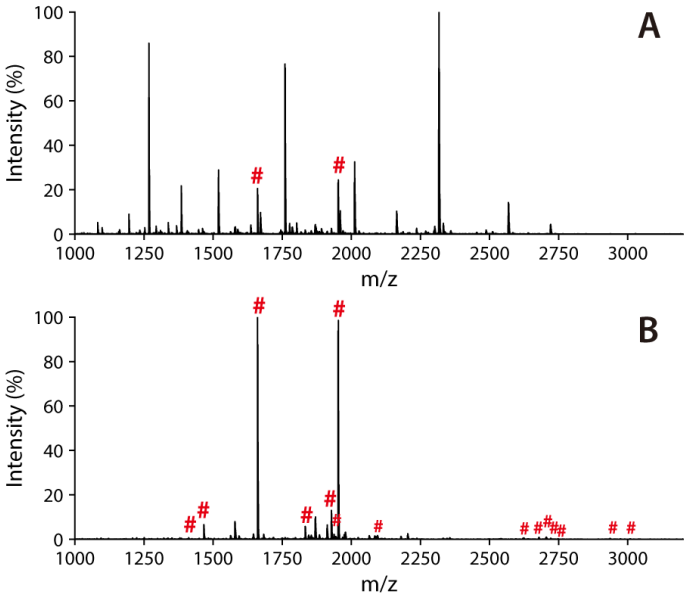


Figure 4. MALDI-TOF mass spectra of tryptic digests from α-casein (10 μg) before (A) and after (B) phosphopeptide enrichment by Cotton Ti-IMAC. Phosphopeptides are labeled with red "#".

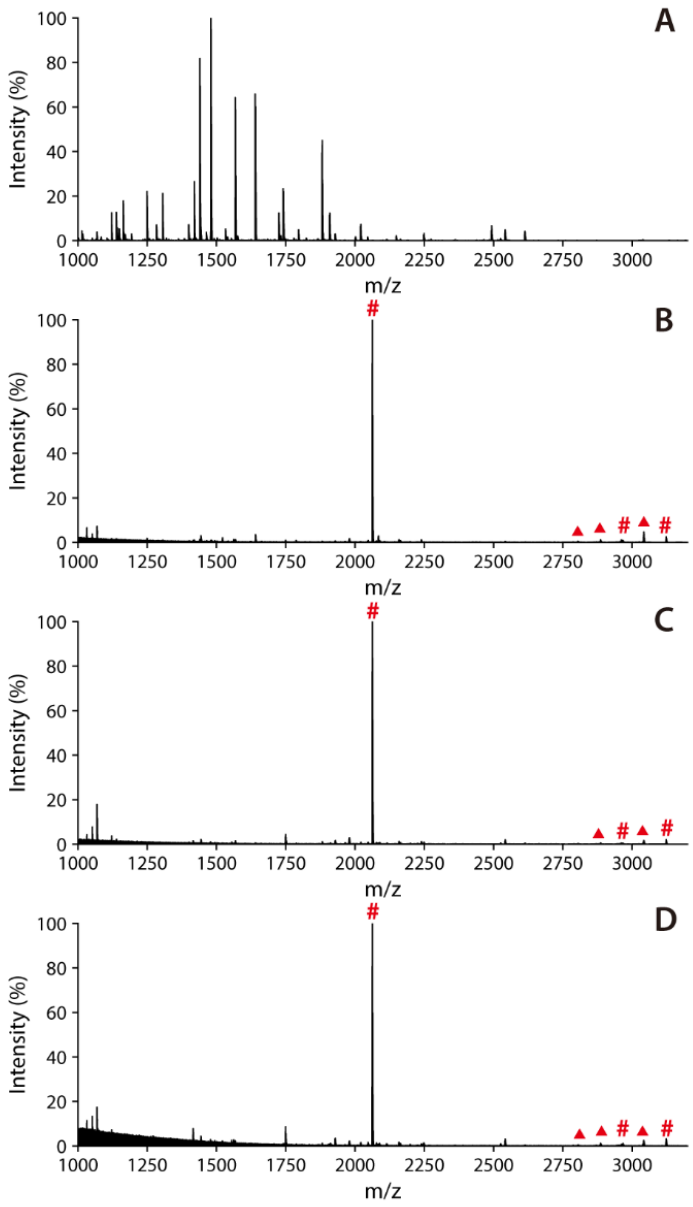


Figure 5. MALDI-TOF mass spectra of phosphopeptides enriched from mixture of BSA and β-casein digests at different molar ratios. (A) Direct analysis at 100:1 ratio; After enrichment at (B) 100:1, (C) 500:1 and (D) 1000:1 ratio. Phosphopeptides are labeled with red "#", neutral loss is labeled with red "▲".

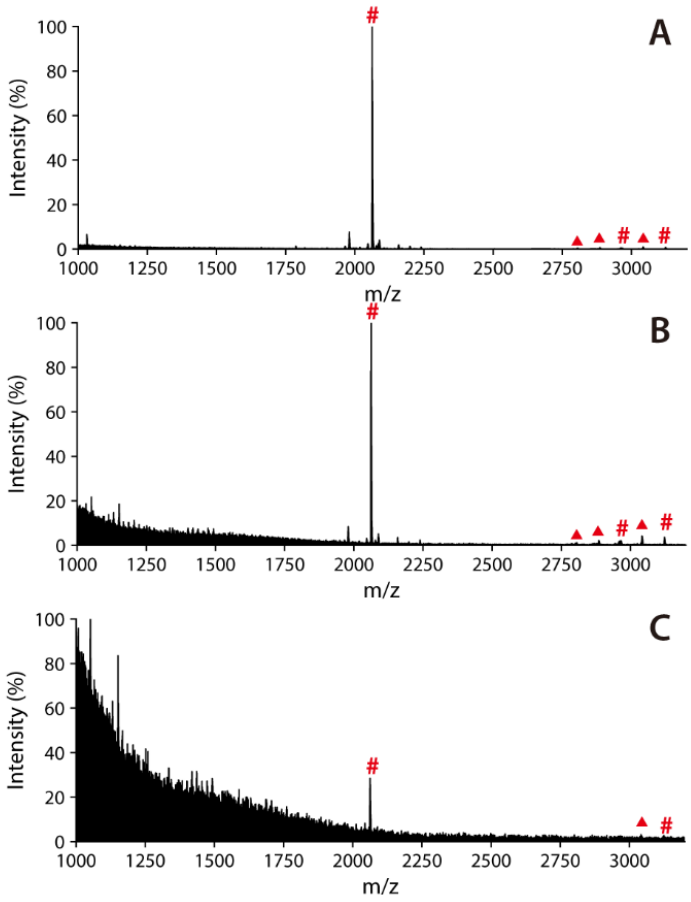


Figure 6. MALDI-TOF mass spectra of phosphopeptides enriched from β-casein digests with Cotton Ti-IMAC at different concentrations: (A) 10 fmol/μL, (B) 1 fmol/μL, and (C) 0.1 fmol/μL. Phosphopeptides are labeled with red "#", neutral loss is labeled with "▲".

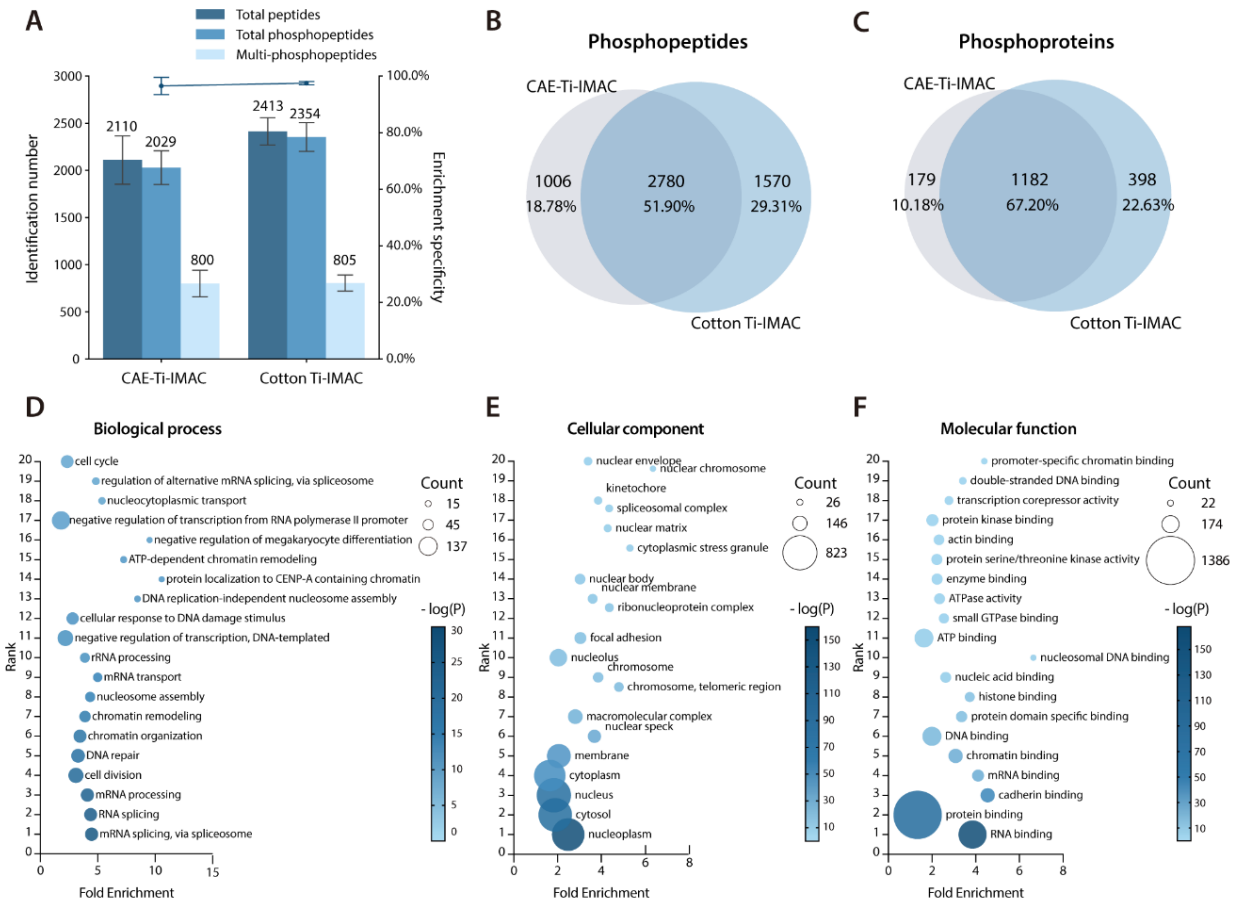


Figure 7. LC-MS/MS analysis of phosphopeptides enriched from HeLa cell tryptic digests. (A) Comparison of phosphopeptide identification number and enrichment specificity by CAE-Ti-IMAC and Cotton Ti-IMAC enrichment, (B) Overlap of phosphopeptides enriched by the two methods, (C) Overlap of phosphoproteins identified by the two methods. Gene ontology analysis of phosphoproteins identified in the Cotton Ti-IMAC method based on (D) biological process, (E) cellular component, and (F) molecular function. Top 20 most significant categories were plotted.

Supplemental Information

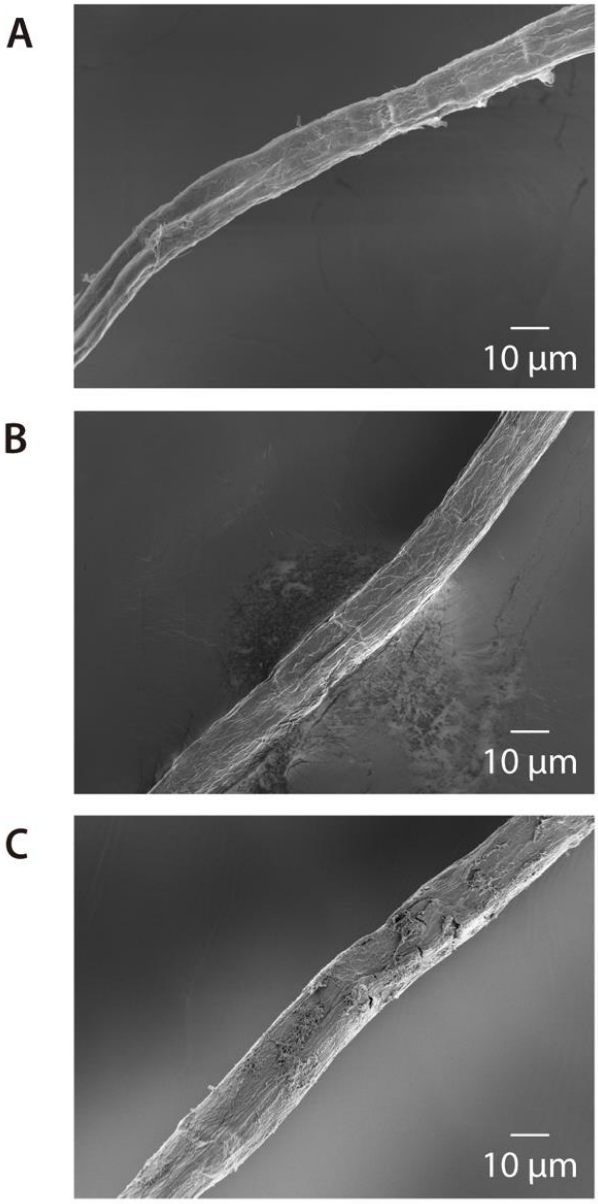
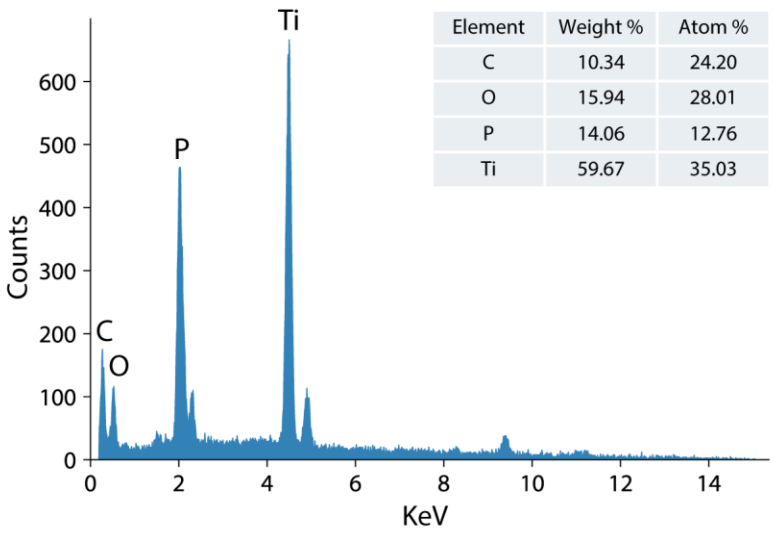


Figure S1. SEM images of (A) unmodified cotton, (B) phosphorylated cotton, and (C) Ti^{4+} immobilized phosphorylated cotton.



Element	Weight %	Atom %
C	10.34	24.20
O	15.94	28.01
P	14.06	12.76
Ti	59.67	35.03

Figure S2. EDS spectrum and chemical composition of cotton Ti-IMAC.

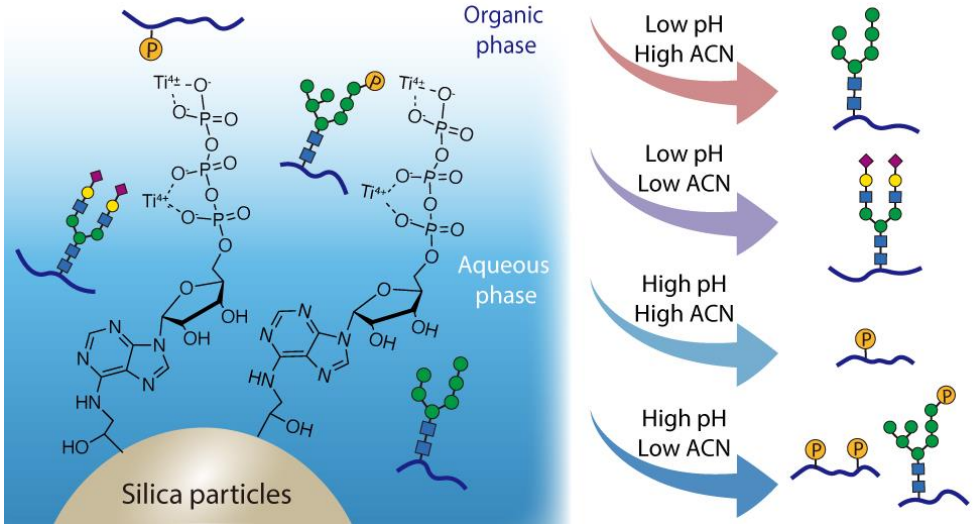
Table S1. MALDI-TOF MS analysis of phosphopeptides enriched from α-casein tryptic digests.

Protein	Amino acid sequence	No. of phosphosites	Theoretical $[M+H]^+$	Observed $[M+H]^+$
α-casein	EQL[pS]T[pS]EENSK	2	1411.5026	1411.5323
	TVDME[pS]TEVFTK	1	1466.6120	1466.6721
	VPQLEIVPN[pS]AEER	1	1660.7941	1660.8747
	YLGEYLIVPN [pS]AEER	1	1832.8465	1832.9387
	DIG[pS]E[pS]TEDQAMEDIK	2	1927.6915	1927.7422
	DIG[pS]E[pS]TEDQA[Mo]EDIK	2	1943.6864	1943.7266
	YKVPQLEIVPN[pS]AEER	1	1951.9524	1952.0486
	NMAINP[pS]KENLCSTFCK	1	2093.8855	2093.9559
	NTMEHV[pS][pS][pS]EE[pS]IISQETYK	4	2618.9047	2618.8628
	VNEL[pS]KDIG[pS]E[pS]TEDQAMEDIK	3	2678.0228	2677.8536
	Q*MEAE[pS]I[pS] [pS] [pS]EEIVPN[pS]VEAQK	5	2703.8864	2703.7165
	QMEAE[pS]I[pS][pS][pS]EEIVPNPN[pS]VEQK	5	2720.9129	2720.7054
	NTMEHV[pS][pS][pS]EE[pS]IISQETYKQ	4	2746.9997	2746.7533
	EKVNEL[pS]KDIG[pS]E[pS]TEDQAMEDIK	3	2935.1604	2934.7824
	NANEEEYSIG[pS][pS][pS]EE[pS]AEVATEEVK	4	3008.0295	3007.5083

[a][pS], phosphorylated site; [b][Mo], oxidation on methionine; [c]Q*, pyroglutamylation on the N-terminal Gln.

Chapter 3

ATP-Coated Dual-functionalized Titanium (IV) IMAC Material for Simultaneous Enrichment and Separation of Glycopeptides and Phosphopeptides



Adapted from: **Wang, D.**[#]; Huang, J.[#]; Zhang, H.; Ma, M.; Xu, M.; Cui, Y.; Shi, X.; Li, L. ATP-Coated Dual-functionalized Titanium (IV) IMAC Material for Simultaneous Enrichment and Separation of Glycopeptides and Phosphopeptides. *J. Proteome Res.* **2023**, Accepted. ([#]co-first authors) Author contribution: study was designed by D.W and J.H under the supervision of L.L.; experiment was performed by D.W. and H.Z.; data was analyzed by D.W.; samples were provided by M.M. and X.S.; manuscript was written by D.W. and J.H., and edited by M.X., Y.C., and L.L.

Abstract

Protein glycosylation and phosphorylation are two of the most common post-translational modifications (PTMs), which play an important role in many biological processes. However, low abundance and poor ionization efficiency of phosphopeptides and glycopeptides make direct MS analysis challenging. In this study, we developed a hydrophilicity-enhanced bifunctional Ti-IMAC material with grafted adenosine triphosphate (denoted as: epoxy-ATP-Ti^{4+}) to enable simultaneous enrichment and separation of common N-glycopeptides, phosphopeptides, and M6P glycopeptides. The enrichment was achieved through a dual-mode mechanism based on the electrostatic and hydrophilic properties of the material. The epoxy-ATP-Ti^{4+} IMAC material was prepared from epoxy-functionalized silica particles via a convenient two-step process. The ATP molecule provided strong and active phosphate sites for binding phosphopeptides in conventional IMAC mode and also contributed significantly to the hydrophilicity, which permitted the enrichment of glycopeptides via HILIC. The two modes could be implemented simultaneously, allowing glycopeptides and phosphopeptides to be collected sequentially in a single experiment from the same sample. Besides standard protein samples, the material was further applied to glycopeptide and phosphopeptide enrichment and characterization from HeLa cell digests and mouse lung tissue samples. In total, 2928 glycopeptides and 3051 phosphopeptides were identified from mouse lung tissue sample, supporting the utility of this material for large-scale PTM analysis of complex biological samples. Overall, the newly developed epoxy-ATP-Ti^{4+} IMAC material and associated fractionation method enable simple and effective enrichment and separation of glycopeptides and

phosphopeptides, offering a useful tool to study potential crosstalk between these two important PTMs in biological systems.

Introduction

As two of the most commonly studied post-translational modifications (PTMs), protein glycosylation and phosphorylation play an important role in many biological processes. Both PTMs are covalently added to specific protein sites through enzymatic modifications. It is estimated that one-third of all proteins are phosphorylated in mammalian cells at any time.[1] The reversible phosphorylation of proteins is usually involved in the regulation of protein synthesis, cell division, signal transduction, cell growth, development and aging.[2,3] Glycosylation is critical for physiological and pathological cellular functions and almost half of all proteins are glycosylated.[4,5] As abnormal phosphorylation or glycosylation have been implicated in various human diseases, it is critical to study these PTMs in a site-specific manner on the corresponding proteins. Liquid chromatography tandem mass spectrometry (LC-MS/MS) has become an essential tool for profiling PTMs on a large scale.[6] However, low abundance and poor ionization efficiency of phosphopeptides and glycopeptides largely hinder their direct MS analysis.[7] To overcome such limitations, an efficient enrichment is necessary prior to MS analysis. Typical enrichment techniques for phosphopeptides include metal oxide affinity chromatography (MOAC) and immobilized metal affinity chromatography (IMAC).[8] Common enrichment strategies for glycopeptides involve hydrazide chemistry enrichment, lectin affinity chromatography, boronic acid enrichment and hydrophilic interaction chromatography (HILIC).[9,10] Despite their excellent enrichment efficiency, these techniques are mainly used to target just one type of PTM.

It is worth noting that PTMs do not exist in isolation, as proteins are likely to be modified at several sites simultaneously. “Crosstalk” between PTMs reveals additional information about functional protein regulation and cellular processes.[11] For example, phosphorylation and N-linked glycosylation have been implicated in various diseases, including diabetes, cancer, and Alzheimer’s disease (AD). There is involved crosstalk between them, including competitive or noncompetitive occupancy of proximal sites, glycosylation of kinases, and phosphorylation of glycosylation-related enzymes.[12] To investigate the crosstalk of two PTMs, a single enrichment technique may not be capable of capturing both modifications. To address this issue, sequential enrichment was adopted by several studies.[13–16] However, this requires tedious sample preparation, multiple enrichment materials, and a large sample quantity by applying two different enrichment strategies. It is highly desirable to develop a new strategy to achieve simple and effective enrichment of both glycopeptides and phosphopeptides from the same sample. Previously, we explored the hydrophilic properties of commercially available centrifuge-assisted-extraction Titanium (IV) IMAC (CAE-Ti-IMAC) material and its utility for simultaneously enriching phosphopeptides, glycopeptides, and mannose-6-phosphate (M6P) glycopeptides through a dual-mode enrichment.[17,18] While we achieved great glyco- and phospho-proteome coverage, we found that the hydrophilic properties of the material were inadequate, since the material was not specifically developed for HILIC. New materials with better hydrophilicity and affinity for phosphate groups would be desirable to enhance the simultaneous enrichment workflow. Recent studies have explored new materials such as smart polymer-based or nanocomposite materials that possess great hydrophilicity and excellent affinity for phosphopeptides via switchable hydrogen

bond and surface charge,[19–22] positively charged metal ions,[23–29] or Lewis acid-base interaction.[30–32] These materials can both retain both phosphopeptides and glycopeptides; however, some studies only examined simple samples using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) without assessing their utility for complex biological samples. Additionally, glycopeptides and phosphopeptides coeluted in one fraction for some materials, which may cause interference during MS analysis. Furthermore, many studies performed deglycosylation for glycopeptides and did not discuss the availability of adopting a fractionation strategy for their material, which lost the site-specific information of glycopeptides and limited the PTM coverage, especially for peptides with negatively charged glycans or multiple phosphates. Most importantly, fabrication of these materials requires complicated reactions and expertise in material science, which may limit their accessibility and availability in a MS lab.

To address these issues, we aimed at developing a hydrophilicity-enhanced Ti(IV)-IMAC material which is easy to prepare and applicable to large-scale profiling of glycosylation and phosphorylation via LC-MS/MS. Adenosine triphosphate (ATP) is an important signaling biomolecule that also possesses properties desirable for this application. ATP has three phosphate groups that can chelate heavy metal ions, which in turn can bind to phosphopeptides. Previous studies have used ATP as a ligand in IMAC materials and achieved excellent selectivity and sensitivity in phosphopeptide enrichment.[33–36] However, the strong hydrophilicity of ATP provided by its hydrophilic adenine base and ribose sugar groups has not been exploited, which can superimpose the HILIC-mode enrichment upon the electrostatic attraction. The structure and good biocompatibility of ATP make it a promising ligand for the purpose.

To take advantage of the properties of ATP, we have developed an ATP-grafted IMAC material (Epoxy-ATP-Ti^{4+}) with facile preparation as shown in **Scheme 1A**. The epoxy-ATP-Ti^{4+} IMAC material was prepared by coupling ATP to a commercially available epoxy-functionalized silica and then immobilizing titanium ions. The preparation can be achieved in only two steps and does not require specialized techniques or equipment. Enrichment of glycopeptides and phosphopeptides were tested and evaluated using both standard protein digests and complex biological samples, including cell lines and mouse tissues (**Scheme 1B**). The epoxy-ATP-Ti^{4+} IMAC material was found to be a promising tool for simultaneous enrichment and separation of glycopeptides and phosphopeptides, with minimal interference between the fractions. Its considerable hydrophilicity and metal ion affinity make it an excellent platform for analyzing phosphorylation and N-glycosylation, and their potential crosstalk.

Experimental Section

Chemicals. Dithiothreitol (DTT) and sequencing grade trypsin were from Promega (Madison, WI). Optima LC/MS grade solvents, formic acid (FA), urea, sodium chloride, phosphoric acid, ammonium hydroxide, sodium carbonate and 2,5-Dihydroxybenzoic acid (DHB) were from Fisher Scientific (Pittsburgh, PA). Trifluoroacetic acid (TFA), iodoacetamide (IAA), triethylammonium bicarbonate (TEAB) and sodium dodecyl sulfate (SDS), bovine serum albumin (BSA), ribonuclease B from bovine pancreas (RNase B), and α-casein and β-casein from bovine milk were purchased from Sigma-Aldrich (St Louis, MO). Adenosine-5'-triphosphate (ATP) was from DOT Scientific Inc (Burton, MI). Titanium sulfate was purchased from Sinopharm Chemical Reagent Co., Ltd (Shanghai, China). Oasis HLB 1 cc (10 mg) extraction cartridges were purchased from

Waters Corporation (Milford, MA). Epoxy-functionalized silica bulk material (12 µm, 300 Å) was obtained from PolyLC (Columbia, MD). Empty 200 µL TopTips were from Glygen Corp (Columbia, MD). Protease inhibitor cocktail tablets and phosphatase inhibitor cocktail tablets were from Roche (Mannheim, Germany). CAE-Ti-IMAC microspheres were obtained from J&K Scientific Ltd. (Beijing, China).

Synthesis of Epoxy-ATP-Ti^{4+} IMAC Material. 1 g epoxy-functionalized silica was incubated in 1.5 mL ATP solution (0.1 g/mL in 0.5 M Na_2CO_3, pH > 8) at 65°C for 16 hours. The material was then washed with 50 mM TEAB buffer and water three times. Ti^{4+} ions were immobilized onto the epoxy-ATP functionalized silica beads by incubating the material in 100 mM $Ti(SO_4)_2$ in 0.1% TFA for 2 h. The material was washed with 0.1% TFA solution for three times and stored in 0.1% TFA solution at 4°C for future use.

Characterization. Scanning electron microscope (SEM) images were obtained using Zeiss Gemini 450 field emission scanning electron microscope (Carl Zeiss, Germany). Fourier-transform infrared spectroscopy (FT-IR) spectra were measured with an Equinox 55/S FT–IR spectrophotometer using KBr pellets (Bruker, Germany). Energy-dispersive spectroscopy (EDS) images were obtained by Thermo Noran Energy dispersive X-ray microanalysis system (Thermo Fisher Scientific, San Jose, CA).

Sample Preparation. The samples were prepared per previous work.[18] The details of cell culture, protein extraction and protein digestion are provided in Supplemental Information.

Simultaneous Enrichment of N-glycopeptides and Phosphopeptides. The protocol of conventional single-mode phosphopeptide enrichment followed previous publications with slight

modification.[18,37] The details are provided in Supplemental Information. For simultaneous glycopeptide and phosphopeptide enrichment, protein standard digest mixtures (10 µg each sample), HeLa cell digest (100 µg) or mouse lung tissue digest (500 µg) were loaded onto a Ti-IMAC-cotton tip in loading buffer (80% ACN, 3% TFA). The flow-through was collected and reloaded five more times. The IMAC-cotton tip was washed with 300µL loading buffer and centrifuged at 200 g for 2 min, which was repeated for six times to remove non-modified peptides. Different numbers of elution fractions were collected at 200 xg centrifugation speed depending on the sample type. For protein standard digests, two fractions were collected sequentially: 1) 300 µL 0.1% TFA and 2) 300 µL 10% NH_4OH. For HeLa cell digests, three fractions were collected in acidic conditions and one fraction was collected in basic conditions: 1) 300 µL 50% ACN/ 0.1% FA; 2) 300 µL 0.1% FA; 3) 300 µL 40% ACN/3% TFA, 300 µL 50% ACN/6% TFA/200 mM NaCl, 300 µL 30% ACN/0.1% TFA (these three elutions were pooled together as one fraction and desalted by HLB cartridge) and 4) 300 µL 10% NH_4OH. For mouse lung sample, six fractions were collected for analysis. Fraction 1 was eluted by 300 µL 60% ACN/0.1% FA. Fraction 2 was eluted by 300 µL 40% ACN/0.1% FA. Fraction 3 was eluted by 300 µL 20% ACN/0.1% FA and 300 µL 0.1% FA solution. Fraction 4 contained pooled eluates of 300 µL 40% ACN/3% TFA, 300 µL 50% ACN/6% TFA/200 mM NaCl and 300 µL 30% ACN/0.1% TFA and this fraction was desalted by a HLB cartridge. Fraction 5 was eluted by 300 µL 40% ACN/10% NH_4OH and fraction 6 consisted of eluates from 300 µL 20% ACN/10% NH_4OH, 300 µL 10% ACN/10% NH_4OH and 300 µL 10% NH_4OH. The IMAC-cotton tip was reconditioned with 300 µL 90% ACN/2.5%

NH_4OH for three times between fraction 4 and 5. Samples were dried down *in vacuo* before MS analysis.

MALDI-TOF and NanoLC-MS/MS analysis. Standard samples were analyzed on a Bruker Rapiflex MALDI-TOF/TOF instrument (Bruker Daltonik, Bremen, Germany) with DHB matrix (25 mg/mL in 50%ACN/1%H_3PO_4 (v/v)). Ion source parameters: laser energy 80%, frequency 5000 Hz and each data collection of 2000 shots. HeLa cell samples and mouse lung tissue samples were analyzed by LC–MS/MS with an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Glycopeptide fractions eluted in acidic conditions were reconstituted in 0.1% FA and phosphopeptide fractions collected in basic conditions were reconstituted in 20 mM citric acid/1% FA solution prior to MS analysis. Peptides were separated on a 15 cm length, 75 μm i.d. in-house packed BEH C18 (1.7 μm, 130 Å, Waters) capillary column with an 80 min gradient from 0 to 30% ACN (0.1% FA). The flow rate was set at 0.3 μL/min. For N-glycopeptides analysis, data acquisition was performed in top speed mode with a cycle time of 3s. Survey scans of peptide precursors from m/z 400 to 2000 were performed at resolving power of 120 K and AGC target of 4E5 with a maximum injection time (IT) of 100 ms. The precursors were selected for higher-energy C-trap dissociation (HCD) analysis with a normalized collision energy (NCE) of 30 and ± 3% stepped HCD collision energy. Tandem MS acquisition was at resolving power of 60 K, lower mass limit of 120 m/z, and dynamic exclusion of 12 s with 10 ppm mass tolerance. For phosphopeptide analysis, the MS scan was acquired from m/z 300 to 1800 at 120 K resolution and AGC target of 2E5 with maximum IT of 50 ms. MS/MS method was a top 20 data-dependent

acquisition (DDA) mode in which all MS/MS dissociations were performed with NCE of 30 and ± 8% stepped HCD collision energy. MS/MS parameters include resolution of 30 K, AGC target of 5E4, and maximum IT of 150 ms.

Data Analysis. Byonic software (version 2.9.38, Protein Metrics Inc, San Carlos, CA) was used to process intact N-glycopeptide data. Raw files were searched against UniProt *Homo sapiens* reviewed database (August 2020, 20311 sequences) or UniProt *Mus musculus* reviewed database (January 2021, 17033 sequences). 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. Fixed modifications were specified as carbamidomethylation (+57.02146 Da) on cysteine residues. Dynamic modifications included oxidation of methionine (+15.99492 Da, rare1) and N-glycosylation (common1). Glycan modifications were searched against a glycan database expanded from Byonic embedded human N-glycan database (182 entries) to include N-linked M6P glycans consisting of HexNAc (2-4) Hex (3-9) Phospho (1-2) modifications. Peptide identification results were filtered at Byonic score > 150, PEP 2D < 0.05, |Log Prob| > 1 and Delta Mod Score > 10. MS/MS spectra of M6P glycopeptide PSMs were manually inspected to ascertain the existence of phosphorylated hexose diagnostic ions. The same raw files and database were also searched on MaxQuant software (Version 1.6.7.0).[38] Precursor ion mass tolerance of 4.5 ppm and fragment ion mass tolerance of 20 ppm were set. Carbamidomethylation on cysteine was set as a fixed modification, whereas oxidation on methionine and phosphorylation on serine, threonine, and tyrosine were set as the variable modifications. Trypsin was set as the specific proteolytic enzyme with up to two missed

cleavages allowed. Both peptides and proteins were filtered at FDR < 0.01 and a minimum score of 40 was accepted as confident peptide identification. The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029775.[39]

Results and Discussion

Synthesis and Characterization of Epoxy-ATP-Ti^{4+} IMAC material. The synthetic procedure of epoxy-ATP-Ti^{4+} IMAC material is presented in **Scheme 1A.** To develop a facile and widely applicable workflow for synthesizing this IMAC material, we began with commercially available porous epoxy-functionalized silica particles with average diameter of 12 μm and pore size of 300 Å. ATP reacts with the epoxy groups and forms a covalent bond through the amine group of the adenine ring in a single reaction.[35,40] Ti^{4+} ions were then easily chelated onto the ATP-modified material surface via ionic interaction at room temperature.

The morphology of the synthesized material was characterized by SEM. **Figure 1A**, **1B and S1** showed the spherical shape and mesoporous structure of the silica particles after reaction with ATP and after Ti^{4+} ion immobilization, suggesting that the chemical reactions did not affect the overall shape and surface structure of the modified silica particles. To confirm the successful preparation of epoxy-ATP material, FTIR was used to characterize the functional groups of epoxy-silica particles before and after reaction with ATP (**Figure 1C**). The band at 1650 cm^{-1} was attributed to the C=N stretching vibration from the adenine ring in the ATP molecule.[41] The band at 1417 cm^{-1} further confirmed the presence of the CH_2-N units.[35] The broad peak at approximately 3430 cm^{-1} was enhanced by the stretching vibrations of O-H groups from phosphates and N-H

stretching vibrations.[42,43] The wide and strong absorption bands at around 1105 cm^{-1} were attributed to the asymmetric stretching vibration of Si-O-Si bond from the silica, which overlapped with the characteristic peaks from the P-O bond in the ATP after modification.[40] Overall, the FTIR spectra indicated ATP was covalently attached to the epoxy-silica particles. In addition, high resolution SEM image of the epoxy-ATP-Ti^{4+} silica particles (**Figure 1D**) with corresponding EDS elemental mapping images (**Figure 1E-H**) clearly showed that the layer contained evenly distributed Si, O, P and Ti elements, which implied successful fabrication of the proposed epoxy-ATP-Ti^{4+} IMAC material.

Investigating the Phosphopeptide Enrichment Performance with Standard Proteins. After synthesis, the epoxy-ATP-Ti^{4+} IMAC material was first tested to enrich phosphopeptides from standard protein digests using conventional phosphopeptide enrichment protocol.[37] β-casein and α-casein were chosen as the model sample due to their simple structure and multiple phosphorylation sites. As shown in **Figure 2A** and **Figure 2C**, the signals of phosphopeptides were largely suppressed by highly abundant non-phosphorylated peptides before enrichment. After enriching and eluting with ammonia solution, 3 and 14 phosphopeptides were detected from the β-casein digests (**Figure 2B**) and α-casein digests (**Figure 2D**) with high intensities, respectively, including 2 and 9 multiply phosphorylated peptides. Detailed information of the identified phosphopeptides is provided in **Table S1**. The results well-demonstrated the ability of the epoxy-ATP-Ti^{4+} material to effectively enrich phosphopeptides as an IMAC material.

To further evaluate its performance in phosphopeptide enrichment, a mixture of BSA and β-casein tryptic digests were used for testing the selectivity. As a standard protein with no

phosphorylation modification, BSA was chosen as the background interference sample when enriching phosphopeptides from β-casein tryptic digests. When BSA and β-casein tryptic digests were mixed at 100:1 molar ratio, the MALDI-MS spectrum was dominated by the signals from non-modified peptides from BSA and no signal from phosphopeptides was observed (**Figure 3A**). The signal from phosphopeptides improved dramatically after enrichment with practically no interference from other peptides (**Figure 3B**). Higher molar ratios (500:1 and 1000:1) were also tested, and the material showed excellent selectivity (**Figure 3C, 3D**). To investigate the sensitivity of this method, epoxy-ATP-Ti^{4+} IMAC material was applied to enrich phosphopeptides from 10 fmol/μL of diluted β-casein tryptic digests. The MALDI-MS spectra showed that signal of the phosphopeptides could still be well-detected with S/N over 3 when concentration was as low as 1 fmol/μL (**Figure S2**). As expected, the multiple phosphate groups on the ATP molecule provided strong chelation of Ti^{4+} ions, which in turn bound phosphopeptides strongly, thus enhancing the sensitivity of phosphopeptide enrichment.[33] The outstanding selectivity and sensitivity demonstrated the potential of this material for enriching phosphopeptides from complex biological samples.

Assessing the Separation Efficiency of Epoxy-ATP-Ti^{4+} IMAC Material for Glycopeptide and Phosphopeptide Separation. Typically, IMAC or HILIC methods alone cannot enrich both glycopeptides and phosphopeptides at the same time, and low-abundance modified peptides can experience signal suppression when co-eluted in one fraction.[18] However, we found that certain IMAC materials, possessing hydrophilic functional groups, can act as a HILIC stationary phase to retain glycopeptides in high organic phase while also retaining phosphopeptides through ionic

interaction.[17] To evaluate the ability of epoxy-ATP-Ti^{4+} material to conduct glycopeptide and phosphopeptide enrichment in a HILIC mode, we tested it on a mixture of glycoprotein and phosphoprotein tryptic digests (RNase B and β-casein). The sample loading buffer was switched from 40% ACN/3% TFA to 80% ACN/3% TFA since the conventional buffer for phosphopeptide enrichment did not contain enough organic content to form a water-rich layer with the stationary phase that enabled the retention of hydrophilic analytes.[44] 3% TFA in 80% ACN could effectively protonate the acidic amino acid residues and thereby prevent non-specific binding of acidic peptides when capturing phosphopeptides.[37] As 0.1% TFA aqueous solution could interrupt hydrophilic interaction while keeping the electrostatic interaction between Ti^{4+} ions and phosphopeptides, it was used as elution buffer to collect a fraction mainly consisted of glycopeptides. The remaining phosphopeptides were then eluted in 10% NH_4OH solution where the high concentration of hydroxyl groups would compete with the phosphate groups for forming ionic bonds with metal ions. As shown in **Figure 4A**, no signal of phosphopeptide or glycopeptide was observed when directly analyzing the mixture of RNase B and β-casein digests. However, after enriching the sample in HILIC mode and collecting two fractions under different pH conditions, strong signals of glycopeptides and phosphopeptides were detected in two fractions respectively (**Figure 4B** and **4C**). In total, 18 glycopeptides from RNase B were identified in the first fraction and all three phosphopeptides from β-casein were detected in the second fraction. Detailed information about the identified peptides is provided in **Table S2.** It is worth noting that there was practically no overlap between two fractions, in contrast to our previous work on the CAE-Ti-IMAC material where a relatively strong phosphopeptide signal (*m/z* 2061.8284) was also

found in the first fraction together with glycopeptides.[18] This suggested that epoxy-ATP-Ti^{4+} IMAC material had greater hydrophilicity for retaining glycopeptides and more affinity for phosphopeptides, permitting these two kinds of peptides to be were well-separated in two distinct fractions, demonstrating enhanced selectivity.

Evaluating the Dual-mode Enrichment Performance of Epoxy-ATP-Ti^{4+} IMAC material with HeLa Cell Lysate. After demonstrating the potential of epoxy-ATP-Ti^{4+} IMAC material for enrichment and separation of glycopeptides and phosphopeptides from protein standards, the new material was adapted to enrich more diverse glycopeptides and phosphopeptides from a HeLa cell sample using dual-mode enrichment. The core principle of a dual-mode enrichment is enriching one class of peptides through single interaction of enrichment material while enriching the other class of peptides by other synergistic interactions. In this way, dual-mode elution can be applied: one disrupts the single interaction, while the other can disrupt the synergistic interactions, which allows the two classes of peptides to be eluted separately. Here, 100 µg of HeLa cell tryptic digests were loaded onto the epoxy-ATP-Ti^{4+} IMAC material in HILIC mode. At the same time, a parallel experiment was performed with CAE-Ti-IMAC material. Four fractions were collected for a larger scale analysis. The first two fractions, eluted in weakly acidic conditions with decreasing ACN concentration, were expected to release neutral glycopeptides retained by hydrophilic interaction. The third fraction, eluted with different concentrations of ACN under strongly acidic conditions, mimicked the conventional loading and washing buffer of the IMAC method for phosphopeptide enrichment. The sialylglycopeptides captured through synergistic hydrophilic and electrostatic interactions were expected to be thoroughly released in this step since the strong acid in the elution

buffer could protonate sialic acid residues and disrupt the electrostatic interaction. The phosphopeptides and M6P glycopeptides would still be retained on the IMAC material, because the phosphate group has a lower pKa value (0.7–1.0) than sialic acid (~2.6), which cannot be protonated by current strong acid condition.[18] The fourth elution with ammonia solution was intended to wash off all remaining phosphopeptides and M6P glycopeptides. **Figure 5A** and **Figure 5B** show the identification numbers of the different types of modified peptides for the two IMAC materials. The results suggest that the epoxy-ATP-Ti^{4+} IMAC material outperformed the CAE-Ti-IMAC material in all aspects of a dual-mode enrichment. First, the epoxy-ATP-Ti^{4+} IMAC material better separated sialylglycopeptides from neutral glycopeptides in the first three fractions. The ratio of sialylglycopeptides in four fractions increased gradually, indicating the successful separation enabled by the dual-mode elution (**Figure S3**). This separation is necessary to enhance the detection sensitivity for sialylated glycopeptides, as ionization of sialylglycopeptides tends to be suppressed by co-eluting neutral glycopeptides due to their negatively charged ions.[45] With lower ACN concentration and pH, more multi-sialylglycopeptides were eluted in later fractions, which are usually underexplored in conventional HILIC enrichment. The epoxy-ATP-Ti^{4+} IMAC material exhibited better retention of these peptides as more of them were distributed in the last two fractions (**Figure 5C** and **5d**). Second, the epoxy-ATP-Ti^{4+} IMAC material had better affinity for phosphopeptides. In the last fraction, 3425 phosphopeptides were identified with enrichment specificity as high as 81.6% (**Figure 5F**), which was more than two times the phosphopeptides identified with CAE-Ti-IMAC material. Additionally, very few phosphopeptides were eluted in the first three fractions with epoxy-ATP-Ti^{4+} IMAC material while

171 and 866 phosphopeptides were found in the second and the third fractions in CAE-Ti-IMAC enrichment, leading to lower enrichment specificity of phosphopeptides in each fraction (**Figure 5E**). Overall, the total identified modified peptides with epoxy-ATP-Ti^{4+} IMAC material exceeded those with CAE-Ti-IMAC in all categories (**Figure S4**). These findings demonstrate the advantageous properties of the epoxy-ATP-Ti^{4+} IMAC material for dual-mode PTM enrichment, benefitting from the low interference between fractions and extremely high enrichment specificity.

Applying the Epoxy-ATP-Ti^{4+} IMAC Material to Mouse Lung Sample Analysis. The material and dual-mode elution strategy were further applied to enrich PTM peptides from mouse lung tissue, a more complex biological sample. Starting with a larger sample amount, the workflow was expanded to collect six fractions, including four fractions under acidic conditions and two in basic conditions. **Figure 6A** shows that a good separation of glycopeptides and phosphopeptides was achieved. Specifically, 1640, 1383, 1450 and 980 glycopeptides were found in the acidic fractions and 2746 and 1138 phosphopeptides were found in the basic fractions, respectively. Notably, almost no phosphopeptides were eluted in the first four fractions and only a small number of glycopeptides were eluted in the last two fractions, which mainly contained negatively charged sialylated or phosphorylated peptides. Because multi-sialylglycopeptides had stronger affinity for Ti^{4+} ions, these peptides were not eluted until a high-pH solvent was added (**Figure S5**), which suggested that conventional HILIC enrichment and single-mode elution were likely to miss these peptides. In accordance with the results with HeLa cell samples, the ratio of enriched sialylglycopeptides went up gradually when lowering ACN concentration and pH (**Figure 6B**). The decreasing ACN concentration at high pH enabled separation of phosphopeptides and M6P

glycopeptides. Given the low abundance of M6P glycopeptides, they often suffer from signal suppression when analyzed together with phosphopeptides. Nevertheless, these two peptides could be separated in the HILIC mode at high pH with IMAC material based on their different hydrophilicity.[17] As reflected in **Figure S6**, the percentage of phosphopeptides and multi-phosphorylated peptides eluted increased with decreasing ACN concentration. In addition, 58.3% of M6P glycopeptides were found in the last fraction. These might have remained undiscovered without the dual-mode elution strategy at different pH conditions. In total, this method enabled identification of 2928 unique N-glycopeptides and 3051 unique phosphopeptides, of which 896 were sialylglycopeptides, 24 were M6P glycopeptides and 493 were multi-phosphorylated peptides (**Figure 6C**). This large-scale simultaneous enrichment and separation of diverse glycopeptides and phosphopeptides shed light on the PTM crosstalk analysis. These PTM peptides were mapped to 491 glycoproteins and 1471 phosphoproteins, with 99 proteins found to be both glycosylated and phosphorylated, and 4 M6P glycoproteins also found to be phosphorylated (**Figure 6D**). **Figure 6E** visualized the site numbers of two PTMs on the same protein and classified the glycoproteins based on their glycan type. N-glycoproteins were exclusively classified into four glycosylation type categories based on the glycan composition identified: (1) M6P (containing M6P glycan); (2) sialylated (containing sialic acid); (3) fucosylated (containing fucose); (4) other type (containing only complex, hybrid, high-mannose or paucimannose glycans). The scatter plot revealed several hypermodified proteins that contained both PTMs and the microheterogeneity of each glycoprotein as well. Sialylation was found to be the predominant

glycosylation type, with 53.5% of these proteins being sialylated, suggesting that sialylation may play a crucial role in PTM crosstalk with phosphorylation.

Conclusions

In this work, we have introduced a novel epoxy-ATP-Ti^{4+} IMAC material that is easy and facile to prepare in any MS lab. With ATP as the functional ligand, the material possessed enhanced hydrophilicity and superior metal ion binding affinity, which made it an effective stationary phase for both N-glycopeptide and phosphopeptide enrichment. In conventional IMAC mode, the material showed excellent selectivity and sensitivity in phosphopeptide enrichment. When using loading buffer containing high organic content and TFA concentration, the material enabled simultaneous enrichment and separation of N-glycopeptides and phosphopeptides, with practically no overlap between the two sets. The material outperformed the well-developed CAE-Ti-IMAC material in parallel experiments, highlighting the importance of enhanced hydrophilicity in dual-mode enrichment. The novel material was successfully applied to different PTM enrichment from complex biological samples, including HeLa cells and mouse lung tissues. The results showed that the material was not only able to separate N-glycopeptides and phosphopeptides, but also able to separate sialylglycopeptides and M6P glycopeptides with more fractions due to their synergistic electrostatic and hydrophilic interaction with the stationary phase. Large-scale analysis of different PTMs can be achieved at the same time with less sample loss and the need for more than one enrichment material. We anticipate that this newly developed material and the associated fractionation strategy will serve as a useful tool for studying the N-glycoproteome and phosphoproteome, as well as their potential crosstalk, in complex biological samples.

Acknowledgements

The authors wish to thank Dr. Andrew Alpert from PolyLC Inc. for providing the epoxy-functionalized silica material and his critical reading and revision of this manuscript. The authors thanks Gary Girdaukas from Analytical Instrumentation Center (AIC) at School of Pharmacy, University of Wisconsin-Madison for providing technical assistance with FT-IR analysis. This work was supported, in part, by the National Science Foundation (CHE-2108223) and National Institutes of Health Grants RF1AG052324, R21AG065728, R01AG078794, and R01 DK071801 (to L.L.). Some of the mass spectrometers were acquired using NIH shared instrument grants S10 OD028473, S10 RR029531, and S10 OD025084. L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Reference

(1) Zolnierowicz, S.; Bollen, M. Protein Phosphorylation and Protein Phosphatases. *EMBO J.* **2000**, *19* (4), 483–488.

(2) Yang, C.; Zhong, X.; Li, L. Recent Advances in Enrichment and Separation Strategies for Mass Spectrometry-based Phosphoproteomics. *Electrophoresis* **2014**, *35* (24), 3418–3429.

(3) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. L. The Crucial Role of Protein Phosphorylation in Cell Signaling and Its Use as Targeted Therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271–280.

(4) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.

(5) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1* (1), 90.

(6) Doll, S.; Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **2015**, *10* (1), 63–71.

(7) Huang, J.; Wang, F.; Ye, M.; Zou, H. Enrichment and Separation Techniques for Large-Scale Proteomics Analysis of the Protein Post-Translational Modifications. *J. Chromatogr. A* **2014**, *1372*, 1–17.

(8) Qiu, W.; Evans, C. A.; Landels, A.; Pham, T. K.; Wright, P. C. Phosphopeptide Enrichment for Phosphoproteomic Analysis - A Tutorial and Review of Novel Materials. *Anal. Chim. Acta* **2020**, *1129*, 158–180.

(9) Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J. A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-Based Glycoproteomics. *Mol. Cell Proteomics* **2021**, *20*, 100029.

(10) Chen, Z.; Huang, J.; Li, L. Recent Advances in Mass Spectrometry (MS)-Based Glycoproteomics in Complex Biological Samples. *Trends Analyt Chem.* **2018**, *118* (Cell 126 2006), 880–892.

(11) Leutert, M.; Entwisle, S. W.; Villén, J. Decoding Post Translational Modification Crosstalk with Proteomics. *Mol. Cell Proteomics* **2021**, *20*, 100129.

(12) Cui, Y.; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L. Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Anal. Chem.* **2021**, *93* (22), 7908–7916.

(13) Aguilar, H. A.; Iliuk, A. B.; Chen, I.-H.; Tao, W. A. Sequential Phosphoproteomics and N-Glycoproteomics of Plasma-Derived Extracellular Vesicles. *Nat. Protoc.* **2020**, *15* (1), 161–180.

(14) Silbern, I.; Fang, P.; Ji, Y.; Christof, L.; Urlaub, H.; Pan, K.-T. Quantitative Methods in Proteomics. *Methods Mol. Biology* **2021**, *2228*, 185–203.

(15) Zhou, Y.; Lih, T.-S. M.; Yang, G.; Chen, S.-Y.; Chen, L.; Chan, D. W.; Zhang, H.; Li, Q. K. An Integrated Workflow for Global, Glyco-, and Phospho-Proteomic Analysis of Tumor Tissues. *Anal. Chem.* **2019**, *92* (2), 1842–1849.

(16) Cho, K.-C.; Chen, L.; Hu, Y.; Schnaubelt, M.; Zhang, H. Developing Workflow for Simultaneous Analyses of Phosphopeptides and Glycopeptides. *ACS Chem. Biol.* **2018**, *14* (1), 1–9.

(17) Huang, J.; Dong, J.; Shi, X.; Chen, Z.; Cui, Y.; Liu, X.; Ye, M.; Li, L. Dual-Functional Titanium(IV) Immobilized Metal Affinity Chromatography Approach for Enabling Large-Scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. *Anal. Chem.* **2019**, *91* (18), 11589–11597.

(18) Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

(19) Lu, Q.; Chen, C.; Xiong, Y.; Li, G.; Zhang, X.; Zhang, Y.; Wang, D.; Zhu, Z.; Li, X.; Qing, G.; Sun, T.; Liang, X. High-Efficiency Phosphopeptide and Glycopeptide Simultaneous Enrichment by Hydrogen Bond–Based Bifunctional Smart Polymer. *Anal Chem* **2020**, *92* (9), 6269–6277.

(20) Xiong, F.; Jia, J.; Ma, J.; Jia, Q. Glutathione-Functionalized Magnetic Thioether-COFs for the Simultaneous Capture of Urinary Exosomes and Enrichment of Exosomal Glycosylated and Phosphorylated Peptides. *Nanoscale* **2021**, *14* (3), 853–864.

(21) Luo, B.; Yan, S.; Zhang, Y.; Zhou, J.; Lan, F.; Wu, Y. Bifunctional Magnetic Covalent Organic Framework for Simultaneous Enrichment of Phosphopeptides and Glycopeptides. *Anal. Chim. Acta* **2021**, *1177*, 338761.

(22) Shang, D.; Chen, C.; Dong, X.; Cui, Y.; Qiao, Z.; Li, X.; Liang, X. Simultaneous Enrichment and Sequential Separation of Glycopeptides and Phosphopeptides with Poly-Histidine Functionalized Microspheres. *Front. Bioeng. Biotechnol.* **2022**, *10*, 1011851.

(23) Zou, X.; Jie, J.; Yang, B. Single-Step Enrichment of N-Glycopeptides and Phosphopeptides with Novel Multifunctional Ti 4+-Immobilized Dendritic Polyglycerol Coated Chitosan Nanomaterials. *Anal. Chem.* **2017**, *89* (14), 7520–7526.

(24) Hong, Y.; Zhao, H.; Pu, C.; Zhan, Q.; Sheng, Q.; Lan, M. Hydrophilic Phytic Acid-Coated Magnetic Graphene for Titanium(IV) Immobilization as a Novel Hydrophilic Interaction Liquid Chromatography–Immobilized Metal Affinity Chromatography Platform for Glyco- and Phosphopeptide Enrichment with Controllable Selectivity. *Anal. Chem.* **2018**, *90* (18), 11008–11015.

(25) Hong, Y.; Zhan, Q.; Zheng, Y.; Pu, C.; Zhao, H.; Lan, M. Hydrophilic Phytic Acid-Functionalized Magnetic Dendritic Mesoporous Silica Nanospheres with Immobilized Ti4+: A Dual-Purpose Affinity Material for Highly Efficient Enrichment of Glycopeptides/Phosphopeptides. *Talanta* **2019**, *197*, 77–85.

(26) Wang, Z.; Wang, J.; Sun, N.; Deng, C. A Promising Nanoprobe Based on Hydrophilic Interaction Liquid Chromatography and Immobilized Metal Affinity Chromatography for Capture of Glycopeptides and Phosphopeptides. *Anal. Chim. Acta* **2019**, *1067*, 1–10.

(27) Zhang, Y.; Li, J.; Yu, Y.; Xie, R.; Liao, H.; Zhang, B.; Chen, J. Coupling Hydrophilic Interaction Chromatography Materials with Immobilized Fe 3+ for Phosphopeptide and Glycopeptide Enrichment and Separation. *RSC Adv.* **2020**, *10* (37), 22176–22182.

(28) Zheng, H.; Jia, J.; Li, Z.; Jia, Q. Bifunctional Magnetic Supramolecular-Organic Framework: A Nanoprobe for Simultaneous Enrichment of Glycosylated and Phosphorylated Peptides. *Anal Chem* **2020**, *92* (3), 2680–2689.

(29) Tang, R.; Yu, Y.; Dong, J.; Yao, Y.; Ma, S.; Ou, J.; Ye, M. Facile Preparation of Bifunctional Adsorbents for Efficiently Enriching N-Glycopeptides and Phosphopeptides. *Anal. Chim. Acta* **2021**, *1144*, 111–120.

(30) Sun, N.; Wu, H.; Shen, X. Magnetic Titanium Dioxide Nanomaterial Modified with Hydrophilic Dicarboxylic Ligand for Effective Enrichment and Separation of Phosphopeptides and Glycopeptides. *Microchim. Acta* **2020**, *187* (3), 195.

(31) Sheng, Q.; Xue, C.; Zhou, Y.; Li, J.; Yuan, H.; Ke, Y.; Lan, M. Synthesis of Al3+-Doping-TiO2 Monodisperse Microspheres and Their Application for Phosphopeptides and Glycopeptides Enrichment. *Talanta* **2021**, *223* (Pt 2), 121715.

(32) Pan, Y.; Zhang, C.; Xiao, R.; Zhang, L.; Zhang, W. Dual-Functionalized Magnetic Bimetallic Metal-Organic Framework Composite for Highly Specific Enrichments of Phosphopeptides and Glycopeptides. *Anal. Chim. Acta* **2021**, *1158*, 338412.

(33) Zhang, L.; Zhao, Q.; Liang, Z.; Yang, K.; Sun, L.; Zhang, L.; Zhang, Y. Synthesis of Adenosine Functionalized Metal Immobilized Magnetic Nanoparticles for Highly Selective and Sensitive Enrichment of Phosphopeptides. *Chem. Commun.* **2012**, *48* (50), 6274–6276.

(34) Zhang, L.; Liang, Z.; Zhang, L.; Zhang, Y.; Shao, S. Facile Synthesis of Gallium Ions Immobilized and Adenosine Functionalized Magnetic Nanoparticles with High Selectivity for Multi-Phosphopeptides. *Anal. Chim. Acta* **2015**, *900* (C), 46–55.

(35) Su, J.; He, X.; Chen, L.; Zhang, Y. Adenosine Phosphate Functionalized Magnetic Mesoporous Graphene Oxide Nanocomposite for Highly Selective Enrichment of Phosphopeptides. *ACS Sustain. Chem. Eng.* **2018**, *6* (2), 2188–2196.

(36) Yan, Y.; Lu, Y.; Chen, M.; Liang, H. A Novel IMAC Platform – Adenosine Coupled Functional Magnetic Microspheres for Phosphoproteome Research. *Anal. Methods* **2018**, *10* (10), 1190–1195.

(37) Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J. R.; Zou, H.; Mohammed, S. Robust Phosphoproteome Enrichment Using Monodisperse Microsphere–Based Immobilized Titanium (IV) Ion Affinity Chromatography. *Nat. Protoc.* **2013**, *8* (3), 461–480.

(38) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301–2319.

(39) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (Database issue), D442–D450.

(40) Zhang, W.; Lai, C.-K.; Huang, W.; Li, W.; Wu, S.; Kong, Q.; Hopkinson, A. C.; Fernie, A. R.; Siu, K. W. M.; Yan, S. An Eco-Friendly, Low-Cost, and Automated Strategy for Phosphoproteome Profiling. *Green Chem.* **2022**, *24* (24), 9697–9708.

(41) Mello, M. L. S.; Vidal, B. C. Changes in the Infrared Microspectroscopic Characteristics of DNA Caused by Cationic Elements, Different Base Richness and Single-Stranded Form. *PLoS One* **2012**, *7* (8), e43169.

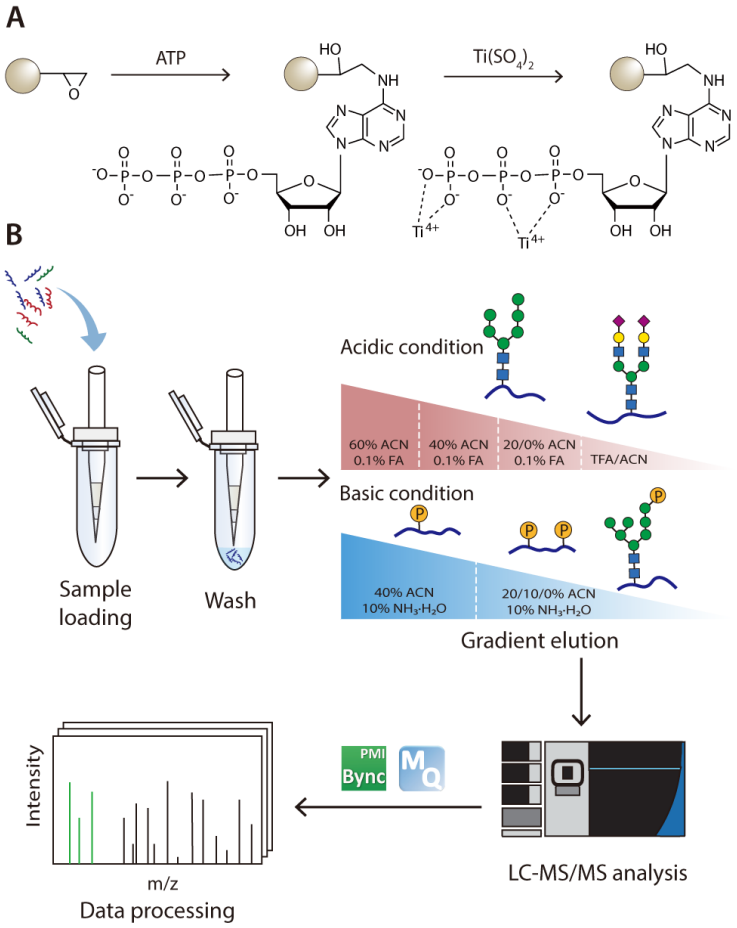
(42) Zhang, K.; Hu, D.; Deng, S.; Han, M.; Wang, X.; Liu, H.; Liu, Y.; Xie, M. Phytic Acid Functionalized Fe3O4 Nanoparticles Loaded with Ti(IV) Ions for Phosphopeptide Enrichment in Mass Spectrometric Analysis. *Microchim. Acta* **2019**, *186* (2), 68.

(43) Nowak, M. J.; Lapinski, L.; Kwiatkowski, J. S.; Leszczyński, J. Molecular Structure and Infrared Spectra of Adenine. Experimental Matrix Isolation and Density Functional Theory Study of Adenine 15 N Isotopomers. *J Phys Chem* **1996**, *100* (9), 3527–3534.

(44) Buszewski, B.; Noga, S. Hydrophilic Interaction Liquid Chromatography (HILIC)—a Powerful Separation Technique. *Anal. Bioanal. Chem.* **2012**, *402* (1), 231–247.

(45) Huang, J.; Wang, D.; Shipman, R. D.; Zhu, Z.; Liu, Y.; Li, L. Simultaneous Enrichment and Separation of Neutral and Sialyl Glycopeptides of SARS-CoV-2 Spike Protein Enabled by Dual-Functionalized Ti-IMAC Material. *Anal. Bioanal. Chem.* **2021**, 1–9.

Figures



Scheme 1. (A) Synthesis of epoxy-ATP-Ti^{4+} IMAC material; (B) Workflow of simultaneous enrichment of N-glycopeptides and phosphopeptides using epoxy-ATP-Ti^{4+} IMAC material.

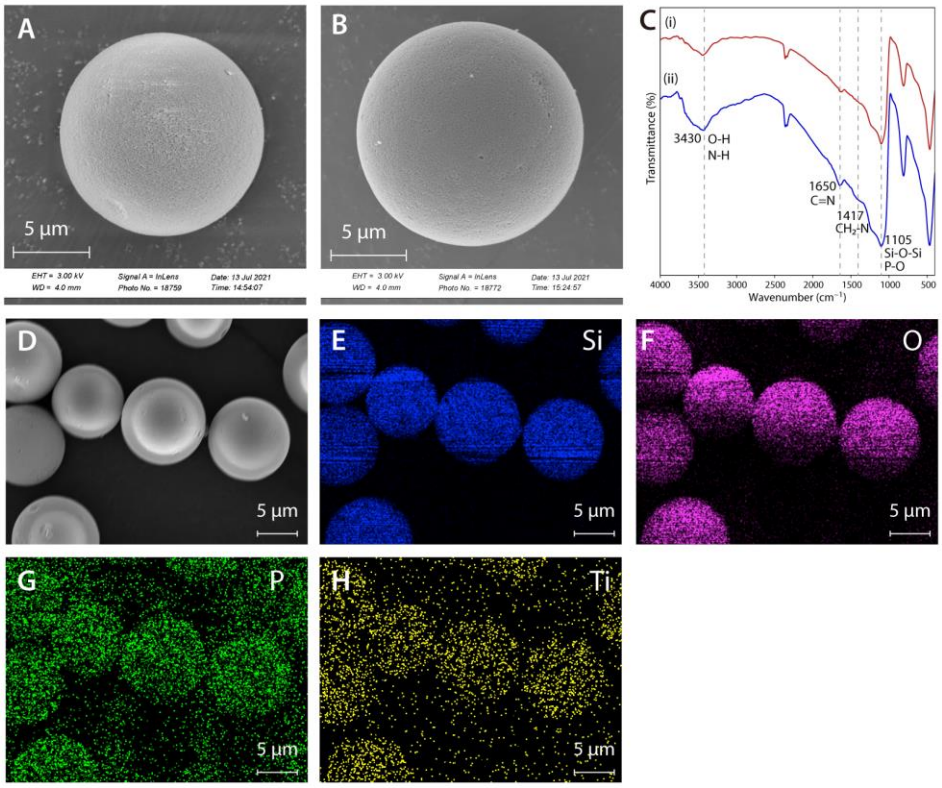


Figure 1. SEM images of (A) epoxy-ATP silica particles and (B) epoxy-ATP-Ti^{4+} silica particles; (C) FTIR spectra of epoxy-functionalized silica particles (i) before and (ii) after ATP coupling reaction; (D) SEM image of epoxy-ATP-Ti^{4+} silica particles with EDS elemental mapping images of (E) Si, (F) O, (G) P, and (H) Ti.

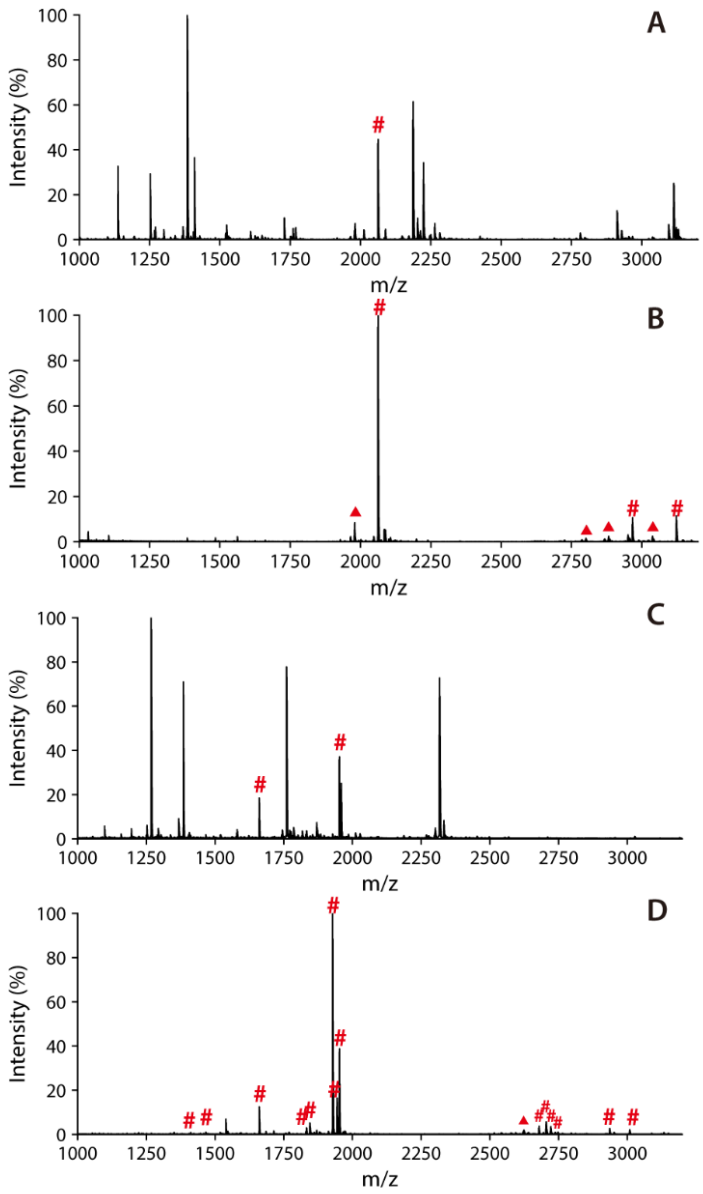


Figure 2. MALDI-TOF mass spectra of tryptic digests from standard protein (10 μg) before and after phosphopeptide enrichment by epoxy- ATP-Ti^{4+} IMAC material: (A)(B) β-casein; (C)(D) α-casein. Phosphopeptides are labeled with red "#", neutral loss is labeled with "▲".

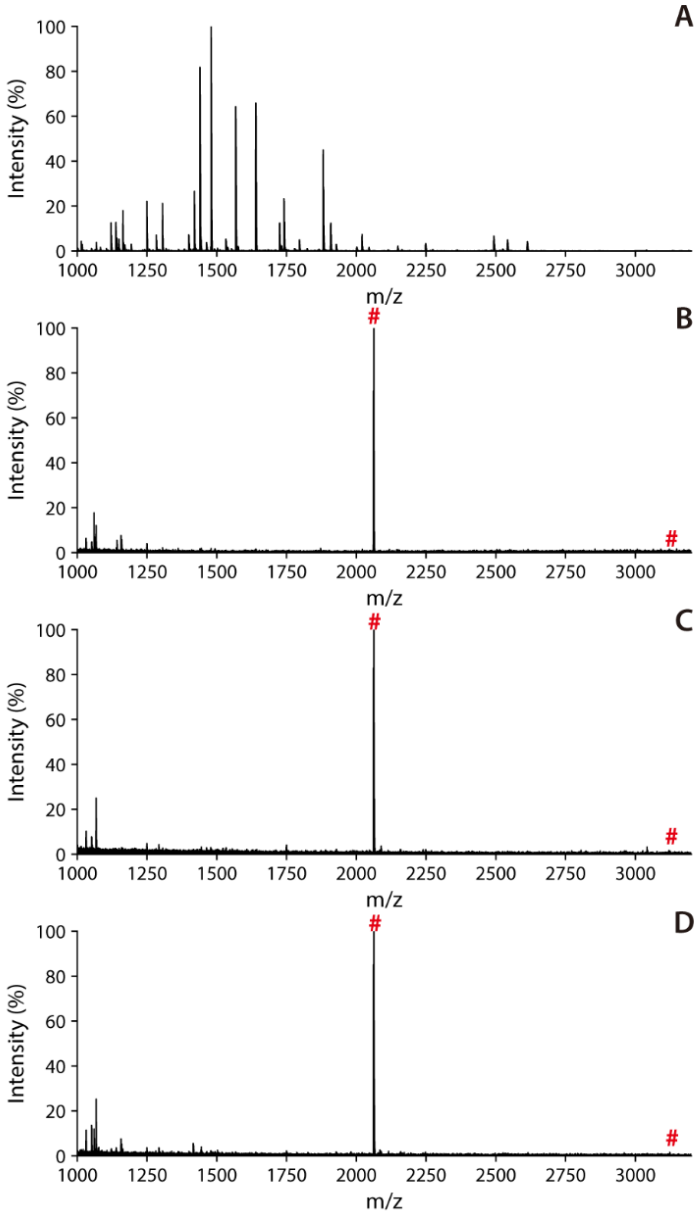


Figure 3. MALDI-TOF mass spectra of phosphopeptides enriched from mixture of BSA and β-casein digests at different molar ratios. (A) Direct analysis at 100:1 ratio; After enrichment at (B) 100:1, (C) 500:1 and (D) 1000:1 ratio. Phosphopeptides are labeled with red "#", neutral loss is labeled with red "▲".

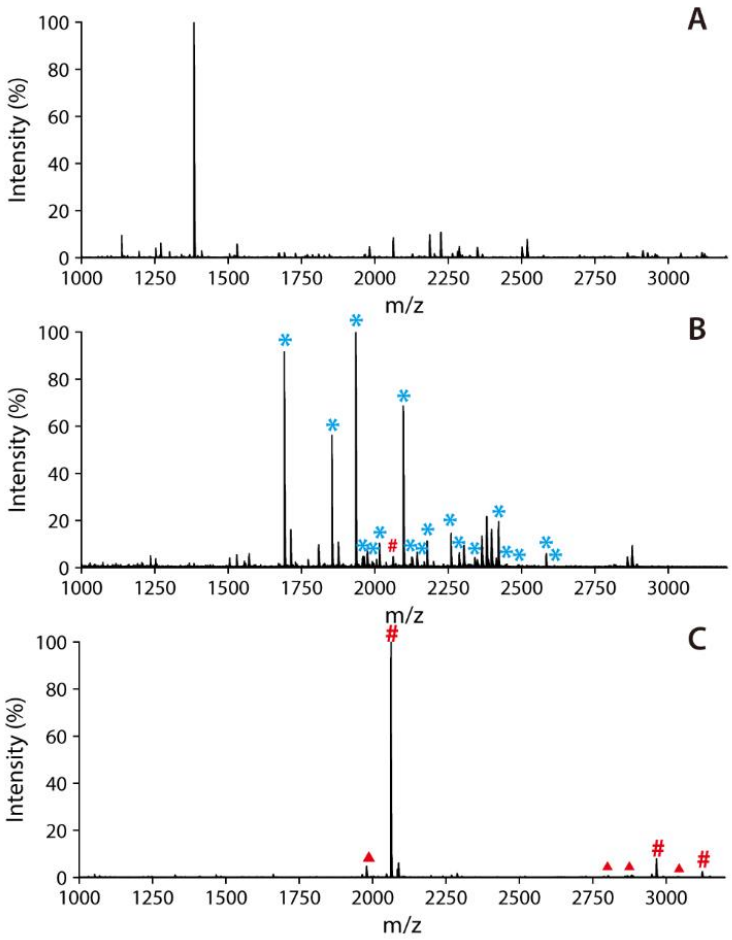


Figure 4. MALDI-TOF mass spectra of tryptic digests from β-casein (10 μg) and RNase B (10 μg) (A) before enrichment by epoxy-ATP-Ti^{4+} IMAC material; (B) eluted with 0.1% TFA; (C) eluted with 10% ammonia. Phosphopeptides are labeled with red "#" and neutral loss is labeled with red "▲"; glycopeptides are labeled with blue "*".

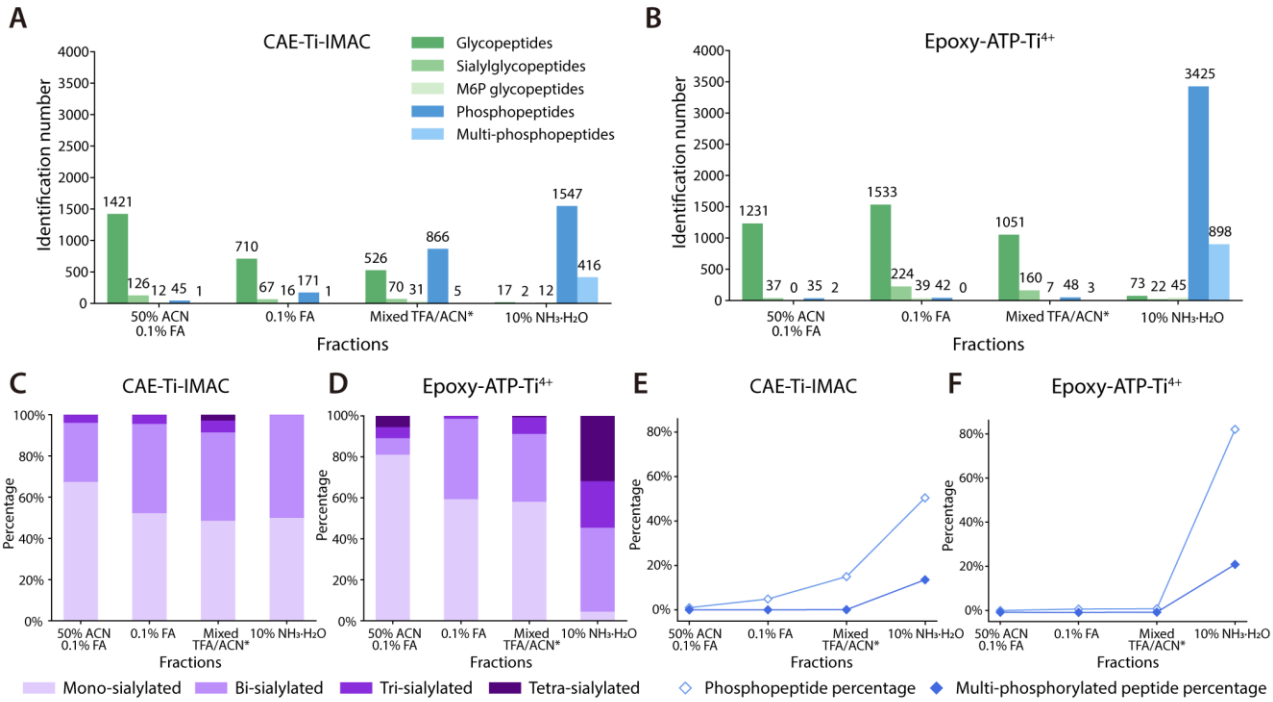


Figure 5. Simultaneous enrichment of glycopeptides and phosphopeptides from HeLa cells with two IMAC materials. Identification numbers refer to peptides in each fraction enriched by (A) CAE-Ti-IMAC and (B) epoxy-ATP-Ti^{4+} IMAC material; Distribution of glycopeptides with different number of sialic acids enriched by (C) CAE-Ti-IMAC and (D) epoxy-ATP-Ti^{4+} IMAC material; phosphopeptide ratio and multi-phosphorylated peptide percentages in fractions enriched by (E) CAE-Ti-IMAC and (F) epoxy-ATP-Ti^{4+} IMAC material. *: The mixed TFA/ACN fraction was pooled from three eluates: 40% ACN/3% TFA, 50% ACN/6% TFA/200 mM NaCl, and 30% ACN/0.1% TFA, and was desalted before LC-MS/MS analysis.

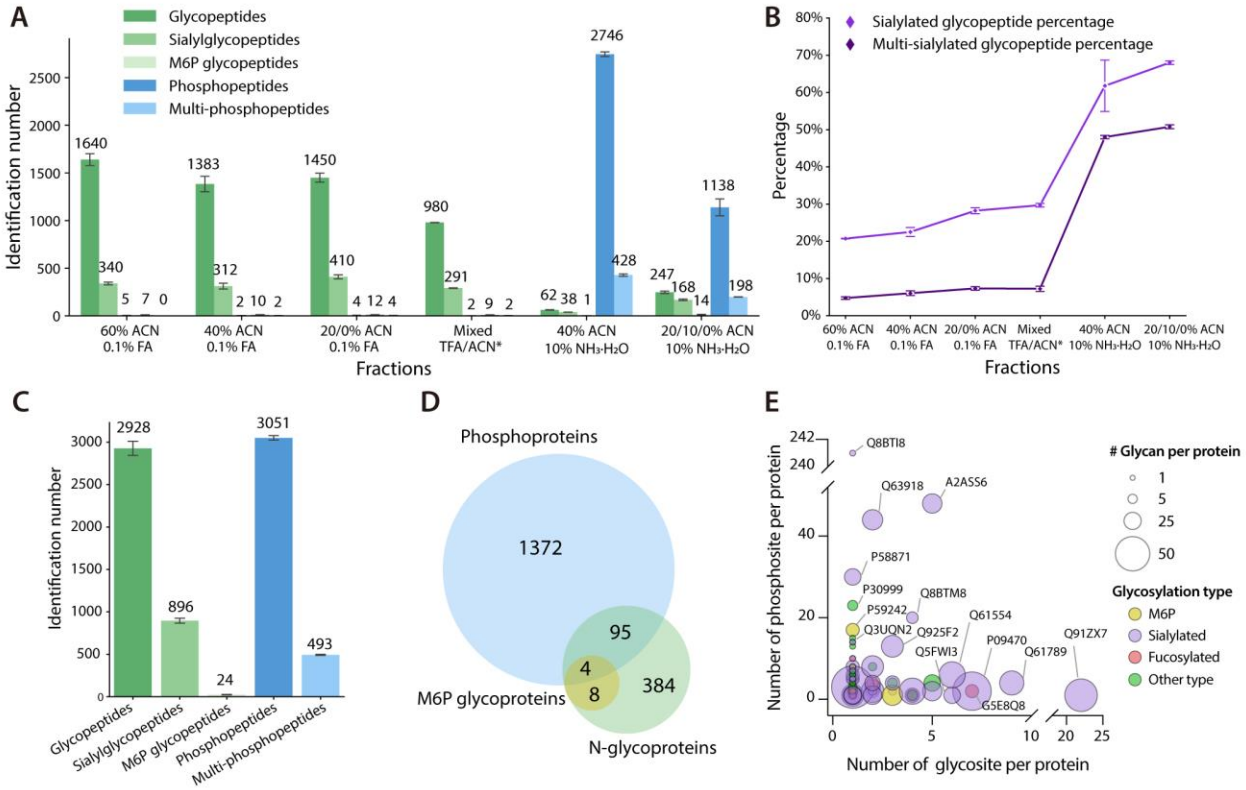


Figure 6. Simultaneous enrichment of glycopeptides and phosphopeptides from mouse lung tissue. (A) Number of peptides identified in each fraction; (B) Sialylated glycopeptide and multi-sialylated glycopeptide percentages in six fractions; (C) Total identification number of modified peptides; (D) Venn diagram showing coexistence of N-glycosylation and phosphorylation at the protein level; (E) Scatter plot of the number of modified sites within proteins that contain both PTMs. Error bar represented sample standard deviation. *: The mixed TFA/ACN fraction was pooled from three eluates: 40% ACN/3% TFA, 50% ACN/6% TFA/200 mM NaCl, and 30% ACN/0.1% TFA, and was desalted before LC-MS/MS analysis.

Supplemental Information

Experimental Section.

Cell Culture and Sample Preparation. For standard proteins, 1 mg protein was dissolved in 100 µL 8M urea/50 mM TEAB buffer. HeLa cells were cultured in DMEM (Hyclone) supplemented with 10% fetal bovine serum (Gibco), 1% penicillin/streptomycin (Sigma Aldrich), and incubated at 37 ºC in a humidified chamber with 5% CO_2. Cells were harvested by washing with PBS buffer (Gibco, pH 7.4) and the pellets were lysed in buffer consisting of 50 mM Tris base (pH 7.4), 4% SDS, 65 mM DTT, 175 mM NaCl. 1 protease inhibitor tablet and 1 phosphatase inhibitor tablet were added per every 10 mL of the lysis buffer. The lysate was sonicated with a probe sonicator in ice water bath at 50% power with pulse 5s on 5s off for 12 cycles. Cell lysates were centrifuged, and supernatant was collected and poured into ice-cold precipitation buffer (acetone: ethanol: acetic acid=50: 50: 0.1) at a ratio of 1:5. The precipitation went overnight in -20ºC. Protein pellets were collected through centrifugation at 18000 xg and washed twice with ice-cold precipitation buffer. The pellets were dried in the fume hood for 10 min and redissolved in 8M urea/50 mM TEAB buffer (pH 8.0). Protein concentrations were measured by a BCA assay kit (Thermo Fisher Scientific, San Jose, CA). For mouse lung tissue, the tissue was collected from male adult C57BL/6 mice. The tissue sample was sliced into 1-2 mm^3 pieces in a clean dish in an ice bath and washed with 150 mM ice-cold PBS buffer for 3 times. The sample was then lysed and sonicated as described above. Proteins were precipitated and the concentration was measured.

Protein Digestion. Proteins were reduced in 20 mM DTT at 37 ºC for 2h, followed by alkylation with 40 mM IAA at room temperature in the dark for 30 min. Alkylation was quenched by adding 20 mM DTT for another 10 min. The urea buffer was diluted to 1.6 M with 50 mM TEAB buffer.

The proteins were first digested with trypsin at a protein-to-enzyme ratio at 100:1 and incubated at 37 ^{o}C. After 12 h, the same amount of trypsin was added again and incubated at 37 ^{o}C for another 4 h to make a final protein-to-enzyme ratio of 50:1. Digestion was quenched by adding TFA to a final concentration of 1%. The samples were stored at -80^{o}C for future use.

Phosphopeptide Enrichment. IMAC material was packed into an empty TopTip with 3 mg cotton at bottom. The TopTip was placed on a 2 mL microcentrifuge tube with the help of an adapter unit. Peptide samples were resuspended in 100 µL loading buffer (40% ACN, 3% TFA) and loaded onto the IMAC-cotton tip. The IMAC-cotton tip was centrifuged at 200 xg for 2 min and flow-through of the sample was re-loaded two more times to ensure complete retention. The IMAC-cotton tip was washed with washing buffer I (50% ACN, 6% TFA and 200 mM NaCl) and washing buffer II (30% ACN, 0.1% TFA) at 200 g for 2 min for three times, respectively. Finally, the phosphopeptides were eluted with 150 µL 10% NH_4OH (v/v). Samples were dried down *in vacuo* before MS analysis.

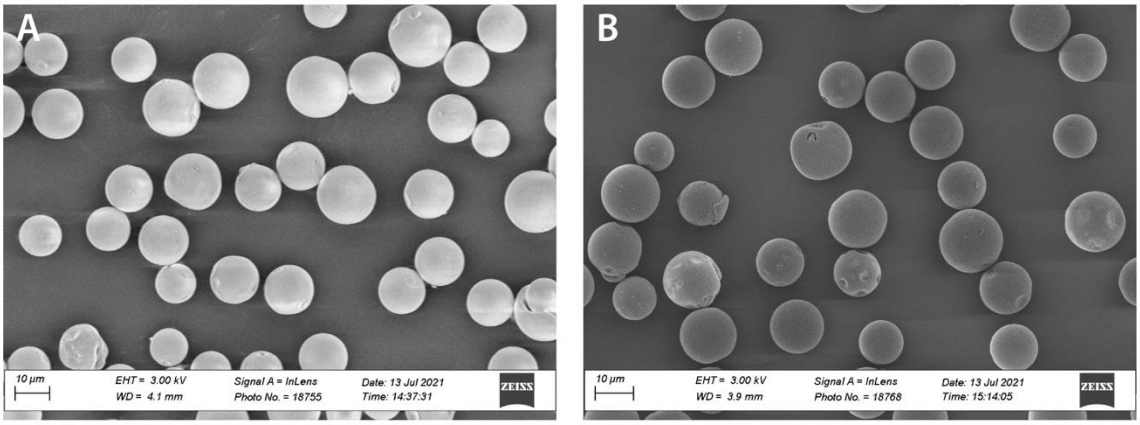


Figure S1. SEM images of (A) epoxy-ATP silica particles and (B) epoxy-ATP-Ti^{4+} silica particles on a smaller scale.

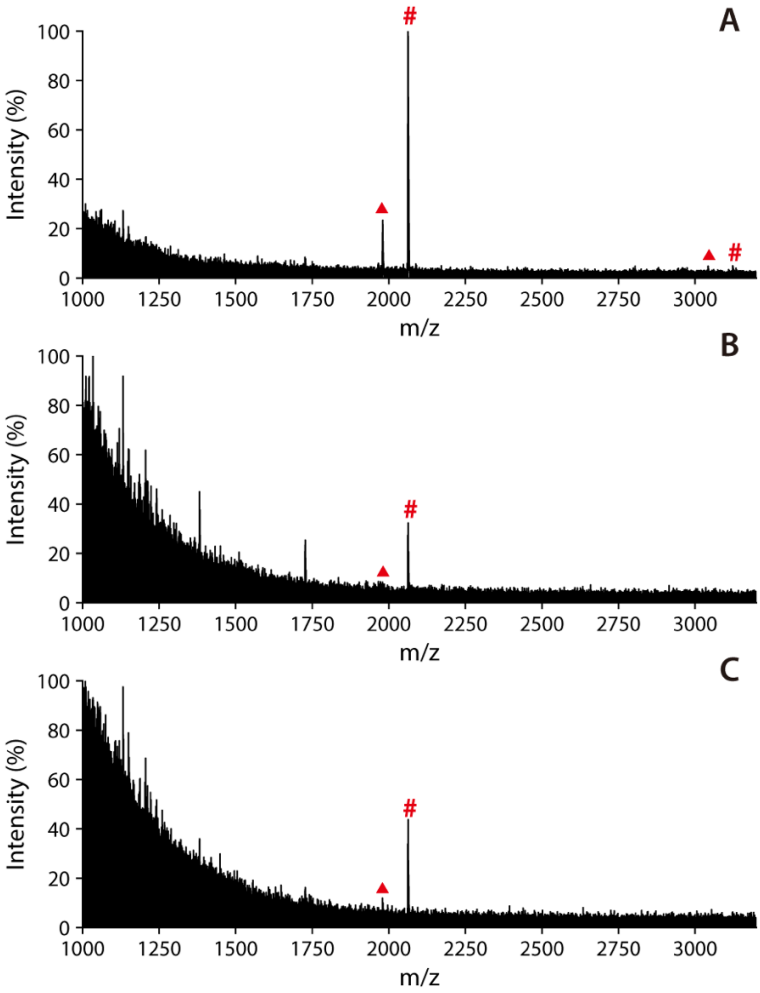


Figure S2. MALDI-TOF mass spectra of phosphopeptides enriched from β-casein digests with epoxy-ATP-Ti^{4+} IMAC material at different concentrations: (A) 10 fmol/μL, (B) 5 fmol/μL, and (C) 1 fmol/μL. Phosphopeptides are labeled with red "#", neutral loss is labeled with "▲".

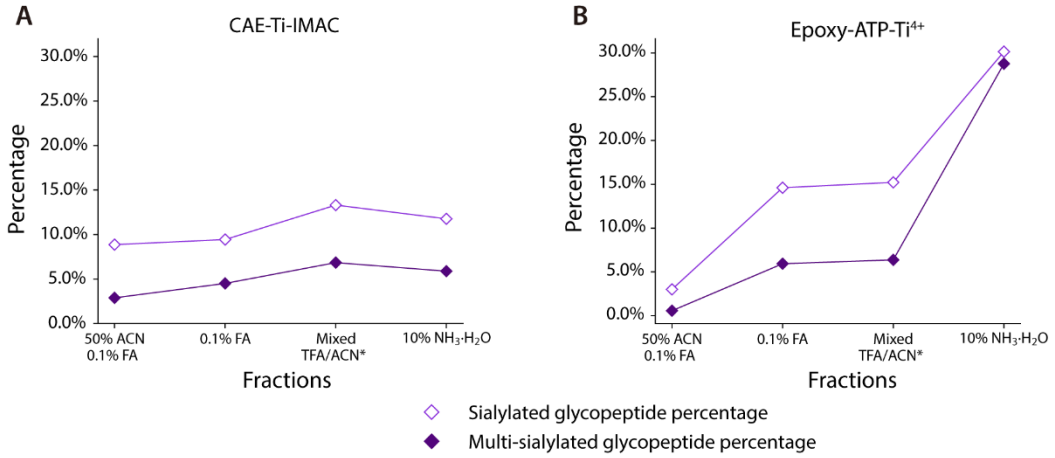


Figure S3. Simultaneous enrichment of glycopeptides and phosphopeptides from HeLa cells with the two IMAC materials. Sialylated glycopeptide percentages with (A) CAE-Ti-IMAC and (B) epoxy-ATP-Ti^{4+} IMAC material.

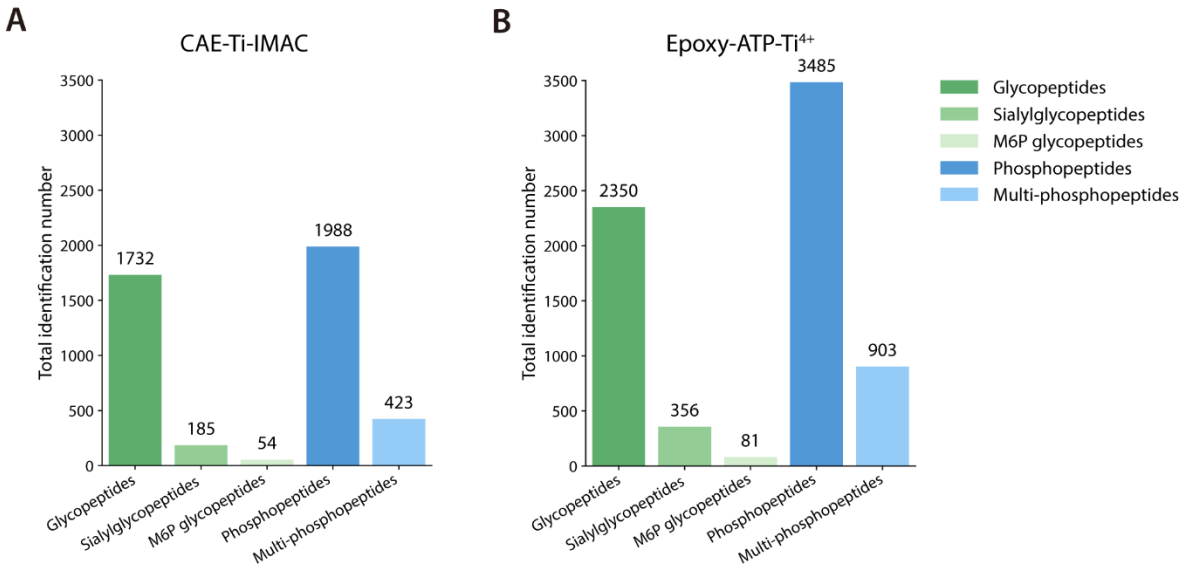


Figure S4. Total number of glycopeptides and phosphopeptides identified from HeLa cells among four fractions in LC-MS/MS analysis with (A) CAE-Ti-IMAC material and (B) epoxy-ATP-Ti^{4+} IMAC material.

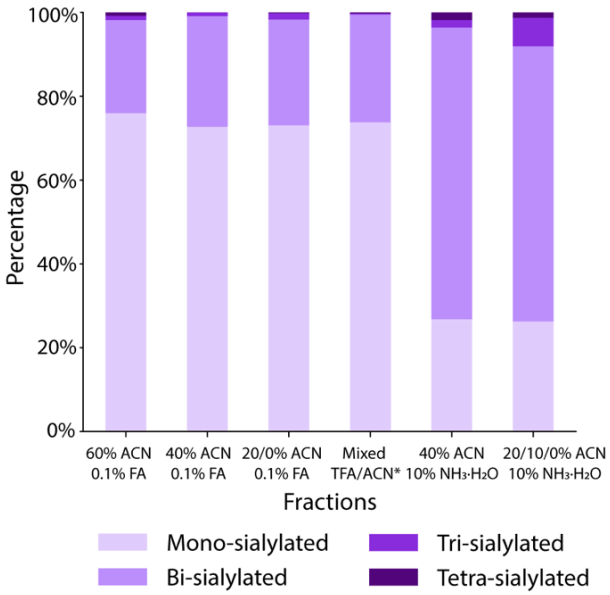


Figure S5. Distribution of glycopeptide with different numbers of sialic acids in six fractions of simultaneous enrichment of glycopeptides and phosphopeptides from mouse lung tissue.

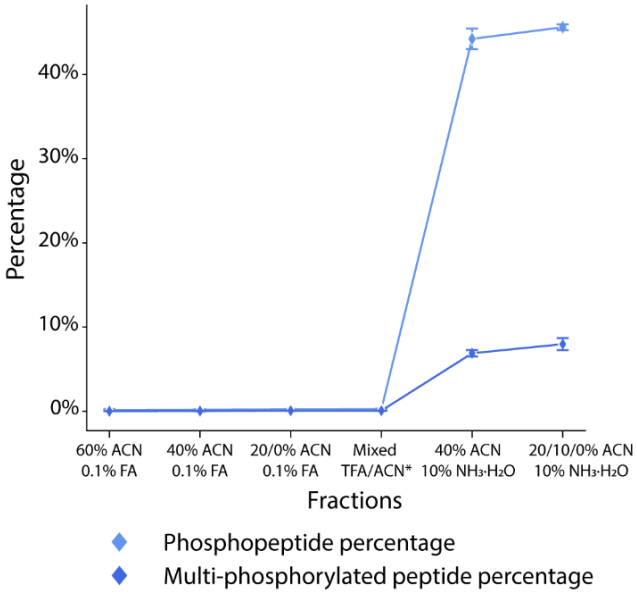


Figure S6. Phosphopeptide and multi-phosphorylated peptide percentages in six fractions with simultaneous enrichment of glycopeptides and phosphopeptides from mouse lung tissue.

Table S1. MALDI-TOF MS analysis of phosphopeptides enriched from β-casein and α-casein tryptic digests.

Protein	Amino acid sequence	No. of phosphosites	Theoretical $[M+H]^+$	Observed $[M+H]^+$
β-casein	FQ[pS]EEQQQTEDELQDK	1	2061.8284	2061.8844
	ELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR	4	2966.1645	2966.2463
	RELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR	4	3122.2656	3122.3700
α-casein	EQL[pS]T[pS]EENSK	2	1411.5026	1411.5397
	TVDME[pS]TEVFTK	1	1466.6120	1466.6463
	VPQLEIVPN[pS]AEER	1	1660.7941	1660.8523
	YLGEYLIVPN [pS]AEER	1	1832.8465	1832.9064
	DIGSE[pS]TEDQAMEDIK	1	1847.7252	1844.9436
	DIG[pS]E[pS]TEDQAMEDIK	2	1927.6915	1927.7441
	DIG[pS]E[pS]TEDQA[Mo]EDIK	2	1943.6864	1943.7354
	YKVPQLEIVPN[pS]AEER	1	1951.9524	1952.0156
	VNEL[pS]KDIG[pS]E[pS]TEDQAMEDIK	3	2678.0228	2678.0426
	Q*MEAE[pS]I[pS] [pS] [pS]EEIVPN[pS]VEAQK	5	2703.8864	2703.9050
	QMEAE[pS]I[pS][pS][pS]EEIVPNPN[pS]VEQK	5	2720.9129	2720.9132
	Q[Mo]EAE[pS]I[pS][pS][pS]EEIVPNPN[pS]VEQK	5	2736.9077	2736.9211
	EKVNEL[pS]KDIG[pS]E[pS]TEDQAMEDIK	3	2935.1604	2935.1602
	NANEEEYSIG[pS][pS][pS]EE[pS]AEVATEEVK	4	3008.0295	3008.0250

[a][pS], phosphorylated site; [b][Mo], oxidation on methionine; [c]Q*, pyroglutamylation on the N-terminal Gln.

Table S2. MALDI-TOF MS analysis of glycopeptides and phosphopeptides enriched from mixture of RNase B and β-casein tryptic digests.

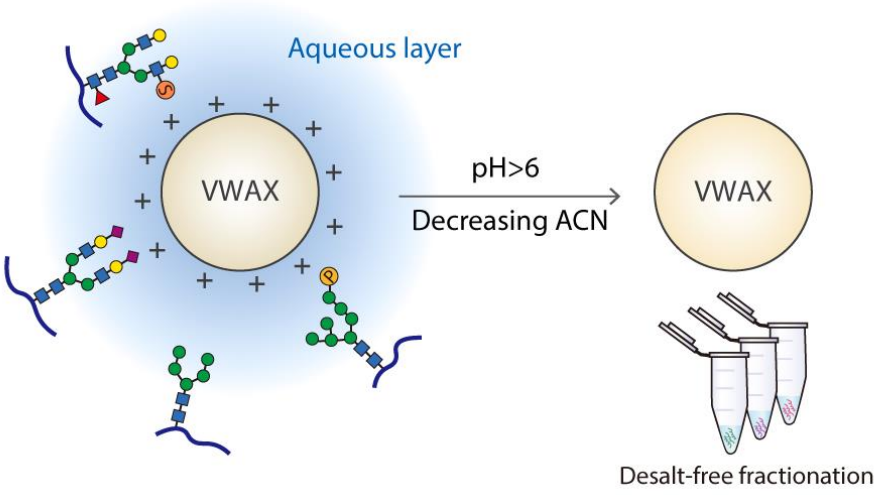
Protein	Enzyme	Peptide sequence	Glycan composition	Glycosylation site	Theoretical $[M+H]^+$	Observed $[M+H]^+$
Rnase B	trypsin	NLTK	GlcNAc2 Man5	Asn-60	1691.7104	1691.6906
Rnase B	trypsin	NLTK	GlcNAc2 Man6	Asn-60	1853.7632	1853.7420
Rnase B	trypsin	SRNLTK	GlcNAc2 Man5	Asn-60	1934.8435	1934.8459
Rnase B	trypsin	NLTKDR	GlcNAc2 Man5	Asn-60	1962.8384	1962.8361
Rnase B	trypsin	NLTK(cam)DR	GlcNAc2 Man5	Asn-60	2005.8442	2005.8435
Rnase B	trypsin	NLTK	GlcNAc2 Man7	Asn-60	2015.8160	2015.7935
Rnase B	trypsin	SRNLTK	GlcNAc2 Man6	Asn-60	2096.8963	2096.9000
Rnase B	trypsin	NLTKDR	GlcNAc2 Man6	Asn-60	2124.8912	2124.8757
Rnase B	trypsin	NLTK(cam)DR	GlcNAc2 Man6	Asn-60	2167.8970	2167.9052
Rnase B	trypsin	NLTK	GlcNAc2 Man8	Asn-60	2177.8688	2177.8453
Rnase B	trypsin	SRNLTK	GlcNAc2 Man7	Asn-60	2258.9491	2258.9512
Rnase B	trypsin	NLTKDR	GlcNAc2 Man7	Asn-60	2286.9440	2286.9121
Rnase B	trypsin	NLTK	GlcNAc2 Man9	Asn-60	2339.9217	2339.8881
Rnase B	trypsin	SRNLTK	GlcNAc2 Man8	Asn-60	2421.0019	2420.9995
Rnase B	trypsin	NLTKDR	GlcNAc2 Man8	Asn-60	2448.9968	2448.9637
Rnase B	trypsin	NLTK(cam)DR	GlcNAc2 Man8	Asn-60	2492.0026	2492.9817
Rnase B	trypsin	SRNLTK	GlcNAc2 Man9	Asn-60	2583.0548	2583.0441
Rnase B	trypsin	NLTKDR	GlcNAc2 Man9	Asn-60	2611.0497	2611.0323

cam: carbamylation at lysine

Protein	Peptide sequence	No. of phosphosites	Theoretical $[M+H]^+$	Observed $[M+H]^+$
β-casein	FQ[pS]EEQQQTEDELQDK	1	2061.8284	2061.8234
	ELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR	4	2966.1645	2966.1020
	RELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR	4	3122.2656	3122.1776

Chapter 4

Very Weak Anion Exchange (VWAX) Chromatography for Glycopeptide Enrichment and Separation



This work was performed in collaboration with Dr. Andrew J. Alpert from PolyLC Inc.

Abstract

The low abundance and poor ionization efficiency of glycopeptides necessitate efficient enrichment prior to mass spectrometry (MS) analysis. Although numerous enrichment methods have been developed, the detection and identification of glycopeptides with negatively charged glycans can still be impaired by the more abundant neutral glycopeptides in positive ion mode LC-MS/MS analysis. Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) addresses this limitation by combining charged stationary phases with hydrophilic interaction chromatography (HILIC) mobile phases. However, conventional anion-exchange materials usually lose the positive charge in the range of 9–12 and require high salt concentration gradients, which are incompatible with MS analysis without additional desalting steps. In this study, we introduce novel Very Weak Anion Exchange (VWAX) chromatography materials that lose their positive charge by pH 6, facilitating ERLIC-mode peptide separation and fractionation with a conventional binding capacity but a pH gradient using volatile salts at much lower concentration. Starting from 75% ACN, glycopeptides are well-retained on the stationary phase. Peptides with negatively charged glycans including sialylated, mannose-6-phosphate (M6P), and sulfated glycopeptides, elute in later fractions, which separates them from neutral glycopeptides and enhances their detection. Successful glycopeptide enrichment is achieved from mixtures of tryptic digests of both glycoprotein standards and cell lysates. This method enables direct MS analysis without desalting, greatly simplifying the experimental workflow and reducing sample loss. In total, 2255 N-glycopeptides were identified across eight fractions from 200 µg of HeLa tryptic digests, including 682 sialyl-glycopeptides, 144 M6P glycopeptides and 17 sulfated glycopeptides.

Our approach offers a powerful tool to profile the glycoproteome from complex biological systems and sheds light on less-studied glycosylation types.

Introduction

Glycosylation is a common post-translational modifications (PTMs) and is an essential determinant of protein activity and function.[1] Over 50% of mammalian proteins are estimated to be glycosylated, and glycoproteins are involved in many intra- and intercellular processes, including cell–cell recognition, immune response, and host–pathogen interactions.[2,3] Aberrant glycosylation has been associated with the pathological processes of various diseases, such as neurodegenerative diseases, autoimmune disease, and cancers.[4–6] However, the complexity of glycosylation has pose a challenge for comprehensive analysis. Glycosylation can occur on multiple glycosites, and multiple glycan structures can exist on the same site, leading to heterogeneity at both the site and glycan levels.[7] Previous studies have focused on either glycan structures or glycosites, which only provides partial information on glycosylation.[8] In contrast, intact glycopeptides has become an ideal research target because they retain information on both glycosylation sites and glycan structures, which enables site-specific investigation this PTM across the whole proteome.

Liquid chromatography tandem mass spectrometry (LC-MS/MS) has emerged as a powerful tool for profiling PTMs on a large scale.[9] However, detection of intact glycopeptides remains challenging, primarily due to their suppressed ionization efficiency and low abundance among the bulk of unmodified peptides after enzymatic digestion.[10,11] Therefore, an efficient enrichment is necessary prior to MS analysis.

To date, many techniques have been reported for N-linked glycopeptide enrichment. Hydrazide chemistry-based enrichment has been used to capture glycans on glycopeptides through oxidation and solid-phase immobilization. Despite the high selectivity, subsequent analyses require the release of glycans, which eliminates the glycan information.[12–14] Affinity-based methods, such as lectin affinity chromatography and metal oxide affinity chromatography, have also been widely used.[15,16] However, these methods tend to target at specific types of glycans and are unable to profile all glycopeptides in a single enrichment. Alternatively, hydrophilic interaction chromatography (HILIC) is another long-established and widely used tool for glycopeptide enrichment, where glycopeptides can be separated from other peptides based on the hydrophilic interaction between the stationary phase and the hydroxyl groups of the glycans.[17,18] HILIC possesses several advantages, including no bias toward diverse glycopeptides, satisfactory MS compatibility, low cost, and high throughput.[19] However, it has been observed that in positive ion mode LC-MS/MS analysis, signals of glycopeptides carrying a negatively charged glycan might still be suppressed by the more abundant neutral glycopeptides, which can impede their detection and identification.[20,21] To address this issue, a variant of HILIC called electrostatic repulsion hydrophilic interaction chromatography (ERLIC) has been developed. ERLIC, which was first introduced by Alpert in 2008,[22] is a mode of separation utilizing charged stationary phases with HILIC mobile phases, where hydrophilic interaction and electrostatic interaction between analytes and stationary phase are superimposed. The charge states of analytes can be adjusted by altering the pH of the mobile phase (MP) to control the electrostatic interaction between the analytes of interest and the stationary phase. For glycopeptides, ERLIC typically employs anion exchange

(AX) materials as stationary phases, and buffer pH to control the charges of amino acid residues on peptide backbones and acidic groups on glycans. This extra degree of tunability provides ERLIC with some advantages over conventional HILIC, particularly for enrichment and separation of negatively charged glycopeptides.[18]

Since its first implication to glycopeptide enrichment by Lewandrowski et al.,[23] ERLIC has been successfully applied to large-scale profiling of N- and O-glycoproteomes, as well as simultaneous enrichment of glycopeptides and phosphopeptides through fractionation in sample preparation.[24–27] These approaches have demonstrated the effectiveness of ERLIC in enriching glycopeptides, especially those with negative charges. However, an increasing salt concentration is required during the fractionation process to disrupt the ionic interaction between the charged analytes and the stationary phase, which is not compatible with MS analysis. Desalting is a necessary step after fractionation, but it is time-consuming and introduces sample loss, especially for hydrophilic or highly charged peptides with PTMs.[28]

Anion-exchange materials are generally categorized as strong anion-exchange (SAX) or weak anion-exchange (WAX) materials, with SAX materials retaining positive charge up to pH 12 and WAX materials retaining positive charge up to pH 9-9.5.[29,30] The terms “weak” and “strong” refer to the range of pH over which the materials lose their positive charge, not the interaction strength with negatively charged analytes. The development of AX materials has historically favored the use of DEAE-groups or equivalents based on their capacity to maintain a high charge density over a wide pH range since the first work of Peterson and Sober.[31] However, this feature may bring up new challenges, as the use of mass spectrometry (MS) for detection is now widespread and the

high salt concentration gradients are not desired for MS analysis. For analysis of highly charged analytes, an alternative to a high salt gradient is a gradient of pH. This either changes the charge on the analytes or uncharges the stationary phase. However, the extensive use of silica-based materials for life science chromatography imposes an upper limit of pH 8 for sustained use, which is inadequate for completely uncharging WAX materials. To address this issue, S. Bäurer et al introduced a material consisting of silica derivatized with a silane containing a pyridylurea ligand.[32] This material exhibits a net positive charge below pH 6.5 and a net negative charge above that pH, allowing for elution of negatively charged analytes with very low salt levels using gradients from pH 5–7. Despite the innovation, two issues persist with its implementation: First, by using a thin, silane-based coating, the underlying negatively charged silanol groups appear to have had access to the pyridylurea groups. This permits the oppositely charged groups to titrate each other, thereby lowering the positive charge density significantly. A similar situation was seen in the work of Mant et al,[33] where a layer of sulfonate groups steadily lost capacity as an underlying polyamine layer's charge density increased with decreasing pH. Second, the proximity of the pyridylurea groups in the coating means that the ease of protonation of one group can be affected by the charge on neighboring groups. Such situations result in titration curves featuring loss of charge in a continuum over a wide pH range.[29]

To circumvent these limitations, we introduce experimental materials termed Very Weak Anion-Exchange (VWAX) chromatography, which become uncharged more readily than conventional anion-exchange materials. We have examined a different set of functional groups, and present here a set of AX materials that have a neutral net charge above pH 6 rather than

negative. With the VWAX column, well-retained peptides are eluted with significantly less salt. We successfully developed a desalting-free fractionation method with volatile MPs, allowing the separation of neutral glycopeptides and negatively charged glycopeptides based on the pH gradient. Peptides with negatively charged glycans eluted in later fractions, which enhanced their detection. With this approach, we mapped intact sialyl, mannose-6-phosphate (M6P) and sulfated glycopeptides from the tryptic digests of both protein standard and cell lysate samples. Our study provides a new tool for in-depth profiling glycoproteomes from complex biological systems and offers new insights into those underexplored glycosylation types.

Experimental Section

Chemicals. Optimal LC/MS grade solvents, Tris base, urea, formic acid (FA), acetic acid (AA), ammonium acetate, ammonium bicarbonate (ABC), calcium chloride ($CaCl_2$) and sodium chloride (NaCl) were purchased from Fisher Scientific (Pittsburgh, PA). Bovine fetuin, thyroglobulin from bovine thyroid (BTG), Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), triethylammonium bicarbonate (TEAB), sodium dodecyl sulfate (SDS), iodoacetamide (IAA), 4-(2-Aminoethyl) morpholine and histamine free base were purchased from Sigma-Aldrich (St Louis, MO). 4-(2-Aminoethyl) pyridine was purchased from Oakwood Products (Estill, SC). Dithiothreitol (DTT) and sequencing grade trypsin were from Promega (Madison, WI). WAX and VWAX columns (100x2.1 mm; 3-µm, 300-Å) was obtained from PolyLC (Columbia, MD). Protease inhibitor cocktail tablets and phosphatase inhibitor cocktail tablets were from Roche (Mannheim, Germany). Dulbecco's modified eagle medium (DMEM) and PBS buffer (pH 7.4) was from Cytiva

(Marlborough, MA). Synthetic phosphopeptide standards were from United BioSystems Inc. (Herndon, VA, USA).

VWAX Material Preparation. The VWAX column materials were prepared in accordance with a previous publication, with some modifications.[34] The initial step involved attaching the aminopropyl-silane to the silica surface. Subsequently, a solution of the reactive polymer polysuccinimide was introduced to the silane-coated silica, to form covalent amide bonds between the amino groups of the silane and some of the succinimide rings of the polymer. Finally, the remaining succinimide rings were reacted with a reagent containing a primary, aliphatic amine, to incorporate the functional group into the coating. which included derivatives of pyridine, histamine, and morpholine,

Cell Culture. HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum (Gibco), 1% penicillin/streptomycin (Sigma Aldrich), and incubated at 37 °C in a humidified chamber with 5% CO_2. Cells were harvested by washing with PBS buffer.

Sample Preparation. HeLa cell pellet was lysed in extraction buffer composed of 50 mM Tris base (pH 7.4), 4% SDS, 65 mM DTT, 175 mM NaCl, and 1% (v/v) protease and phosphatase inhibitor cocktail. The lysates were subjected to probe sonication in an ice water bath at 50% power, with a pulsing cycle of 5 seconds on and 5 seconds off for a total of 12 cycles. Following sonication, the cell lysates were centrifuged at 18000 ×g for 5 min, and the resulting supernatant was mixed with ice-cold precipitation buffer (acetone: ethanol: acetic acid=50: 50: 0.1) at a 1:5 ratio. The mixture was incubated overnight at -20°C for protein precipitation. Protein pellets were obtained by centrifuging at 18000 ×g for 10 min and washing twice with ice-cold precipitation buffer. The

pellets were air-dried in the fume hood for 10 min before being reconstituted in an 8M urea/50 mM TEAB buffer (pH 8.0). Protein concentrations were determined using a BCA assay kit from Thermo Fisher Scientific (San Jose, CA).

Protein digestion. The protein digestion process was adapted from the filter-aided sample preparation (FASP) protocol with some modifications.[35] For standard proteins, 1 mg of protein was dissolved in 500 µL of 50 mM TEAB buffer and heat-denatured at 95°C for 10 min. Proteins were reduced in 10 mM TCEP at room temperature for 30 min and alkylated with 20 mM IAA in the dark for 30 min. The protein solution was transferred to a 30k MWCO filter unit and centrifuged at 14000 ×g for 10 min. The filter was washed thrice with 250 µL of 50 mM TEAB and centrifuged at 14000 ×g for 20 min each time. Afterward, the filter was moved to a new collection tube, and the sample was resuspended with 250 µL of 50 mM TEAB. Trypsin digestion was performed at a 50:1 protein-to-enzyme ratio and incubated at 37 °C for 18 h. Peptide digests were collected via centrifugation at 14000 ×g for 10 min, and the process was repeated once with an additional 50 µL of water. The resulting samples were dried down *in vacuo* and stored at -80°C for future use.

Offline ERLIC Fractionation. For both VWAX and WAX experiments, binary MPs were used for LC separation. Mobile phase A consisted of 10 mM FA with a high concentration of ACN in water. Various concentrations of ACN, ranging from 70% to 80%, were evaluated. Mobile phase B comprised 50 mM ammonium acetate in 10% ACN. 50 µg of peptides from protein standard digests or 200 µg of HeLa cell digests were fractionated with the following gradient at a flow rate of 0.2 mL/min: 0–10 min, 0% B; 10–35 min, 0-100% B; 35–40 min, 100% B; 40–50 min, 0% B.

The eluent was collected using a FC-4 fraction collector (Rainin Dynamax) at 2-min intervals. The fractions were then combined into fewer tubes for subsequent MS analysis.

NanoLC-MS/MS analysis. An in-house packed nano-C18 column (15 cm length, 75 μm i.d.,) filled with Bridged Ethylene Hybrid C18 material (1.7 μm, 130 Å, Waters) was used for sample separation. Initial MS analysis of bovine fetuin and BTG samples on the Q Exactive mass spectrometer are provided in Supplemental Information. Both BTG and HeLa cell samples were analyzed via LC–MS/MS using an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) connected to a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). The BTG and HeLa peptides were eluted using a 36-min and 80-min gradient from 0 to 30% ACN (0.1% FA), respectively, with a flow rate of 0.3 μL/min. Data acquisition was conducted in top-speed mode with a cycle time of 3s. Survey scans of peptide precursors from *m/z* 400 to 2000 were performed at a resolving power of 60 K and an AGC target of 4E5, with a maximum injection time of 100 ms. Precursors were chosen for higher-energy C-trap dissociation (HCD) analysis with a normalized collision energy (NCE) of 30 and ± 3% stepped HCD collision energy. Tandem MS acquisition parameters included a resolving power of 60 K, lower mass limit of 120 *m/z*, and dynamic exclusion of 12 s with 10 ppm mass tolerance.

Data Analysis. Byonic software (version 4.5.2, Protein Metrics Inc, San Carlos, CA) was employed to process intact N-glycopeptide data. Raw files were searched against the bovine fetuin protein sequences, the BTG sequence, or the UniProt *Homo sapiens* reviewed database (August 2020, 20311 sequences). Trypsin was chosen as the enzyme, and a maximum of two missed cleavages were permitted. Searches were performed with a precursor mass tolerance of 10 ppm

and a fragment mass tolerance of 0.01 Da. Fixed modifications were specified as carbamidomethylation (+57.02146 Da) on cysteine residues. Dynamic modifications included oxidation of methionine (+15.99492 Da, rare2) and N-glycosylation (common1). Glycan modifications were searched against a glycan database expanded from Byonic embedded mammalian N-glycan database (309 entries). For BTG, potential sulfated glycan compositions were included based on previous report.[36] For HeLa cell samples, N-linked M6P glycans comprising HexNAc (2–4) Hex (3–9) Phospho (1–2) modifications were incorporated. Possible sulfated glycan compositions were generated by adding 1–2 sulfate groups to the original complex/hybrid glycans in the database. For protein standards, the peptide identification results were filtered at FDR < 1% and Byonic score > 150. For complex biological samples, peptide identification results were filtered at FDR < 1%, Byonic score > 150, PEP 2D < 0.05, and |Log Prob| > 1. MS/MS spectra of M6P glycopeptide PSMs were manually inspected to confirm the presence of phosphorylated hexose diagnostic ions.

Results and Discussion

Rationale of VWAX. To facilitate the elution of highly charged analytes using a pH gradient rather than a high salt concentration gradient, our objective was to develop an AX material that loses its positive charge at lower pH values appreciably lower than those of conventional WAX materials, and the opposite of SAX materials. In this study, we prepared three VWAX materials containing aliphatic amine groups, which are expected to lose their positive charge at around pH 6. These materials include: PolyHISTAMINE A, prepared with histamine; PolyAE-PYRIDINE A,

prepared with 4-Aminoethyl-pyridine; and PolyAE-MORPHOLINE A, prepared with 4-Aminoethyl-morpholine.

To avoid the interference to the charge density caused by the underlying negatively charged silanol groups and the proximity of functional groups, the VWAX materials were designed with anionic (positively charged) functional groups within a thick polymeric coating covalently attached to the surface. This design ensures that the functional groups are situated far enough from the surface to avoid titration by the silanols' negative charge. Furthermore, the polymeric coating shields the underlying silanols, a feature not present in the coating prepared by S. Bäurer et al.[32] Additionally, silane-based coatings are two-dimensional, which restricts their capacity. Integrating positively charged groups into a thick polymeric coating overcomes this limitation, enabling the anionic groups to outnumber the cationic silanols. As a result, the chromatography is dominated by the properties of the positively charged functional group embedded the coating.

Figure 1 displays the titration curves of the three VWAX materials, measured with salicylic acid (pKa 2.97) and a mobile phase of 20 mM Na_3PO_4 at different pH levels with 10% ACN. Retention peaks around pH 4 and declines sharply above pH 5. Compared to the three VWAX materials, the WAX material (PolyWAX LP) maintains significant charge density to much higher pH values. This suggests that the VWAX material holds the potential to facilitate the elution of well-retained peptides with considerably less salt than required for standard WAX or SAX materials. It can also be observed that the pyridine-containing material loses its positive charge cleanly at lower pH values than other two materials.

Evaluation of IEX/ERLIC with Peptide Standards. Figure 2 presents a comparison of the separation of synthetic peptides with 1, 2 or 3 phosphate groups on a regular PolyWAX LP column versus the three VWAX columns, using a salt and pH gradient. MP A consisted of 20 mM ammonium formate in either 65% or 70% ACN with pH adjusted to 3.0, while MP B contained 0.4 M ammonium acetate in 10% ACN (pH 7). In 65% ACN, the difference in retention times between the PolyWAX LP column and the three VWAX columns is dramatic, particularly for peptides with 2 or 3 phosphates. All phosphopeptides were eluted within 30 min on VWAX columns, whereas the WAX column retained the phosphopeptides for longer time and did not elute all phosphopeptides during the gradient. This demonstrates the significant advantage of VWAX materials in eluting strongly retained analytes with considerably less salt. Similar comparisons were also conducted at 70% ACN. It is noteworthy that increasing ACN from 65% to 70% resulted in a significant increase in the retention of the peptide with 1 phosphate on any column. Its retention reflects a combination of electrostatic attraction and hydrophilic interaction.[22] However, increasing ACN from 65% to 70% ACN had minimal effect on the retention of peptides with 2 or 3 phosphates, suggesting their retention was dominated by electrostatic effects.

Evaluation with Glycoprotein Standards. Given the ERLIC-mode retention of phosphopeptide standards on the VWAX columns, we reasoned that this mode would also be suitable for N-glycopeptides, especially for those carrying negatively charged glycans. Considering that pyridine-containing material lost its positive charge at lower pH than other two materials, the PolyAE-PYRIDINE A column was chosen for subsequent evaluation of glycopeptide enrichment. Tryptic digests prepared from bovine fetuin and thyroglobulin from bovine thyroid (BTG) were employed

as glycoprotein standards, as bovine fetuin is known to exhibit sialylation at multiple sites,[37–39] and BTG features more diverse glycan types, including neutral, sialylated, and sulfated glycans.[36,40] Since multiple interactions, such as hydrophilic interaction and electrostatic interaction, work simultaneously in ERLIC, it is essential to select optimized MPs for glycopeptide separation. To enable a desalt-free offline fractionation method, only volatile components were considered in MPs.

For initial testing, we compared two MP A compositions at 80% and 70% ACN, both with 10 mM FA added, while consistently using 50 mM ammonium acetate as MP B. The inclusion of FA in MP A generated a pH between 2.5–2.6, which was expected to protonate the majority of peptides with acidic residues while leaving the sialic acid, phosphate, and sulfate groups charged, as aspartic acid and glutamic acid have pKa values of 3.9 and 4.1, respectively. The gradient was 0–5 min at 0% B followed by 0–100% B in 25 min. Fractions were pooled into 6 tubes, as illustrated in **Figure S1** for direct LC-MS analysis on the Q Exactive orbitrap instrument. **Figure 3** displays the number of total PSMs, total glycol-PSMs (GPSMs), sialyl-glycopeptide PSMs (Neu-PSMs), and sulfated glycopeptide PSMs (Sulfated-PSMs). Though 80% ACN was proven to be optimal for glycopeptide enrichment in our previous study on SAX materials,[26] MP A with 70% ACN appeared to perform better than the one with 80% ACN in this VWAX study, as more peptides were eluted in the first five minutes and more GPSMs were detected in the later fractions. Since a higher ACN concentration might be overly retentive, causing many non-modified peptides with minimal hydrophilicity to be co-retained with glycopeptides,[26] it is plausible that 80% ACN results in such strong retention for the current VWAX material. **Figure S2** shows the distribution of

identified peptides in fetuin with varying numbers of Asp and Glu residues. It can be observed that more peptides with multiple acidic residues were retained in 70% ACN, while no obvious trend was seen in 80% ACN. This suggests that a MP with lower ACN concentration could wash off more non-specific binding that would present in 80% ACN and the electrostatic interaction played a more significant role in 70% ACN on this VWAX column.

Following the initial profiling, BTG was selected as a model to further optimize the MPs and gradient. To minimize the interference from non-specific binding, the duration of 100% MP A was increased from 5 to 10 minutes, allowing more time to remove the bulk unmodified peptides from the column. The separation of BTG digests in 80% ACN MP A was repeated with an adjusted gradient: 0–10 min, 0% B; 10–20 min, 0-14% B; 20–45 min, 100% B. The collected samples were analyzed on the Fusion Lumos orbitrap instrument. The resulting chromatogram and identification numbers are shown in **Figure S3**. Although the bulk peak was eluted in the first 10 min, the identification of GPSMs remained low, following a similar pattern to the previous experiment. Adjusting the gradient time did not significantly enhance glycopeptide enrichment, suggesting that ACN content was the major variable in this case. We further compared MP A with 70% and 75% ACN using the following gradient: 0–10 min, 0% B; 10–35 min, 0-100% B. **Figures S4A** and **S4B** present the corresponding chromatograms, where the majority of peptides were eluted in the first 10 min under both conditions. Surprisingly, unlike the considerable disparity between 70% ACN and 80% ACN, the identification of PSMs (**Figures 4A, 4B**) shows a similar pattern. There is a clear trend that peptides with a neutral glycan eluted first, and the peptides with negatively charged glycans tended to elute in latter fractions due to their additional electrostatic interaction with the

VWAX stationary phase. When examining the total unique glycopeptides identified (**Figures S4C, S4D**), it turns out that the 75% ACN enriched more unique glycopeptides, especially for neutral glycopeptides, as a result of the stronger hydrophilic interaction. This also suggests that 75% ACN in MP A established a better balance between electrostatic attraction and hydrophilic interaction.

It is noteworthy that the VWAX column successfully enriched both sulfated and sialylated glycopeptides in the last two fractions. Proteins carrying sulfated glycans are associated with various diseases, such as cancer,[41] osteoarthritis,[42] and rheumatoid arthritis.[43] Sulfation typically occurs on complex N-glycans, with a sulfate localized on C-6 of N-acetylglucosamine (GlcNAc), C-4 of N-acetylgalactosamine (GalNAc), and C-3 of galactose (Gal) of terminal GalNAc-GlcNAc and Gal-GlcNAc residues.[44] Analyzing intact sulfated glycopeptides remains challenging due to the lack of efficient enrichment methods and MS signal suppression caused by their negative charge. Ion pairing methods have been introduced to enhance the signal of sulfated glycopeptides from hormone proteins.[45,46] Toyoda et al. proposed a method combining protease digestion, chemical derivatization, and SAX separation to characterize intact sulfated N-glycopeptides.[47,48] However, these methods either focus on single proteins or require multiple processing steps, making them unsuitable for high-throughput analysis. In contrast, the VWAX method demonstrated its potential for large-scale profiling of intact sulfated glycopeptides from peptide mixtures. The sulfate group has a pKa of 1, maintaining its charge under the current MP A conditions.[49] These peptides were attracted to the VWAX stationary phase until the stationary phase became uncharged as pH increased. Among the two conditions (70% ACN and 75% ACN), a total of 24 sulfated-PSMs were identified. Six out of these contained the diagnostic ion

(SulfoHexNAc, *m/z* 284.044) in the spectra, which significantly enhanced the identification confidence. One example spectrum is shown in **Figure 4C.** Our observation also aligns with the previous report that only a small portion of sulfated-PSMs contained the diagnostic oxonium ion upon HCD fragmentation, and the occurrence became fewer when sialic acids were present on the same glycopeptide.[36]

Glycopeptide Enrichment from Complex Biological Samples. The VWAX approach was adapted to enrich more diverse glycopeptides from a more complex biological sample. 200 µg of HeLa tryptic digests were loaded to the VWAX column with the optimized gradient and MPs, starting either from 75% ACN or 70% ACN with 10 mM FA. The fractions were pooled into eight tubes for the MS analysis (**Figure S5A** and **Figure S6A**). During the database search, another less studied glycosylation, M6P glycosylation, was included in the glycan database along with common N-glycans and sulfated N-glycans. M6P glycosylation results from phosphorylation of glycans at the C-6 position of mannose in high-mannose glycans, playing an important role in transferring lysosomal hydrolases to the lysosome and many other important biological processes.[50,51] Similar to sulfated glycopeptides, the phosphate group has a negative charge and a pKa value of 0.7–1.0, which can cause signal suppression during MS detection if not well-separated from other glycopeptides during enrichment.[20,51] It is expected that M6P glycopeptides will also be captured on the VWAX stationary phase, and elute after the neutral glycopeptides, once the electrostatic interaction is disrupted.

The identification results are displayed in **Figure 5A, Figure S5B** and **Figure S6B,** and **Figure S6C**. Consistent with previous observation, the MP with 75% ACN resulted in a higher

identification number of glycopeptides compared to the MP with 70% ACN, particularly for neutral glycopeptides, suggesting that enhanced hydrophilic interaction is crucial for retaining these glycopeptides. There was a clear trend that neutral glycopeptides eluted first, and the negatively charged glycopeptides only eluted in the later fractions, demonstrating an ERLIC-mode separation mechanism. All M6P and sulfated GPSMs were manually inspected to screen for the presence of signature oxonium ions. While M6P glycopeptides commonly generate phosphorylated hexose diagnostic ions, the identified sulfated glycopeptides rarely have the sulfate-containing diagnostic ions. This might be due to the low sulfate-oxonium ion generation rate or low intensity, as observed from the BTG standards. Therefore, we categorized the identified sulfated glycopeptides as tentative matches, and further validation is necessary to confirm their existence in HeLa cells.

Another interesting observation was made during the manual inspection of sulfated glycopeptides. We found that several sulfated PSMs actually belonged to M6P peptides, which contained the diagnostic phosphorylated oxonium ions at *m/z* 243.027. Sulfation (SO_3, +79.9569) and phosphorylation (HPO_3, +79.9663) have almost identical mass, making them difficult to distinguish by MS. Since glycan phosphorylation was typically considered to occur on high mannose glycans,[51] only these glycans were included as M6P glycans in the glycan database during the search, leading Byonic to mismatch the PSMs to the isobaric sulfated glycopeptides. Although previous studies have reported the detection of phosphorylated hybrid glycans in gastric cancer cells and human serum,[52,53] no study has reported the detection of intact glycopeptides with such glycans due to their extremely low abundance.[54] Here we successfully identified a glycoprotein,

PLOD2 (Uniprot accession: O00469), carrying both phosphorylated high mannose and hybrid glycans on site N63. One example PSM is shown in **Figure 5B**. This protein has been found to be overexpressed in different cancers and closely related to a poor prognosis.[55] Further investigation of its M6P glycosylation may provide deeper insights into its biological roles in disease progression.

Overall, 2255 N-glycopeptides were identified among the eight fractions, including 682 sialyl glycopeptides, 144 M6P glycopeptides and 17 tentative sulfated glycopeptides. The total glycoproteome coverage was comparable with our previously reported dataset generated by the Epoxy-ATP-Ti^{4+} dual-functional IMAC enrichment approach (see **Chapter 3**), with a higher number of sialyl glycopeptides and M6P glycopeptides identified, highlighting the advantages of the VWAX method in mapping these negatively charged glycopeptides.

Comparison Between WAX and VWAX. The optimized fractionation method was tested on both the conventional WAX column and the VWAX column with 200 µg of HeLa tryptic digests. As depicted in **Figure S7**, the VWAX column enriched a considerably higher number of glycopeptides in each fraction, exhibiting greater specificity. The difference was more pronounced for negatively charged species eluted in the last three fractions. VWAX significantly outperformed WAX in enriching sialyl and M6P glycopeptides. Among all sialyl glycopeptides enriched by VWAX, 25.2% and 2.9% of them possessed two or three sialic acids, respectively, while only 6.1% of sialyl glycopeptides had two sialic acids and no tri-sialylated glycopeptides were detected when enriched by WAX. **Figure S8** illustrates the distribution of identified peptides with varying numbers of Asp and Glu residues for the two methods. With the current MPs, the WAX column

may be prone to a higher degree of non-specific binding from peptides containing acidic amino acid residues. Using the current MP B, the WAX column is unable to fully neutralize its positive charge, thus continuing to retain more acidic targets of interest and resulting in a lower glycoproteome coverage.

Conclusions

In this work, we introduced new experimental materials for VWAX chromatography, which possess the similar binding capacity but neutralizes at lower pH compared to conventional AX materials. This allows the use of a pH gradient for elution instead of a high salt concentration gradient, enabling MS-compatible, desalt-free sample preparation. We assessed the ERLIC properties of the VWAX column and optimized the MPs and gradient for glycopeptide enrichment. Glycopeptides were successfully enriched from mixture of tryptic digests from both glycoprotein standards and cell lysates. The ERLIC-mode chromatography mechanism not only separated the glycopeptides from the bulk of unmodified peptides but also enabled simultaneous separation of neutral glycopeptides, sialyl glycopeptides, M6P glycopeptides, and sulfated glycopeptides. Detection of these negatively charged glycopeptides was enhanced, thereby improving the glycoproteome coverage. Notably, we detected intact M6P glycopeptides with hybrid glycan structures from HeLa cell digests, which, to our knowledge, is the first report of such findings in complex biological mixtures like cell lysates. VWAX demonstrated considerably higher specificity for glycopeptides and facilitated easier elution of glycopeptides with negatively charged glycans compared to the conventional WAX column using the optimized mobile phases composed of low concentrations of volatile salts. This approach allows for direct sample analysis on MS

without the need for desalting, greatly simplifying the experimental workflow and reducing sample loss. Collectively, we believe this new material and workflow will serve as a useful tool for profiling the glycoproteome from complex biological systems and offer new insights into those less-studied glycosylation types.

Acknowledgements

This work was supported, in part, by the National Institutes of Health Grants RF1AG052324, R01 AG078794, and R01 DK071801 (to L.L.). The Orbitrap instruments were purchased through the support of an NIH Shared Instrument Grant (NIH-NCRR S10RR029531 to L.L.) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Reference

(1) Patrie, S. M.; Roth, M. J.; Kohler, J. J. Introduction to Glycosylation and Mass Spectrometry. In *Mass Spectrometry of Glycoproteins*; Methods in Molecular Biology; Humana Press: Totowa, NJ, **2013**; Vol. 951, pp 1–17.

(2) Apweiler, R.; Hermjakob, H.; Sharon, N. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database11Dedicated to Prof. Akira Kobata and Prof. Harry Schachter on the Occasion of Their 65th Birthdays. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473* (1), 4–8.

(3) Gagneux, P.; Hennet, T.; Varki, A. Biological Functions of Glycans. In *Essentials of Glycobiology. 4th edition.*; Cold Spring Harbor Laboratory Press: NY, **2022**.

(4) Pinho, S. S.; Reis, C. A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* **2015**, *15* (9), 540–555.

(5) Abou-Abbass, H.; Abou-El-Hassan, H.; Bahmad, H.; Zibara, K.; Zebian, A.; Youssef, R.; Ismail, J.; Zhu, R.; Zhou, S.; Dong, X.; Nasser, M.; Bahmad, M.; Darwish, H.; Mechref, Y.; Kobeissy, F. Glycosylation and Other PTMs Alterations in Neurodegenerative Diseases: Current Status and Future Role in Neurotrauma. *Electrophoresis* **2016**, *37* (11), 1549–1561.

(6) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.

(7) Fang, P.; Ji, Y.; Oellerich, T.; Urlaub, H.; Pan, K.-T. Strategies for Proteome-Wide Quantification of Glycosylation Macro- and Micro-Heterogeneity. *Int. J. Mol. Sci.* **2022**, *23* (3), 1609.

(8) An, H. J.; Froehlich, J. W.; Lebrilla, C. B. Determination of Glycosylation Sites and Site-Specific Heterogeneity in Glycoproteins. *Curr. Opin. Chem. Biol.* **2009**, *13* (4), 421–426.

(9) Doll, S.; Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **2015**, *10* (1), 63–71.

(10) Nwosu, C. C.; Strum, J. S.; An, H. J.; Lebrilla, C. B. Enhanced Detection and Identification of Glycopeptides in Negative Ion Mode Mass Spectrometry. *Anal. Chem.* **2010**, *82* (23), 9654–9662.

(11) Huang, J.; Wang, F.; Ye, M.; Zou, H. Enrichment and Separation Techniques for Large-Scale Proteomics Analysis of the Protein Post-Translational Modifications. *J. Chromatogr. A* **2014**, *1372*, 1–17.

(12) Zhang, H.; Li, X.; Martin, D. B.; Aebersold, R. Identification and Quantification of N-Linked Glycoproteins Using Hydrazide Chemistry, Stable Isotope Labeling and Mass Spectrometry. *Nat. Biotechnol.* **2003**, *21* (6), 660–666.

(13) Chen, J.; Shah, P.; Zhang, H. Solid Phase Extraction of N-Linked Glycopeptides Using Hydrazide Tip. *Anal. Chem.* **2013**, *85* (22), 10670–10674.

(14) Zhang, Z.; Sun, D.; Cong, Y.; Mao, J.; Huang, J.; Qin, H.; Liu, J.; Huang, G.; Wang, L.; Ye, M.; Zou, H. Amine Chemistry Method for Selective Enrichment of N-Linked Glycopeptides for Glycoproteomics Analysis. *J. Proteome Res.* **2015**, *14* (9), 3892–3899.

(15) Goumenou, A.; Delaunay, N.; Pichon, V. Recent Advances in Lectin-Based Affinity Sorbents for Protein Glycosylation Studies. *Frontiers Mol. Biosci.* **2021**, *8*, 746822.

(16) Larsen, M. R.; Jensen, S. S.; Jakobsen, L. A.; Heegaard, N. H. H. Exploring the Sialiome Using Titanium Dioxide Chromatography and Mass Spectrometry *. *Mol. Cell Proteomics* **2007**, *6* (10), 1778–1787.

(17) Jandera, P. Stationary and Mobile Phases in Hydrophilic Interaction Chromatography: A Review. *Anal. Chim. Acta* **2011**, *692* (1–2), 1–25.

(18) Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J. A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-Based Glycoproteomics. *Mol. Cell Proteomics* **2021**, *20*, 100029.

(19) Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X. Recent Advances in Hydrophilic Interaction Liquid Interaction Chromatography Materials for Glycopeptide Enrichment and Glycan Separation. *TrAC - Trends Anal. Chem.* **2020**, *124*, 115570.

(20) Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

(21) Chen, Y.-J.; Yen, T.-C.; Lin, Y.-H.; Chen, Y.-L.; Khoo, K.-H.; Chen, Y.-J. ZIC-CHILIC-Based StageTip for Simultaneous Glycopeptide Enrichment and Fractionation toward Large-Scale N-Sialoglycoproteomics. *Anal. Chem.* **2021**, *93* (48), 15931–15940.

(22) Alpert, A. J. Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Anal. Chem.* **2008**, *80* (1), 62–76.

(23) Lewandrowski, U.; Lohrig, K.; Zahedi, R. P.; Wolters, D.; Sickmann, A. Glycosylation Site Analysis of Human Platelets by Electrostatic Repulsion Hydrophilic Interaction Chromatography. *Clin. Proteom.* **2008**, *4* (1–2), 25–36.

(24) Zhang, H.; Guo, T.; Li, X.; Datta, A.; Park, J. E.; Yang, J.; Lim, S. K.; Tam, J. P.; Sze, S. K. Simultaneous Characterization of Glyco- and Phosphoproteomes of Mouse Brain Membrane Proteome with Electrostatic Repulsion Hydrophilic Interaction Chromatography. *Mol. Cell Proteomics* **2010**, *9* (4), 635–647.

(25) Hao, P.; Guo, T.; Sze, S. K. Simultaneous Analysis of Proteome, Phospho- and Glycoproteome of Rat Kidney Tissue with Electrostatic Repulsion Hydrophilic Interaction Chromatography. *PLoS ONE* **2011**, *6* (2), e16884-13.

(26) Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L. Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2019**, 1–11.

(27) Cui, Y.; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L. Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Anal. Chem.* **2021**, *93* (22), 7908–7916.

(28) Larsen, M. R.; Graham, M. E.; Robinson, P. J.; Roepstorff, P. Improved Detection of Hydrophilic Phosphopeptides Using Graphite Powder Microcolumns and Mass Spectrometry. *Mol. Cell Proteomics* **2004**, *3* (5), 456–465.

(29) Alpert, A. J.; Regnier, F. E. Preparation of a Porous Microparticulatee Anion-Exchange Chromatography Support for Proteins. *J. Chromatogr. A* **1979**, *185*, 375–392.

(30) Alpert, A. J.; Hudecz, O.; Mechtler, K. Anion-Exchange Chromatography of Phosphopeptides: Weak Anion Exchange versus Strong Anion Exchange and Anion-Exchange Chromatography versus Electrostatic Repulsion–Hydrophilic Interaction Chromatography. *Anal. Chem.* **2015**, *87* (9), 4704–4711.

(31) Peterson, E. A.; Sober, H. A. Chromatography of Proteins. I. Cellulose Ion-Exchange Adsorbents. *J. Am. Chem. Soc.* **1956**, *78* (4), 751–755.

(32) Bäurer, S.; Polnick, S.; Sánchez-Muñoz, O. L.; Kramer, M.; Lämmerhofer, M. N-Propyl-N′-2-Pyridylurea-Modified Silica as Mixed-Mode Stationary Phase with Moderate Weak Anion

Exchange Capacity and PH-Dependent Surface Charge Reversal. *J. Chromatogr. A* **2018**, *1560*, 45–54.

(33) Mant, C. T.; Hodges, R. S. Separation of Peptides by Strong Cation-Exchange High-Performance Liquid Chromatography. *J. Chromatogr. A* **1985**, *327*, 147–155.

(34) Alpert, A. J. Hydrophilic-Interaction Chromatography for the Separation of Peptides, Nucleic Acids and Other Polar Compounds. *J. Chromatogr. A* **1990**, *499*, 177–196.

(35) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6* (5), 359–362.

(36) Kuo, C.-W.; Guu, S.-Y.; Khoo, K.-H. Distinctive and Complementary MS2 Fragmentation Characteristics for Identification of Sulfated Sialylated N-Glycopeptides by NanoLC-MS/MS Workflow. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (6), 1166–1178.

(37) Lin, Y.-H.; Franc, V.; Heck, A. J. R. Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* **2018**, *17* (8), 2861–2869.

(38) Di, Y.; Zhang, L.; Zhang, Y.; Zhao, H.; Yan, G.; Yao, J.; Zhang, S.; Lu, H. MdCDPM: A Mass Defect-Based Chemical-Directed Proteomics Method for Targeted Analysis of Intact Sialylglycopeptides. *Anal. Chem.* **2019**, *91* (15), 9986–9992.

(39) Li, M.; Huang, J.; Ma, M.; Shi, X.; Li, L. Selective Enrichment of Sialylglycopeptides Enabled by Click Chemistry and Dynamic Covalent Exchange. *Anal. Chem.* **2022**, *94* (18), 6681–6688.

(40) Feng, Y.; Chen, B.; Yu, Q.; Zhong, X.; Frost, D. C.; Ikonomidou, C.; Li, L. Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound (SUGAR) Tags: A Probe for Quantitative Glycomic Analysis. *Anal. Chem.* **2019**, *91* (4), 3141–3146.

(41) Magro, G.; Perissinotto, D.; Schiappacassi, M.; Goletz, S.; Otto, A.; Müller, E.-C.; Bisceglia, M.; Brown, G.; Ellis, T.; Grasso, S.; Colombatti, A.; Perris, R. Proteomic and Postproteomic Characterization of Keratan Sulfate-Glycanated Isoforms of Thyroglobulin and Transferrin Uniquely Elaborated by Papillary Thyroid Carcinomas. *Am. J. Pathol.* **2003**, *163* (1), 183–196.

(42) Bayliss, M. T.; Osborne, D.; Woodhouse, S.; Davidson, C. Sulfation of Chondroitin Sulfate in Human Articular Cartilage. *J. Biol. Chem.* **1999**, *274* (22), 15892–15900.

(43) Wang, J.-R.; Gao, W.-N.; Grimm, R.; Jiang, S.; Liang, Y.; Ye, H.; Li, Z.-G.; Yau, L.-F.; Huang, H.; Liu, J.; Jiang, M.; Meng, Q.; Tong, T.-T.; Huang, H.-H.; Lee, S.; Zeng, X.; Liu, L.;

Jiang, Z.-H. A Method to Identify Trace Sulfated IgG N-Glycans as Biomarkers for Rheumatoid Arthritis. *Nat. Commun.* **2017**, *8* (1), 631.

(44) She, Y.-M.; Li, X.; Cyr, T. D. Remarkable Structural Diversity of N-Glycan Sulfation on Influenza Vaccines. *Anal. Chem.* **2019**, *91* (8), 5083–5090.

(45) Jiang, H.; Irungu, J.; Desaire, H. Enhanced Detection of Sulfated Glycosylation Sites in Glycoproteins. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (3), 340–348.

(46) Irungu, J.; Dalpathado, D. S.; Go, E. P.; Jiang, H.; Ha, H.-V.; Bousfield, G. R.; Desaire, H. Method for Characterizing Sulfated Glycoproteins in a Glycosylation Site-Specific Fashion, Using Ion Pairing and Tandem Mass Spectrometry. *Anal. Chem.* **2006**, *78* (4), 1181–1190.

(47) Toyoda, M.; Narimatsu, H.; Kameyama, A. Enrichment Method of Sulfated Glycopeptides by a Sulfate Emerging and Ion Exchange Chromatography. *Anal. Chem.* **2009**, *81* (15), 6140–6147.

(48) Toyoda, M.; Kaji, H.; Sawaki, H.; Togayachi, A.; Angata, T.; Narimatsu, H.; Kameyama, A. Identification and Characterization of Sulfated Glycoproteins from Small Cell Lung Carcinoma Cells Assisted by Management of Molecular Charges. *Glycoconjugate J.* **2016**, *33* (6), 917–926.

(49) Kim, J.; Lee, B.; Lee, J.; Ji, M.; Park, C. S.; Lee, J.; Kang, M.; Kim, J.; Jin, M.; Kim, H. H. N-Glycan Modifications with Negative Charge in a Natural Polymer Mucin from Bovine Submaxillary Glands, and Their Structural Role. *Polymers* **2020**, *13* (1), 103.

(50) Sleat, D. E.; Wang, Y.; Sohar, I.; Lackland, H.; Li, Y.; Li, H.; Zheng, H.; Lobel, P. Identification and Validation of Mannose 6-Phosphate Glycoproteins in Human Plasma Reveal a Wide Range of Lysosomal and Non-Lysosomal Proteins* S. *Mol. Cell Proteomics* **2006**, *5* (10), 1942–1956.

(51) Huang, J.; Dong, J.; Shi, X.; Chen, Z.; Cui, Y.; Liu, X.; Ye, M.; Li, L. Dual-Functional Titanium(IV) Immobilized Metal Affinity Chromatography Approach for Enabling Large-Scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. *Anal. Chem.* **2019**, *91* (18), 11589–11597.

(52) Yamada, K.; Kayahara, H.; Kinoshita, M.; Suzuki, S. Simultaneous Analysis of Sulfated and Phosphorylated Glycans by Serotonin-Immobilized Column Enrichment and Hydrophilic Interaction Chromatography. *Anal. Chem.* **2018**, *90* (14), 8387–8395.

(53) Yamada, K.; Suzuki, K.; Hirohata, Y.; Kinoshita, M. Analysis of Minor Acidic N-Glycans in Human Serum. *J. Proteome Res.* **2020**, *19* (8), 3033–3043.

(54) Čaval, T.; Zhu, J.; Tian, W.; Remmelzwaal, S.; Yang, Z.; Clausen, H.; Heck, A. J. R. Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-Phosphate Modification* [S]. *Mol. Cell Proteomics* **2019**, *18* (1), 16–27.

(55) Du, H.; Pang, M.; Hou, X.; Yuan, S.; Sun, L. PLOD2 in Cancer Research. *Biomed. Pharmacother.* **2017**, *90*, 670–676.

Figures

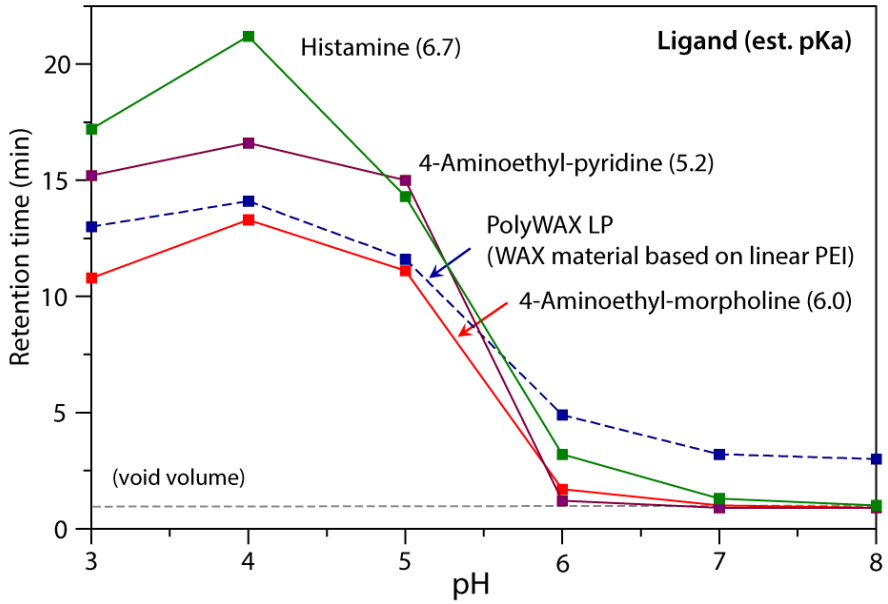


Figure 1. Titration curves for the three VWAX materials and a conventional PolyWAX LP material, determined by measuring the retention time of salicylic acid under varying pH conditions.

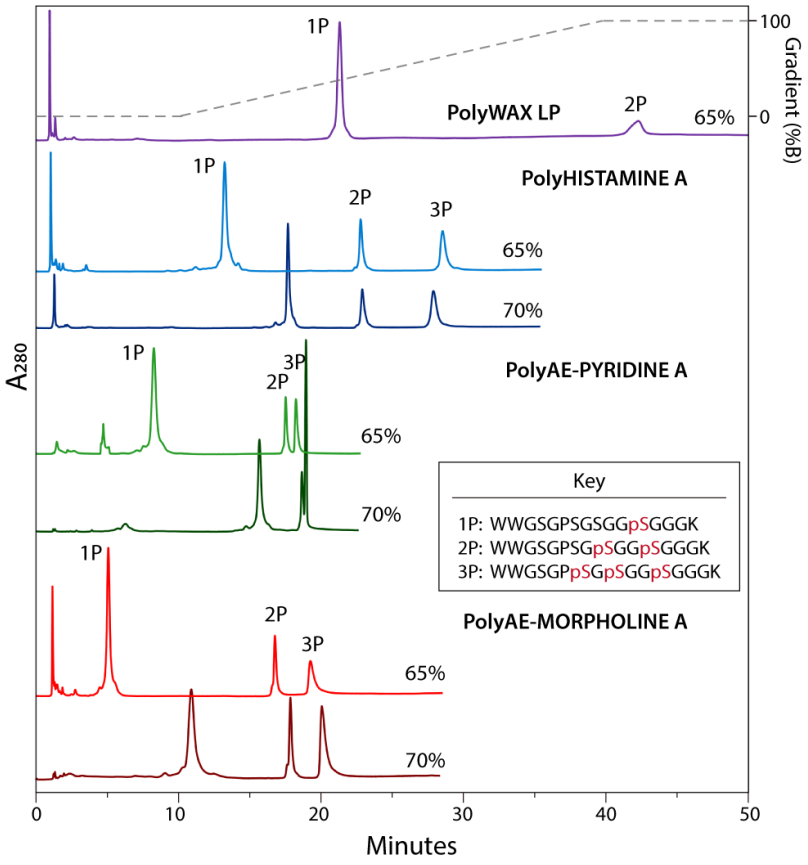


Figure 2. Comparison of the separation of peptides with 1, 2 or 3 phosphate groups on a regular PolyWAX LP column versus the three VWAX columns, using a salt and pH gradient in either 65% ACN or 70% ACN.

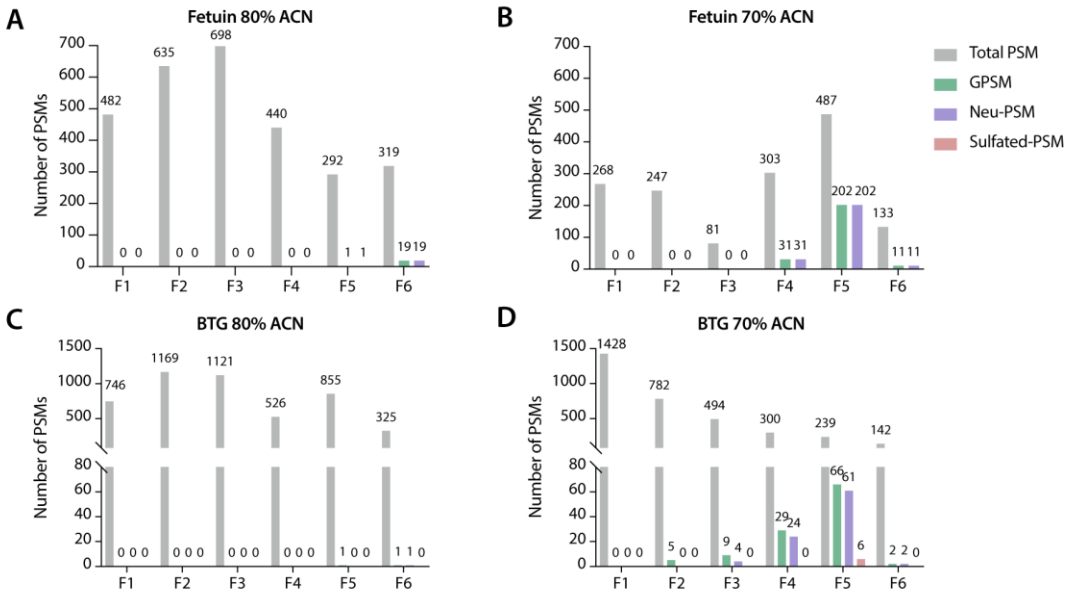


Figure 3. Number of PSMs identified in each fraction from bovine fetuin in (A) 80% ACN and (B) 70% ACN; and from BTG in (C) 80% ACN and (D) 70% ACN.

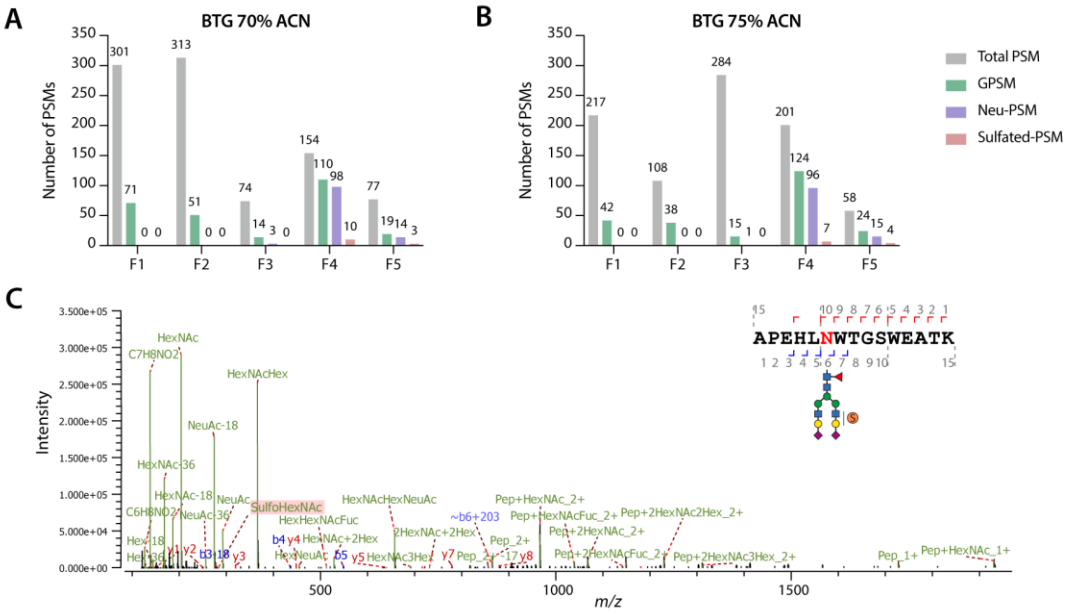


Figure 4. Number of PSMs identified in each fraction from BTG in (A) 70% ACN and (B) 75% ACN. (C) Representative PSM of a sulfated glycopeptide from BTG, as identified by Byonic. The tentative glycan structure is illustrated based on its glycan composition. Sulfate-containing diagnostic oxonium ion is highlighted in red.

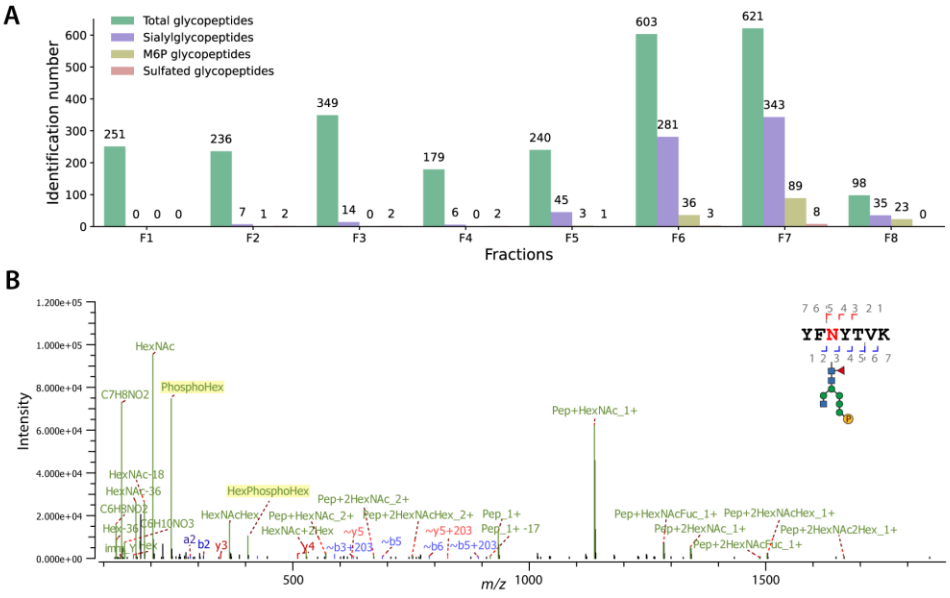


Figure 5. (A) Number of glycopeptides identified in each fraction from HeLa cell digests enriched by VWAX in 75% ACN. (B) Representative PSM of an M6P glycopeptide featuring a hybrid glycan, as identified by Byonic. The tentative glycan structure is illustrated based on its glycan composition. Phospho-containing diagnostic oxonium ion is highlighted in yellow.

Supplemental Information

Experimental Section.

LC-MS/MS analysis. Bovine fetuin and BTG samples were analyzed on an Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled with a Waters nanoAcquity UPLC (Milford, MA). Binary buffers (A, 0.1% FA; B, 100% ACN, 0.1%FA) were used in LC and the linear gradient was from 3 to 75% B for 60 min. MS survey scans of peptides were acquired from 300 to 2000 *m/z* at a resolution of 70K, using an AGC target setting of 1E6 and a maximum injection time of 50 ms. For MS2 scan, the top 15 precursor ions were selected for fragmentation by stepped higher-energy collision dissociation (HCD) with normalized collision energy (NCE) of 25 ± 5%. The MS2 data acquisition was performed at a resolution of 70K, an isolation width of 0.9 Da, a lower mass limit of *m/z* 110, an AGC target of 2E5, and an IT of 100 ms.

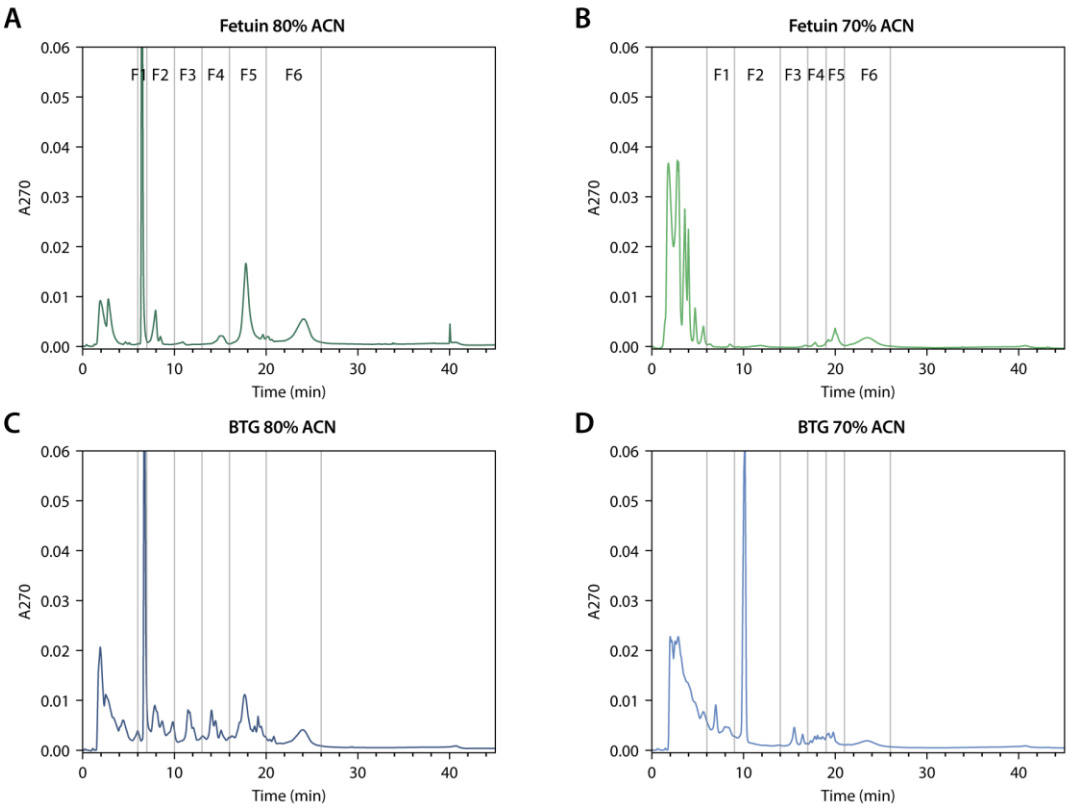


Figure S1. Chromatograms of 50 µg of glycoprotein standard digests separated by VWAX: (A) Bovine fetuin using MP A containing 80% ACN; (B) Bovine fetuin using MP A containing 70% ACN; (C) BTG separated by VWAX using MP A containing 80% ACN; and (D) BTG separated by VWAX using MP A containing 70% ACN.

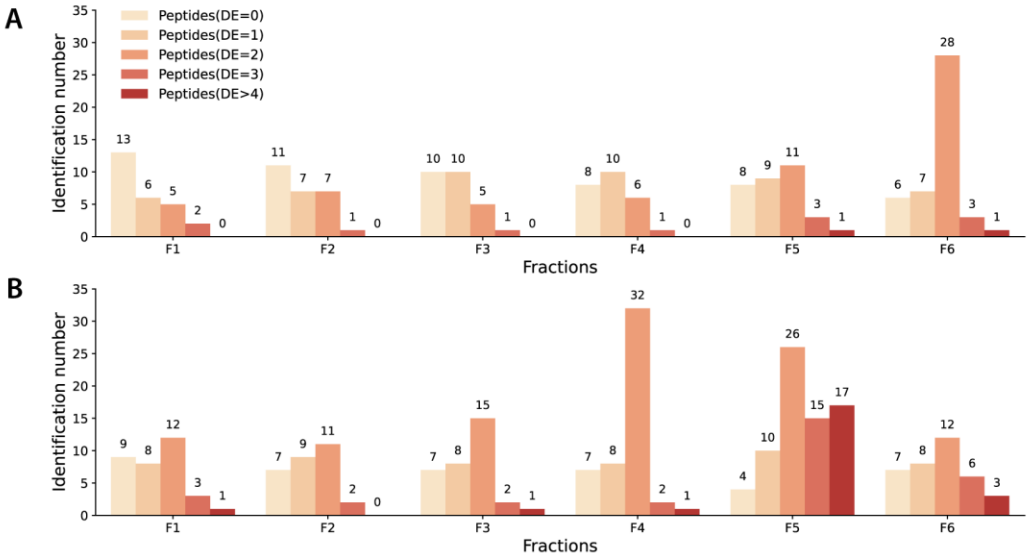


Figure S2. Number of peptides with varying number of Asp and Glu amino acid residues identified in each fraction of bovine fetuin enriched in (A) 80% ACN and (B) 70% ACN.

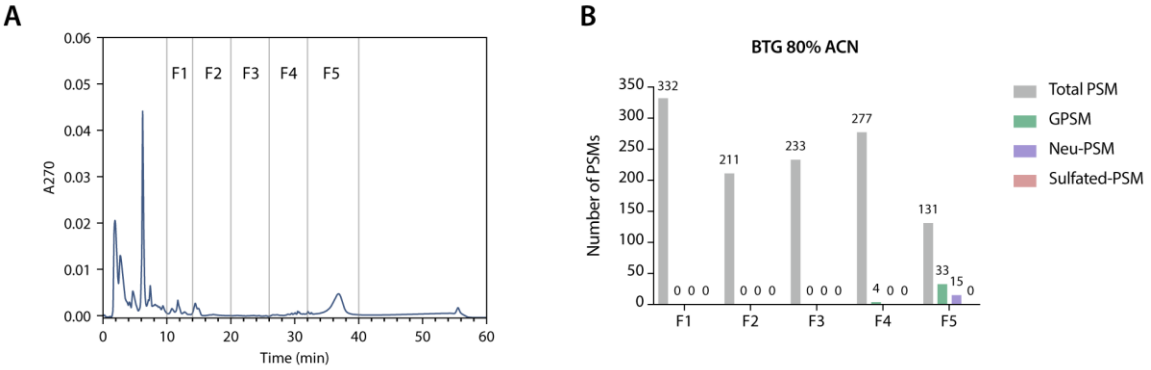


Figure S3. Chromatograms of 50 µg of BTG digests separated by VWAX using MP A containing 80% ACN. (B) Number of PSMs identified in each fraction.

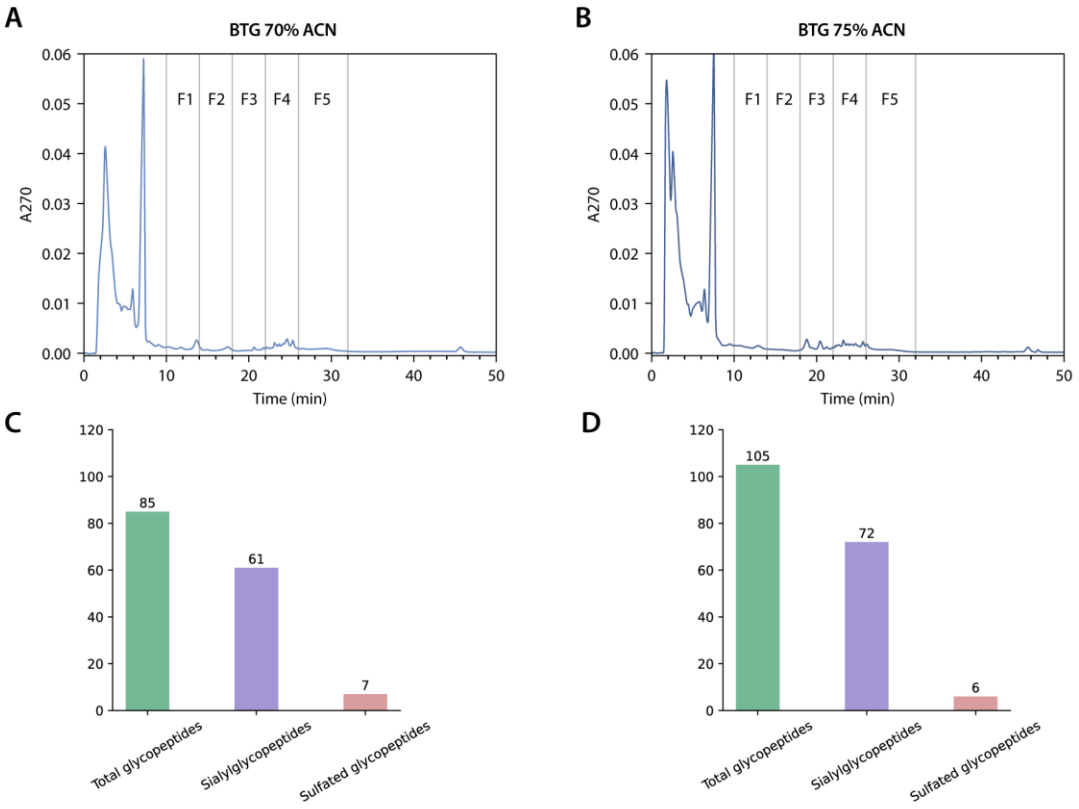


Figure S4. Chromatograms of 50 µg of BTG digests separated by VWAX using MP A containing (A) 70% ACN and (B) 75% ACN. The total number of identified glycopeptides among all fractions enriched by (C) 70% ACN and (D) 75% ACN.

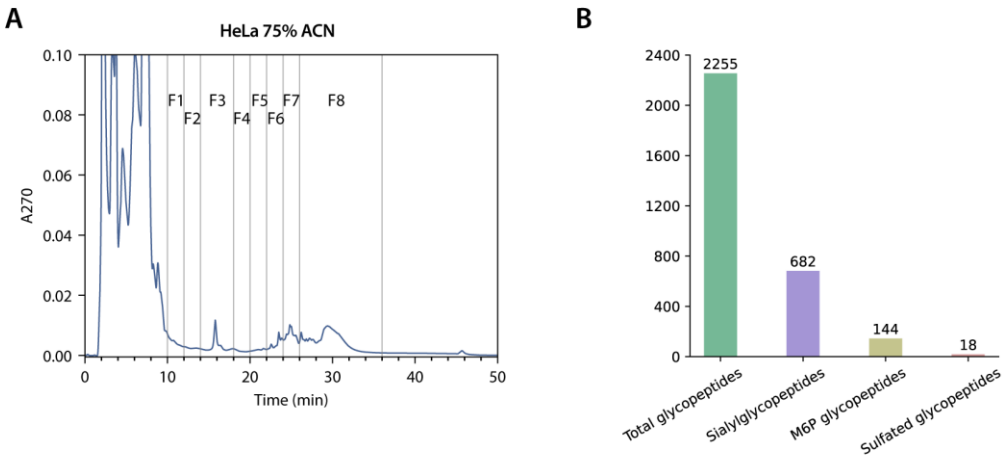


Figure S5. (A) Chromatogram of 200 µg of HeLa cell digests separated by VWAX using MP A containing 75% ACN. (B) The total number of identified glycopeptides among all fractions.

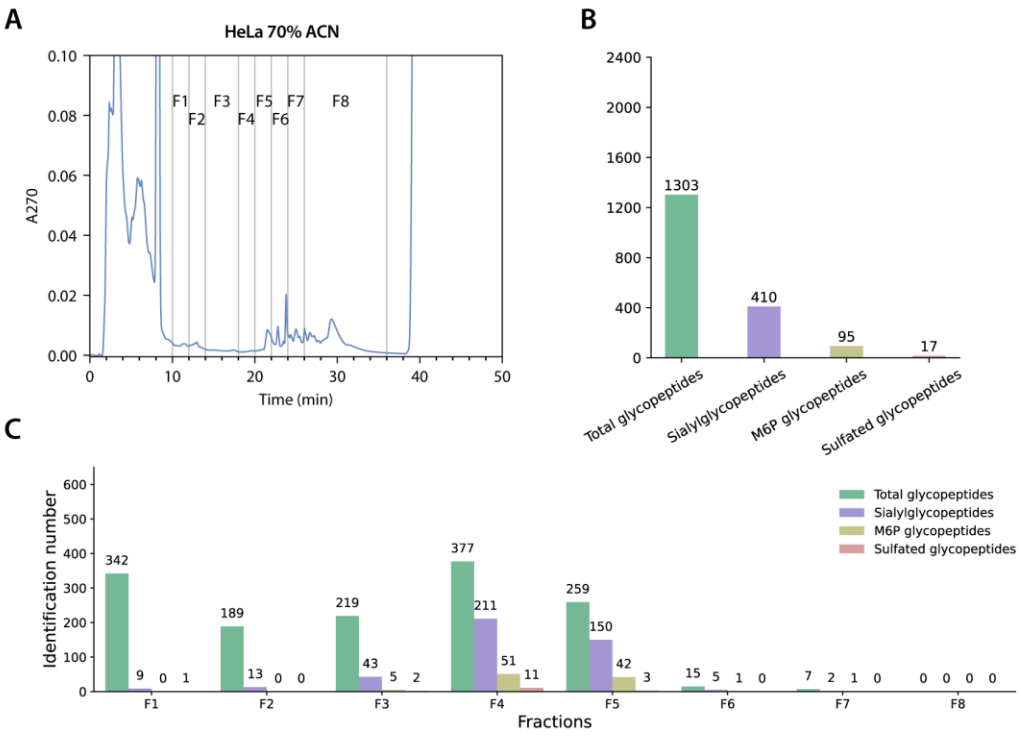


Figure S6. (A) Chromatogram of 200 µg of HeLa cell digests separated by VWAX using MP A containing 70% ACN. (B) The total number of identified glycopeptides among all fractions. (C) Number of glycopeptides identified in each fraction.

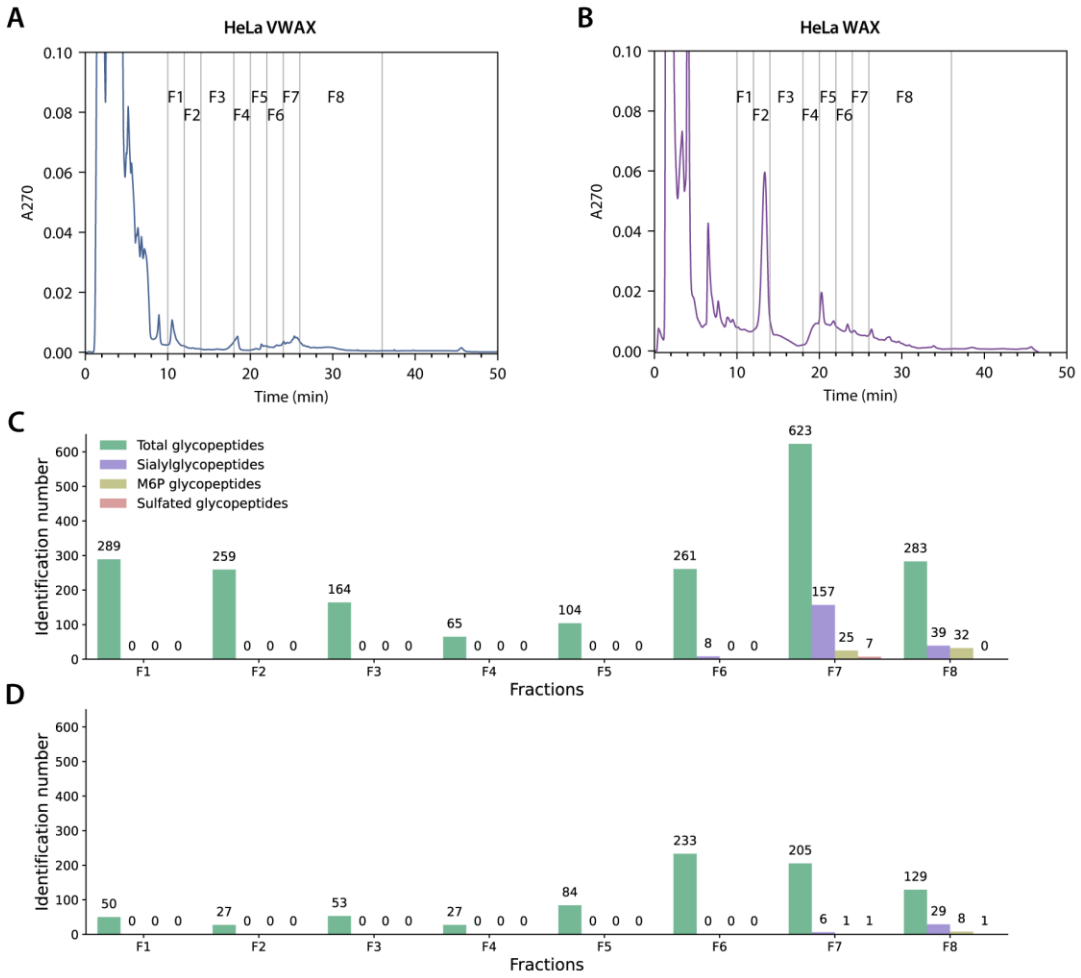


Figure S7. Chromatograms of 200 µg of HeLa cell digests separated by (A) VWAX and (B) WAX using MP A containing 75% ACN. Number of glycopeptides identified in each fraction enriched by (C) VWAX and (D) WAX.

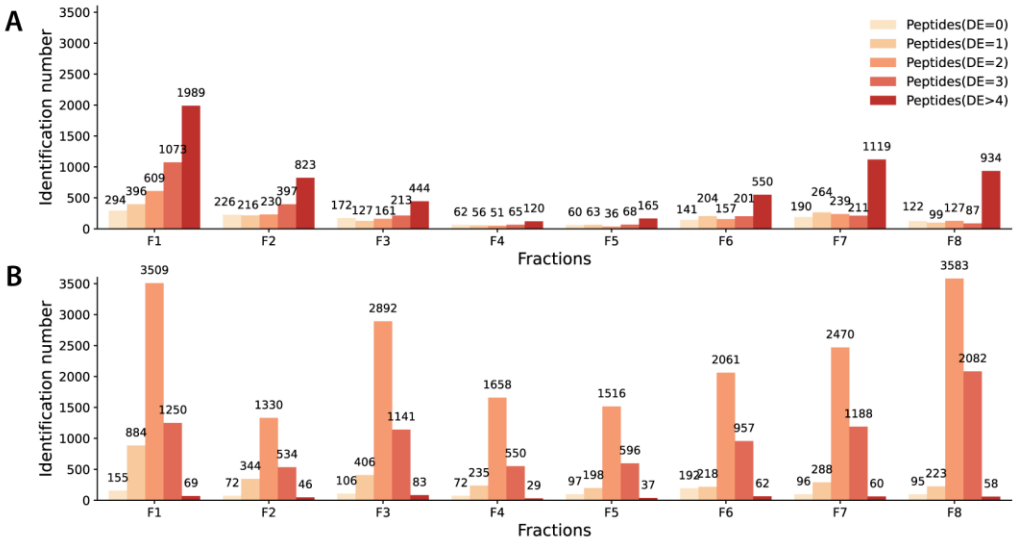
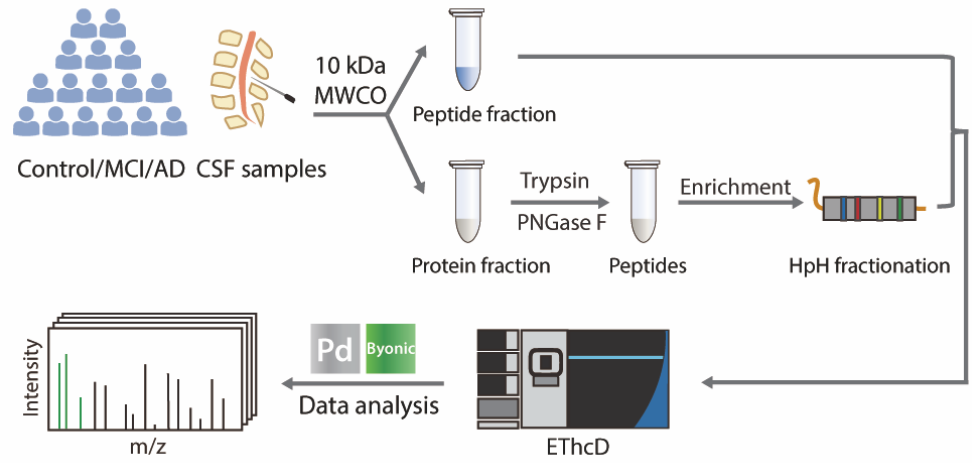


Figure S8. Number of peptides with varying number of Asp and Glu amino acid residues identified in each fraction of HeLa cell digests enriched by (A) VWAX and (B) WAX.

Chapter 5

In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients



Adapted from: Chen, Z.[#]; **Wang, D.[#];** Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Yu, Q.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients. *ACS Chem. Biol.* **2022**, *17* (11), 3059–3068. ([#]co-first authors) Author contribution: study was designed by D.W and Z.C. under the supervision of L.L.; experiment was performed by Z.C.; data was analyzed by D.W and Z.C.; samples were provided by H.Z., A.S.; C.C. and O.O.; manuscript was written by D.W. and Z.C., and edited by Q.Y., J.J., R.S., X.Z., J.H., Q.Y., and L.L.

Abstract

Site-specific O-glycoproteome mapping in complex biological systems provides a 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, while missing information on other glycosylation types. In order to achieve an unbiased O-glycosylation profiling, we developed an integrated strategy combining universal boronic acid enrichment, high-pH fractionation, and electron-transfer and higher-energy collision dissociation (EThcD) for enhanced intact O-glycopeptide analysis. We applied this strategy to analyze the O-glycoproteome in CSF, resulting in the identification of 308 O-glycopeptides from 110 O-glycoproteins, covering both sialylated and nonsialylated glycoforms. To our knowledge, this is the largest data set of O-glycoproteins and O-glycosites reported for CSF to date. We also developed a peptidomics workflow that utilized the EThcD and a three-step database searching strategy for comprehensive PTM analysis of endogenous peptides including N-glycosylation, O-glycosylation and other common peptide PTMs. Interestingly, among the 1411 endogenous peptides identified, 89 were O-glycosylated and only one N-glycosylated peptide was found, indicating that CSF endogenous peptides were predominantly O-glycosylated. Analyses of the O-glycoproteome and endogenous peptidome PTM were also conducted in the CSF of MCI and AD patients to provide a landscape of glycosylation patterns in different disease states. Our results showed a decreasing trend in fucosylation and increasing trend of endogenous peptide O-glycosylation, which may play an important role in AD progression.

Introduction

Human cerebrospinal fluid (CSF) is predominantly produced in choroid plexuses, circulates within the ventricles of the brain, and surrounds the brain and spinal cord.[1] The functional role of CSF includes mechanical protection of the central nervous system (CNS), homeostasis of the interstitial fluid in the brain, and regulation of neuronal functioning.[2,3] Through direct contact with CNS, the CSF reflects the ongoing physiological or pathological state of CNS.[4,5] Metabolites, peptides, proteins, enzymes, and hormones in CSF are involved in many biological processes, and changes in these compositions are viewed as a sign of pathological alterations in CNS. These biological compositional changes provide an opportunity to mine the CSF for biomarker discovery in neurological diseases.[6]

As one of the most complicated protein post-translational modifications (PTMs), glycosylation serves as a key regulatory mechanism controlling protein folding, molecular trafficking, cell adhesion, receptor activation and signal transduction.[7–10] Based on the amino acids that glycans attach to, glycosylation can be classified into two major categories: N-glycosylation and O-glycosylation. Biosynthesis of N-glycosylation is initiated by transferring a preassembled 14 monosaccharide complex glycan to asparagine residue (Asn) within the consensus motif (Asn-X-Ser/Thr, X≠P) followed by sequential addition or removal of certain monosaccharides in a well-defined process.[11,12] In contrast, O-glycosylation synthesis involves the attachment of a single monosaccharide to the serine/threonine (Ser/Thr) residue of a polypeptide without any definable peptide consensus motif and subsequent attachment of numerous diverse monosaccharide residues.

As a result, a higher degree of site occupancy, structural heterogeneity, and diversity have been observed in O-glycosylation.

O-glycosylation is further categorized into mucin and nonmucin types, according to the monosaccharide residue directly linked to polypeptides. The attached monosaccharide residue is N-acetylgalactosamine (GalNAc/HexNAc) in the mucin type, whereas the attached residue can be N-acetylglucosamine (GlcNAc/HexNAc), fucose (Fuc), galactose (Gal/Hex), mannose, and glucose in the nonmucin type.[11] Mucin type O-glycosylation on the cell surface and on secreted proteins has also been shown to modulate recognition, adhesion, communication between cells, and the cell's surrounding environments.[13] As a nutrient- and stress-responsive modification, nonmucin 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 proteins, O-glycosylation can also happen on the endogenous peptides, including neurotransmitters and hormones. In fact, it has been well recognized that after initial peptidase cleavages, endogenous peptides can undergo further post-translational modifications such as amidation, acetylation, phosphorylation, sulfation, and N-/O-glycosylation. As an example, it has been reported that there was an extensive N-/O-glycosylation of gonadotropin, which is a glycoprotein polypeptide hormone.[15] Recently, our group reported O-glycosylation on mouse insulin-1B and -2B chains and the human insulin-B chain, as well as multiple O-glycoforms of signaling peptides.[16] We also discovered 14 O-glycosylated neuropeptides in the crustacean nervous system.[17] These findings highlight the crucial role of O-

glycosylation on neurotransmitters and peptide hormones, and the necessity to characterize the glycosylation state of endogenous peptides.

Though the CSF proteome and endogenous peptidome have been extensively characterized, there are only a few reports on CSF O-glycosylation. One study reported 39 O-glycopeptides from 22 CSF O-glycoproteins,[18] and another study identified 43 O-glycopeptides from 28 CSF O-glycoproteins.[19] By utilizing an enrichment approach based on the oxidation of sialic acid and hydrazide chemistry, both studies lost the information on sialic acid and characterized only a subset of O-glycoproteome. As a result, the current O-glycoproteome depth of human CSF is rather limited, hindering the design of studies to explore more disease-related O-glycosylation alterations. On the other hand, CSF is a valuable source for studying neurodegenerative diseases, with previous studies exploring O-glycosylation changes in CSF samples from Alzheimer's disease (AD) patients. However, they are generally targeted at some well-known glycoproteins such as amyloid precursor protein (APP) or apolipoprotein E (APOE).[20,21] A comprehensive study at a system-wide scale used to explore the O-glycosylation changes in the AD state is still needed. For the endogenous peptidome study in CSF, one study identified 730 endogenous peptides, including 138 peptides with PTMs such as acetylation, amidation, phosphorylation, and Gln to pyro-Glu conversion.[22] However, none of the identified peptides were reported to be glycosylated. In another study with 563 endogenous peptides identified, the presence of glycan oxonium ions was observed in some spectra, and 28 O-glycopeptides were eventually identified with lower collision energy during fragmentation.[23] Nonetheless, only two O-glycan compositions were found. It is highly possible that there are many more O-glycosylated endogenous peptides undiscovered. Therefore,

an approach with a glycosylation-centered analysis workflow needs to be developed to systematically evaluate N-/O- glycosylation on endogenous peptides.

In the present study, we optimized the boronic acid-based enrichment strategy to efficiently enrich both sialylated and nonsialylated O-glycopeptides. High-pH (HpH) fractionation was performed to further improve O-glycoproteome coverage. With EThcD fragmentation, site-specific information was obtained for intact O-glycopeptide characterization. The optimized approach was then applied to O-glycoproteome analysis in CSF from healthy individuals, mild cognitive impairment (MCI) and AD patients. For endogenous peptide analysis, we developed a peptidomics workflow that combined CSF endogenous peptide extraction by 10 kDa molecular weight cutoff (MWCO), EThcD fragmentation, and a three-step database searching strategy for comprehensive PTM analysis. This workflow was further adopted to study the endogenous CSF peptides in MCI and AD patients. Overall, we believe that the developed analytical strategies in this study are readily applicable for site-specific O-glycosylation analysis of both glycoproteins and endogenous peptides in other complex biological systems.

Results and Discussion

Optimization of Boronic Acid Enrichment Strategy for O-Glycoproteomic Analysis. To avoid the interference of highly abundant nonglycosylated peptides, enrichment is a key step to enhance glycopeptide detection. Common enrichment techniques include lectin affinity enrichment chromatography and hydrophilic interaction liquid chromatography (HILIC).[24] However, each enrichment method possesses inherent limitations. Lectin only enriches specific types of glycans, while HILIC-based methods usually suffer from interference from many hydrophilic non-

glycopeptides and is also complicated by the difference in hydrophilicity between O-glycopeptides and N-glycopeptides.[25] Instead, boronic acid enrichment is a more universal method for enriching both N-glycopeptides and O-glycopeptides because the *cis*-diol groups on glycans can react with boronic acids to form a reversible covalent bond. This can later be released under acidic conditions without any side effects.[26,27] Despite its application to a large-scale N-glycoproteome study[28] and a small-scale analysis of O-GlcNAcylation,[29] boronic acid enrichment has not yet been used for comprehensive O-glycoproteome analysis of complex biological samples.

Here, we used a phenylboronic acid (PBA) solid phase extraction cartridge to extract the O-glycopeptides from complex tryptic digests. PNGase F was first used to remove the N-glycans to avoid interference from N-glycopeptides during enrichment and MS detection of O-glycopeptides. Starting from 200 μg of tryptic peptides from PANC1 cells, a total of 213 intact O-glycopeptides were identified right after enrichment in three technical replicates, and 100 O-glycopeptides could be identified in a single replicate on average (**Figure 1a and 1c**). To minimize the interference from coenriched nonglycosylated peptides, we performed off-line HpH fractionation, which has shown high separation orthogonality to low-pH reversed-phase liquid chromatography coupled with MS analysis.[30] A total of 229 intact O-glycopeptides were identified from seven HpH fractions in one replicate, representing nearly 2 times the number of identifications in the direct analysis (**Figure S1a, Table S2)**. A closer look at the distribution of the number of O-glycopeptides among seven fractions showed that O-glycopeptides were mainly in the first four fractions **(Figure S1b)**. This could be explained by the hydrophilicity of O-glycopeptides that shifts the elution to an earlier time frame on a C18 column. To better utilize instrument time and

run more replicates, the first fraction was combined with the last three fractions in another round of experiments. It is well-known that two or more technical replicates are needed to get the maximum coverage of peptides due to the randomness and stochastic sampling of data-dependent analysis (DDA) mode, which is more significant in glycoproteomic analyses.[31–33] Therefore, we evaluated the glycoproteome coverage within the context of multiple replicates. Among all O-glycopeptides identified with four repetitive injections, two replicates only yielded less than 60% cumulative coverage, and at least three technical replicates were needed to reach a coverage of more than 90% of the total identified glycopeptides (**Figure 1b**). As shown in **Figure 1c**, prefractionation improved the glycoproteome coverage by 4 times compared to direct analysis with the same number of replicates. When four replicates were analyzed, the total identification number of O-glycopeptide could further reach up to 987, representing nearly a 5-fold increase.

Site-Specific O-Glycoproteome Analysis in CSF. Benefiting from the optimized O-glycopeptide enrichment strategy and site-specific information provided by EThcD fragmentation (as discussed in the Supplemental Information), we identified 308 intact O-glycopeptides and 292 unique O-glycoforms from 181 O-glycosites and 110 O-glycoproteins in CSF from healthy individuals, which was a large increase compared to previous reports (**Figure 2a**). Overall, the majority of O-glycoproteins (72%) carried only a single O-glycosite, 25% carried two or three O-glycosites and a low percentage (5%) of glycoproteins were found to have more than four glycosites (**Figure 2b**). There were around 30% O-glycosites identified with more than two O-glycans, demonstrating the microheterogeneity of O-glycosylation (**Figure 2c**).

Among different kinds of O-glycoforms, mucin-type core 1 (Galβ1→3GalNAc) is known to be the major component of O-glycans.[34] In the present study, around 48.3% of O-glycoforms had core 1 glycoforms with a large portion of them sialylated (72.9%). This agreed well with the core 1 O-glycoform percentage in human serum (46.4%) as well as the dominant composition of sialylated core 1 glycoforms (~67%), because more than 80% of CSF proteins originated from the plasma filtrate and contained a similar glycosylation pattern.[35,36] Interestingly, as another capping unit to elongate glycan branches, only 5.1% of core 1 O-glycoforms were fucosylated and total fucosylated O-glycoforms only accounted for 29.1%, which was much lower than sialylated forms. As shown in **Figure 2d**, the number of O-glycoforms with two sialic acids far exceeded the number of O-glycoforms with two fucoses, while the number of O-glycoforms with one or three sialic acids/fucoses were similar.

Though there is no consensus motif for O-glycosylation, the role of adjacent proline residues is extensively studied, as it occurs frequently near the glycosylation site and may facilitate the process of protein glycosylation.[37–39] Here, we performed a proline frequency analysis of the ±10 residues surrounding the 181 identified O-glycosites. Though such proline frequency analysis results may differ depending on the origins of the O-glycosites data set and different tissues used, we did observe that the proline residue exhibited conserved sequence at the –1 and +3 positions in CSF (**Figure 3a, c**), which was consistent with previous global statistical analysis[39] and reports in CSF.[19] We also conducted the proline frequency analysis of the experimentally verified 435 O-glycosites from Uniprot (2017_12). The results revealed that the highest frequency of proline was from –4 to +7 positions, except for the +1 position right next to threonine (**Figure 3b, d**). The

conserved proline frequency was most pronounced at –1 and +3 positions in accordance with our findings. This unique conserved sequence information may represent CSF-specific features that resulted from various factors including the origins of the glycoproteins, different biosynthesis routes of these glycoproteins, glycotransferase enzyme activities, etc. Another CSF-specific feature was the higher frequency of glutamic acid (E) adjacent to the O-glycosite, compared with the results from the global O-glycosite database, especially at +1 and +2 positions. A previous study showed that O-glycosylation was markedly reduced with glutamic acid residue substituted at positions –1 and +3, while glutamic acid replacement at +1 and +2 had no such effect.[40] This may explain a higher frequency of glutamic acid at +1 and +2 positions, compared with other positions.

O-Glycosylation Alteration in MCI and AD. In addition to analyzing CSF from healthy individuals, we also applied this strategy to systematically analyze CSF samples from MCI and AD patients to obtain global profiles of O-glycosylation in these disease states. In total, 366 O-glycopeptides mapping to 197 O-glycosites and 128 O-glycoproteins were found in MCI CSF and 358 O-glycopeptides mapping to 214 O-glycosites and 136 O-glycoproteins were found in AD CSF **(Table S3)**. The Venn diagrams demonstrate the diversity and overlap of O-glycosylation at these three stages (**Figure 4a**). 22.8%, 20.9% and 22.1% of unique O-glycopeptides are presented in each group, suggesting the necessity to obtain a global profile of the O-glycosylation landscape before performing further quantitative O-glycoproteomic study. The overlap of O-glycopeptides, O-glycosites and O-glycoprotein between MCI and AD were 27.3%, 30.2% and 39.0%, respectively, which was higher than the overlap between healthy vs. MCI and healthy vs. AD.

These results indicated certain glycosylation pattern changes during disease progression. To highlight more specific features of O-glycosylation patterns at each stage, the percentages of the different disease states of glycoforms were compared (**Figure 4b, Table S3**). Since the core 1 O-glycoform is the most prevalent glycoform, we compared the core 1 percentage across the three states along with their sialylated and fucosylated glycoforms. Most of the percentages were quite comparable among three stages with a slight increase in core 1 and sialylated glycoforms. However, we observed a trend in decreased global fucosylation during disease progression. This was quite interesting because we also found reduced fucosylation patterns in the N-glycosylation study of AD CSF.[41] Future biological mechanistic investigations may be necessary to reveal the functional role of fucosylation in AD pathogenesis and progression.

Selected Examples of O-Glycoproteins. Among all identified O-glycoproteins, we selected some proteins of interest to discuss below. As structural components of lipoprotein particles, apolipoproteins play important roles in maintaining their structure and regulating their metabolism and enzyme activities. Studies have shown that apolipoproteins are often modified by O-glycosylation, including apolipoprotein E, A-I, A-II and C-III.[42–44] In our study, four apolipoproteins, apolipoproteins D, A-I, E, and J were found to be O-glycosylated **(Table S4)**. Although N-glycosylation of apolipoprotein D (APOD) has been extensively studied and there are many potential O-glycosylation sites (8 serine residues and 10 threonine residues) that exist in its protein sequence, O-glycosylation of APOD has not been reported before. In our data sets, APOD was found to be O-glycosylated at Thr86. A total of three O-glycoforms were detected at this O-glycosite, and all of them were fucosylated. In contrast, none of the glycoforms were found in MCI

and AD CSF (**Figure 5a**). Using a lectin-based isolation method, Cubedo et al. reported the O-glycosylation of serum apolipoprotein A-I, but the exact O-glycosite or O-glycoform information was lacking.[45] In our study, one glycoform was detected in healthy and MCI CSF at Thr92 and an additional glycosite was found in AD CSF at Ser 76 (**Figure 5b**). For APOE, two O-glycosites at Thr36 and Thr212 were found in all three states and one additional glycosite at Ser 215 was found in MCI CSF. Over half of the glycoforms were sialylated, and our results suggested a high degree of microheterogeneity of glycosylation at Thr 212 (**Figure 5c**). Despite APOJ having been shown to carry several PTMs such as N-glycosylation, ubiquitination and phosphorylation, O-glycosylation has not been reported before.[46–49] In our study, seven glycosites were found among three states and only Thr 105 and Ser 210 were glycosylated in all three states (**Figure 5d**). Compared to ApoE, glycosylation on ApoJ showed higher degree of macro-heterogeneity and the change in site occupancy may play an important role in AD progression.[50]

APP is another protein of interest in AD research, because its aberrant proteolytic processing into amyloid β (Aβ) in the brain is an essential step during the pathogenesis of AD and O-glycosylation is considered to be related to Aβ formation.[51] Here, we found that Thr 22 was glycosylated at all three states with fucosylated glycoforms, and MCI state had one additional glycosite at Thr 667 (**Figure 5e**), which was one of the previously reported sites.[52]

In addition, we observed several interesting cases of O-glycosylation occurring at Thr residues of the N-glycosylation Asn-X-Thr consensus motif. These proteins are discussed in detail in the Supplemental Information.

Strategy for Comprehensive PTM Analysis of CSF Endogenous Peptides. Glycosylation is not included as a possible dynamic modification in most peptidomic analyses, because it requires an additional glycan database to conduct the search, which requires much larger computational capacities. To overcome such limitation, we utilized EThcD to preserve the labile glycosylation and PTM-centric Byonic searching engine to enable site-specific glycoprofiling.[53] We first tried to search the raw data directly by setting all the possible PTMs (glycosylation, phosphorylation, acetylation etc.) and then using the whole human protein and glycan databases. However, this caused computational issues. A second trial with a "focused" CSF protein database constructed based on the literature also did not work for this data set, suggesting that the size of glycan database might be too large. Therefore, we developed a 3-step searching strategy (**Figure S3**) to facilitate the search; details can be found in the Experimental Section of the Supplemental Information. The key idea of this search method is that a focused protein and glycan database are constructed by the first two rounds of searches. These databases will then be used to perform the final round search, where all the possible PTMs can be searched all at once.

For endogenous peptide separation, our lab previously developed a 10 kDa MWCO-based protocol to achieve an optimal recovery rate.[54,55] Here, we compare this 10 kDa MWCO-based protocol with another 30 kDa MWCO-based protocol reported in literature in terms of their performance on CSF samples.[56] As shown in **Figure S4 (Table S5)**, 10 kDa-based protocol outperformed 30 kDa-based protocol in identification numbers of both total peptides and modified peptides. Therefore, we adapted the 10 kDa-based protocol to separate CSF samples into peptide fraction and protein fraction (**Figure S5**).

CSF Endogenous Peptidome Mapping in Control, MCI, and AD. With an optimized analysis strategy, a total of 1411 endogenous peptides were identified in CSF samples from healthy individuals (**Figure 6a**), which was an over 2-fold increase compared to previous reports.[22,23] Benefiting from EThcD fragmentation, labile PTMs including glycosylation and phosphorylation are well preserved (**Figure S2**). In total, 339 endogenous peptides with PTMs were identified, indicating that endogenous peptides in CSF went through extensive modifications. Among them, 89 and 34 peptides were O-glycosylated and phosphorylated, respectively. It is worth noting that this work is the first report to show extensive O-glycosylation on CSF endogenous peptides, and the number of O-glycosylated peptides exceeded the number of peptides with other PTMs. Overall, these O-glycopeptides were derivatized from 27 protein precursors, and 27 O-glycan compositions were identified. Interestingly, only one N-glycosylated peptide TNSTFVQALVEHVK from precursor protein prosaposin was found with N-glycosite at previously reported Asn215 residue.

As a subset of endogenous peptides, neuropeptides are of interest because of their involvement in various biological processes such as intercellular signaling molecules.[57] By referring to the human neuropeptide database NeuroPep,[58] 276 peptides were identified as neuropeptides. More than half of them (178, 64.5%) belonged to chromogranin/secretogranin family, followed by ProSAAS (28, 10.1%), NPY (18, 6.5%), VGF (17, 6.2%), opioid (11, 4.0%) and 7B2 (7, 2.5%) (**Figure S6**). A total of 88 were modified, including acetylation, amidation, O-glycosylation, phosphorylation, and Gln to pyro-Glu conversion. Among the 15 O-glycosylated neuropeptides, 14 originated from ProSAAS and one was from secretogranin-1. ProSAAS-derived neuropeptides, SAAS, PEN, and LEN, are among the most abundant peptides present in the mouse

hypothalamus,[59–61] which have been implicated in the regulation of food intake and body weight.[62,63] We detected two O-glycosylation sites (Thr53, Thr247) in the LEN and SAAS regions with 11 O-glycosylated LEN peptides and 3 O-glycosylated SAAS peptides (**Table S6**). At site Thr53, two O-glycoforms with sialylated T-antigen were identified, whereas at site Thr247 three O-glycoforms with T-antigen in both sialylated and nonsialylated forms and one nonsialylated Tn-antigen glycoforms were found.

There was a decrease in the number of endogenous peptides in MCI and AD compared with healthy individuals. A closer look into the peptide modifications showed that the decrease mainly came from the unmodified peptides (**Figure 7a, Table S7**). A more detailed analysis of PTM types revealed that there was a trend of increasing O-glycosylation, Gln→pyro-Glu, acetylation, phosphorylation, and a trend for slightly decreased oxidation (**Figure 7b**). These intriguing observations represent a global landscape of endogenous peptidome from control, MCI and AD, exhibiting distinct features at different states of AD progression. Further investigation is needed to explore the biological function, thier potential roles of these endogenous peptides in AD progression, and how each PTM regulates their bioactivities.

Conclusion

In summary, we developed an effective workflow to simultaneously analyze site-specific O-glycoproteome and endogenous peptides with PTMs in CSF. For the O-glycoproteome study, an optimized boronic acid enrichment method enhanced site-specific O-glycopeptide analysis. In total, 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins were identified from healthy individuals, representing the largest data set of site-specific O-glycoproteome study

in CSF to date, including 154 novel O-glycosites reported for the first time. We also profiled the O-glycoproteome in MCI and AD patients. For endogenous peptide study, we developed a peptidomics workflow that enabled comprehensive PTM analysis. From this analysis, a considerable number of CSF endogenous peptides were O-glycosylated in all three states, while only a few notable peptides were N-glycosylated. Although an in-depth glycoproteomic biomarker discovery study on individual control, MCI and AD subjects is ideal for biomarker discovery, the relatively small size of CSF samples, instrument/labor time, and financial constraints limited these experiments in this study. Instead, the samples in each group are pooled to reduce biological variation and allow deeper profiling of low-abundance glycoproteins.[64] Our findings shed light on the CSF O-glycoproteome landscape, dominant glycosylation differences, similarities during AD progression, and PTM focused peptidome mapping. Future investigations will be conducted using strategies developed here to explore potential O-glycosylated or endogenous proteolytic biomarker candidates in a more quantitative manner, such as using 12-plex DiLeu isobaric labeling to analyze individual subjects and pinpoint more specific changes of interesting glycoprotein(s)/glycosite(s)/glycoform(s).[65]

Experimental Section

Chemicals and Materials. See the Supplemental Information for details.

CSF Samples. See the Supplemental Information and **Table S1** for details. In short, 48 enrollees (16 cognitively normal individuals, 16 individuals with MCI and 16 individuals with AD dementia) in the Wisconsin Alzheimer's Disease Research Center (ADRC) participated in this study. The

University of Wisconsin Institutional Review Board approved all study procedures. CSF aliquots from 16 individuals of each group were pooled together for analysis.

Sample Processing and Protein Digestion. Details of PANC1 cell culture, protein extraction and digestion, and CSF sample processing were provided in the Supplemental Information. Briefly, the CSF sample was separated into a peptide fraction and a protein fraction using 10 kDa MWCO protocol. The peptide fraction was injected for LC-MS/MS analysis after desalting. The protein fraction was reduced, alkylated, and digested with trypsin. N-glycans were removed by PNGase F before O-glycopeptide enrichment.

Boronic acid enrichment. See the Supplemental Information for details.

LC-MS/MS Analysis. Details of LC-MS/MS data acquisition are provided in the Supplemental Information. Briefly, samples were analyzed on the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Dionex UPLC system. Data were acquired in data dependent acquisition (DDA) mode with EThcD fragmentation.

Data Analysis. All data were searched by Byonic (version 2.9.38, Protein Metrics Inc., San Carlos, CA) incorporated in Proteome Discoverer 2.1. Details of search parameters, databases, and postsearch processing were provided in the Supplemental Information. Data are available at Mass Spectrometry Interactive Virtual Environment (https://massive.ucsd.edu) with deposit ID of MSV000087160.

Acknowledgements

This research was supported in part by the National Institutes of Health (NIH) grants RF1 AG052324, R21AG065728, U01CA231081, R01 DK071801, and P30AG062715. 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. L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. We thank M.Bern from Protein Metrics for providing 1-year 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*, 369–383.

(2) 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*, 40–47.

(3) 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; Springer, **1998**; pp 59–65.

(4) Zhang, J. Proteomics of Human Cerebrospinal Fluid – the Good, the Bad, and the Ugly. *Proteom. - Clin. Appl* **2007**, *1* (8), 805–819.

(5) 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*, 39–64.

(6) Yuan, X.; Desiderio, D. M. Proteomics Analysis of Human Cerebrospinal Fluid. *J. Chromatogr. B* **2005**, *815*, 179–189.

(7) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126*, 855–867.

(8) 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*, 6527–6534.

(9) 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.

(10) Chen, Z.; Huang, J.; Li, L. Recent Advances in Mass Spectrometry (MS)-Based Glycoproteomics in Complex Biological Samples. *Trends Analyt Chem.* **2018**, *118* (Cell 126 2006), 880–892.

(11) Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology.*; Cold Spring Harber Laboratory Press, New York; **1998**.

(12) Chen, Z.; Glover, M. S.; Li, L. Recent Advances in Ion Mobility–Mass Spectrometry for Improved Structural Characterization of Glycans and Glycoconjugates. *Current opinion in chemical biology* **2018**, *42*, 1–8.

(13) Tran, D. T.; Hagen, K. G. T. Mucin-Type O-Glycosylation during Development. *J. Biol. Chem.* **2013**, *288*, 6921–6929.

(14) Yang, X.; Qian, K. Protein O-GlcNAcylation: Emerging Mechanisms and Functions. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (7), 452–465.

(15) Thotakura, N. R.; Blithe, D. L. Glycoprotein Hormones: Glycobiology of Gonadotrophins, Thyrotrophin and Free α Subunit. *Glycobiology* **1995**, *5*, 3–10.

(16) 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.

(17) Cao, Q.; Yu, Q.; Liu, Y.; Chen, Z.; Li, L. Signature-Ion-Triggered Mass Spectrometry Approach Enabled Discovery of N- and O-Linked Glycosylated Neuropeptides in the Crustacean Nervous System. *J. Proteome Res.* **2019**, *19* (2), 634–643.

(18) 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.

(19) Halim, A.; Rüetschi, U.; Larson, G.; 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.

(20) Flowers, S. A.; Grant, O. C.; Woods, R. J.; Rebeck, G. W. O-Glycosylation on Cerebrospinal Fluid and Plasma Apolipoprotein E Differs in the Lipid-Binding Domain. *Glycobiology* **2019**, *30* (2), 74–85.

(21) Nilsson, J.; Brinkmalm, G.; Ramadan, S.; Gilborne, L.; Noborn, F.; Blennow, K.; Wallin, A.; Svensson, J.; Abo-Riya, M. A.; Huang, X.; Larson, G. Synthetic Standard Aided Quantification and Structural Characterization of Amyloid-Beta Glycopeptides Enriched from Cerebrospinal Fluid of Alzheimer's Disease Patients. *Sci. Rep.* **2019**, *9* (1), 5522.

(22) Hölttä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J. Peptidome Analysis of Cerebrospinal Fluid by LC-MALDI MS. *PLoS One* **2012**, *7*, e42555.

(23) 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*, 386–399.

(24) Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y. Advances in Mass Spectrometry Driven O-Glycoproteomics. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850* (1), 33–42.

(25) Kong, S.; Zhang, Q.; Yang, L.; Huang, Y.; Liu, M.; Yan, G.; Zhao, H.; Wu, M.; Zhang, X.; Yang, P.; Cao, W. Effective Enrichment Strategy Using Boronic Acid-Functionalized Mesoporous Graphene-Silica Composites for Intact N- and O-Linked Glycopeptide Analysis in Human Serum. *Anal. Chem.* **2021**, *93*, 6682–6691.

(26) 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. *J. Biomol. Tech.* **2005**, *16*, 407.

(27) 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*, 503–508.

(28) 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*, 1563–1572.

(29) 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*, 2462–2475.

(30) Batth, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-PH Reversed-Phase Fractionation for in-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, *13*, 6176–6186.

(31) 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*, 4193–4201.

(32) Huang, J.; Dong, J.; Shi, X.; Chen, Z.; Cui, Y.; Liu, X.; Ye, M.; Li, L. Dual-Functional Titanium(IV) Immobilized Metal Affinity Chromatography Approach for Enabling Large-Scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. *Anal. Chem.* **2019**, *91* (18), 11589–11597.

(33) 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.

(34) 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*, 9953–9961.

(35) 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. *Anal. Chem.* **2017**, *89* (3), 1469–1476.

(36) 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*, O111. 016774.

(37) 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*, 10029–10038.

(38) 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*, 25010–25018.

(39) Wilson, I.; Gavel, Y.; Heijne, G. V. Amino Acid Distributions around O-Linked Glycosylation Sites. *Biochem. J.* **1991**, *275*, 529–534.

(40) 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*, 7061–7065.

(41) Chen, Z.; Yu, Q.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific Analysis of N-Glycoproteome in Human Cerebrospinal Fluid and Glycosylation Landscape Changes in Alzheimer's Disease. *Mol. Cell Proteomics* **2021**, *20*, 100081.

(42) 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*, 6785–6790.

(43) 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. *Arterioscler. Thromb. Vasc. Biol.* **1988**, *8*, 288–297.

(44) 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*, 1140–1144.

(45) Cubedo, J.; Padró, T.; Badimon, L. Glycoproteome of Human Apolipoprotein AI: N-and O-Glycosylated Forms Are Increased in Patients with Acute Myocardial Infarction. *Transl. Res.* **2014**, *164*, 209–222.

(46) Woody, S. K.; Zhao, L. Clusterin (APOJ) in Alzheimer's Disease: An Old Molecule with a New Role; InTech, **2016**.

(47) Burkey, B.; Harmony, J. Intracellular Processing of Apolipoprotein J Precursor to the Mature Heterodimer. *J. Lipid Res.* **1991**, *32*, 1039–1048.

(48) 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*, 5063–5076.

(49) 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*, 2120–2133.

(50) Zhang, Q.; Ma, C.; Chin, L.-S.; Li, L. Integrative Glycoproteomics Reveals Protein N-Glycosylation Aberrations and Glycoproteomic Network Alterations in Alzheimer's Disease. *Sci. Adv.* **2020**, *6* (40), eabc5802.

(51) Shi, J.; Ku, X.; Zou, X.; Hou, J.; Yan, W.; Zhang, Y. Comprehensive Analysis of O-Glycosylation of Amyloid Precursor Protein (APP) Using Targeted and Multi-Fragmentation MS Strategy. *Biochim. Biophys. Acta - Gen. Subj.* **2021**, *1865* (10), 129954.

(52) Halim, A.; Brinkmalm, G.; Rüetschi, U.; Westman-Brinkmalm, A.; Portelius, E.; Zetterberg, H.; Blennow, K.; Larson, G.; Nilsson, J. Site-Specific Characterization of Threonine, Serine, and Tyrosine Glycosylations of Amyloid Precursor Protein/Amyloid β-Peptides in Human Cerebrospinal Fluid. *Proc. Natl. Acad. Sci. U. S. A* **2011**, *108* (29), 11848–11853.

(53) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Curr Protoc Bioinform* **2012**, *40* (1), 13.20.1-13.20.14.

(54) 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*, 1327–1332.

(55) 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. *Proteom. - Clin. Appl* **2016**, *10* (12), 1225–1241.

(56) 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. *Nat. Commun.* **2016**, *7*.

(57) Sossin, W. S.; Fisher, J. M.; Scheller, R. H. Cellular and Molecular Biology of Neuropeptide Processing and Packaging. *Neuron* **1989**, *2*, 1407–1417.

(58) Wang, Y.; Wang, M.; Yin, S.; Jang, R.; Wang, J.; Xue, Z.; Xu, T. NeuroPep: A Comprehensive Resource of Neuropeptides. *Database* **2015**, *2015* (0), bav038.

(59) 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*, 639–648.

(60) Fricker, L. D. Neuropeptides and Other Bioactive Peptides: From Discovery to Function; Morgan & Claypool Life Sciences, **2012**; Vol. 1, pp 1–122.

(61) 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*, 6207–6213.

(62) 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. *Sci. Signal.* **2016**, *9*, ra43.

(63) 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. *Proc. Natl. Acad. Sci. U. S. A* **2013**, *110*, 16211–16216.

(64) Kendziorski, C.; Irizarry, R. A.; Chen, K.-S.; Haag, J. D.; Gould, M. N. On the Utility of Pooling Biological Samples in Microarray Experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (12), 4252–4257.

(65) Frost, D. C.; Greer, T.; Li, L. High-Resolution Enabled 12-Plex DiLeu Isobaric Tags for Quantitative Proteomics. *Anal. Chem.* **2015**, *87* (3), 1646–1654.

Figures

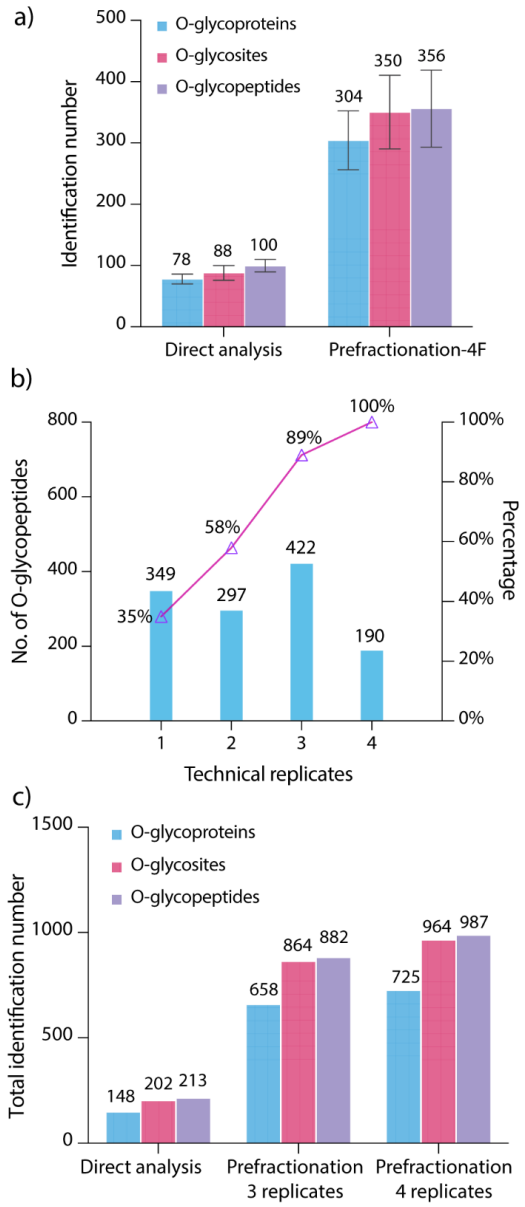


Figure 1. Boronic acid enrichment optimization. a) Comparison of O-glycoproteome coverage between direct analysis and prefractionation with four fractions. (F: fractions); b) The effects of technical replicates on the overall coverage; c) Comparison of O-glycoproteome coverage between direct analysis, prefractionation with three technical replicates and prefractionation with four technical replicates.

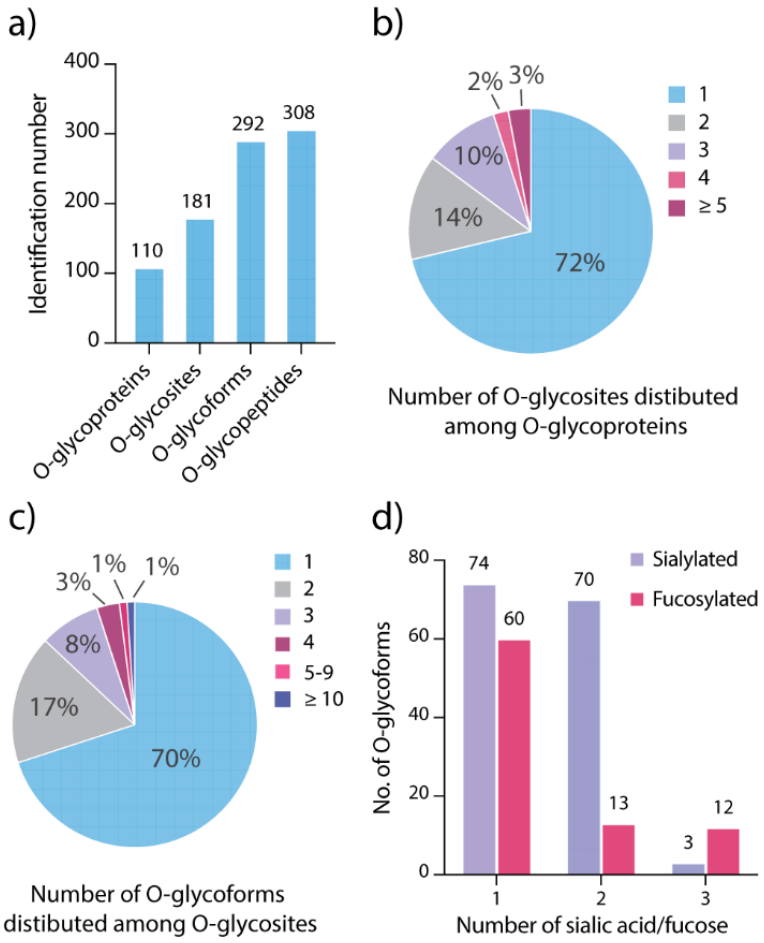


Figure 2. a) The O-glycoproteome mapping in CSF from healthy individuals. b) The distribution of the number of O-glycosites per O-glycoproteins. c) The distribution of the number of O-glycoforms per O-glycosite. d) Number of O-glycoforms with 1, 2 or 3 sialic acid or fucose residues.

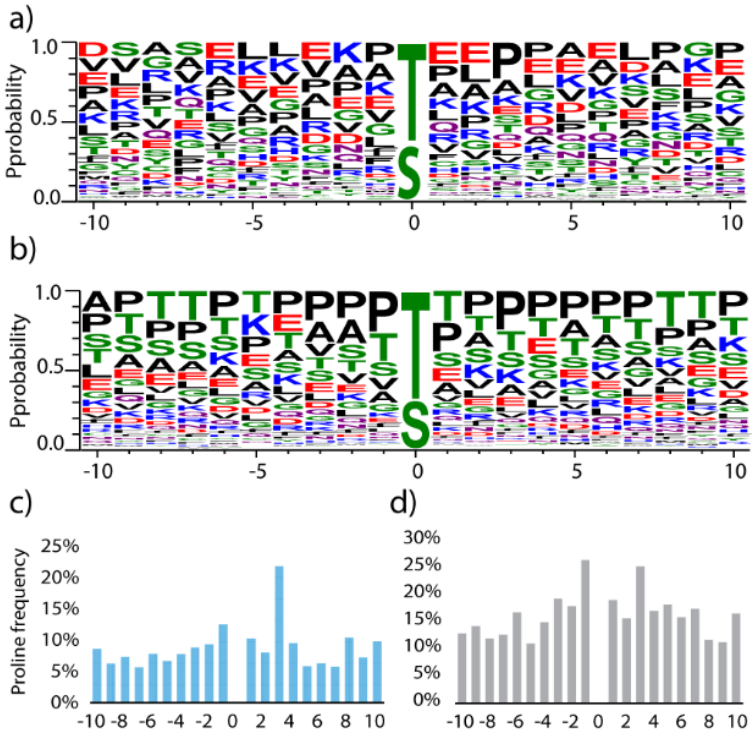


Figure 3. Weblogo probability plots and proline frequency analysis. a) Weblogo plot for 181 identified O-glycosites in the current study. b) Weblogo plot for 435 O-glycosites from Uniprot. c) Proline frequency analysis of 181 identified O-glycosites in the current study. d) Proline frequency analysis of 435 O-glycosites from Uniprot. ±10 amino acid residues surrounding the O-glycosite were analyzed.

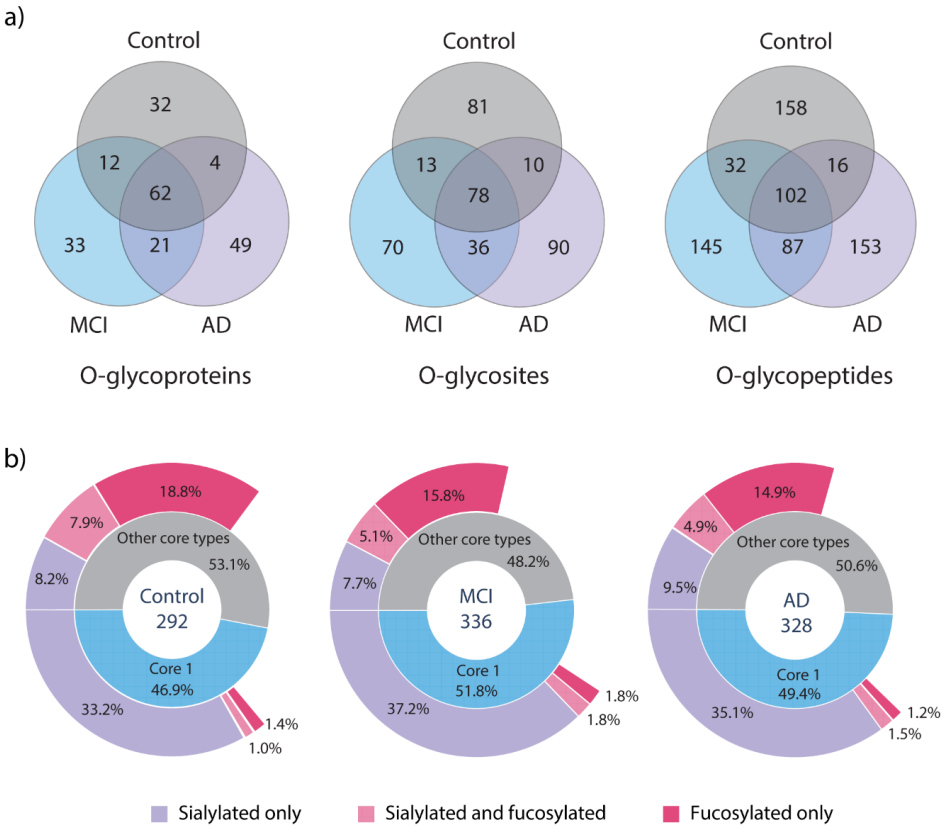


Figure 4. O-glycoproteome comparison between control, MCI, and AD. a) Venn diagram analysis of total O-glycoproteome in control, MCI, and AD. b) Comparison of distribution of different types of O-glycoforms in control, MCI, and AD.

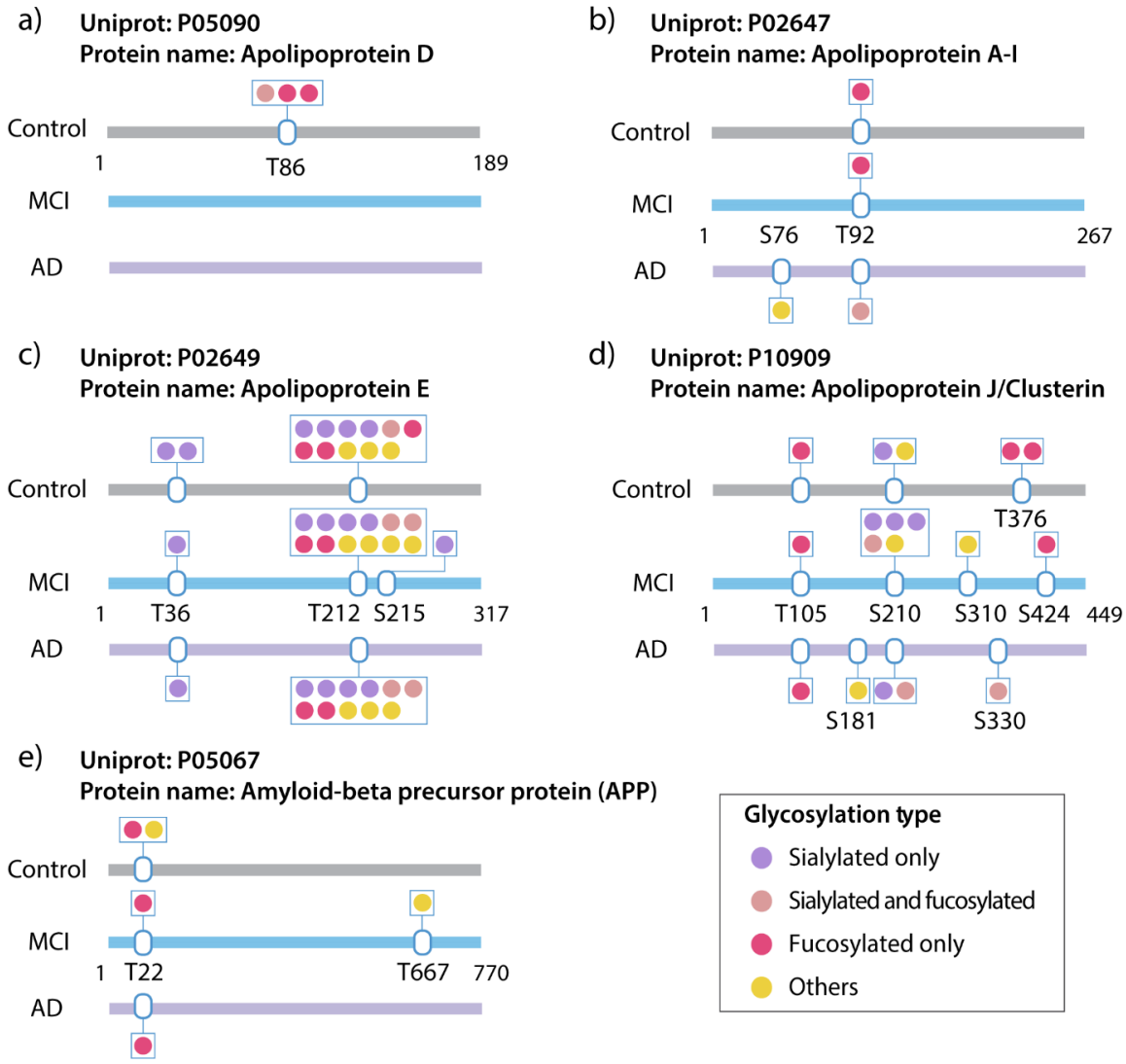


Figure 5. Site-specific O-glycosylation analysis for a) apolipoprotein D, b) apolipoprotein A-I, c) apolipoprotein E, d) apolipoprotein J (clusterin), and e) amyloid-beta precursor protein (APP) in control, MCI, and AD.

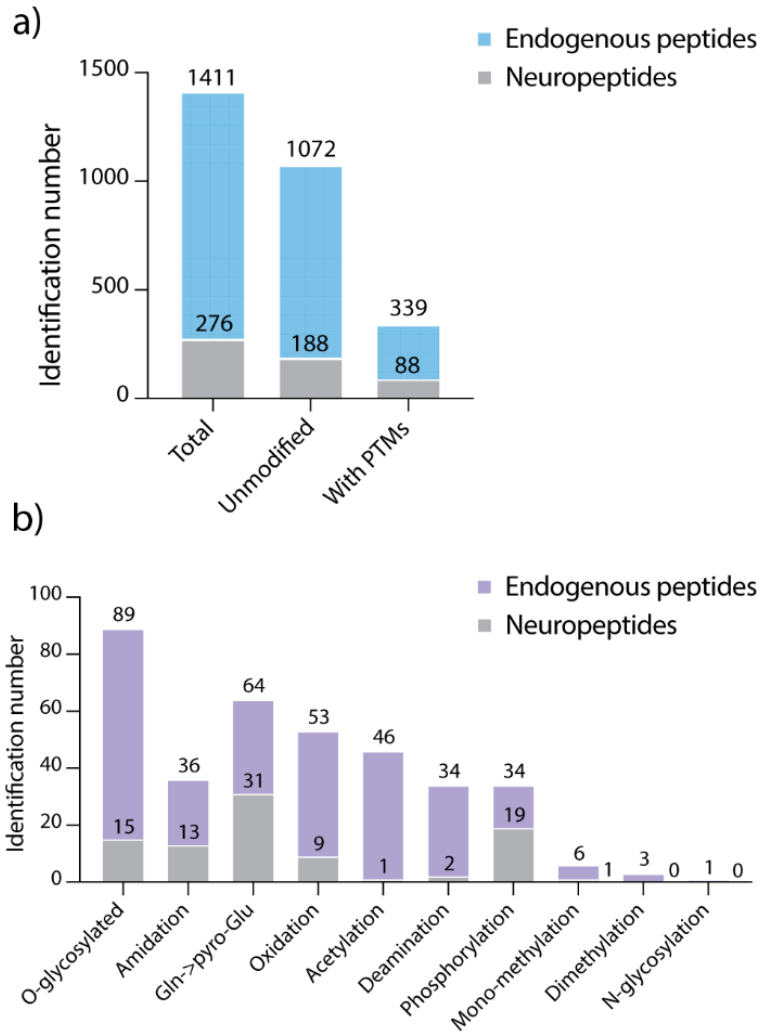


Figure 6. Endogenous peptidome coverage in CSF from healthy individuals. a) The number of endogenous peptides and neuropeptides identified with/without PTMs. b) The number of endogenous peptides and neuropeptides with different PTMs.

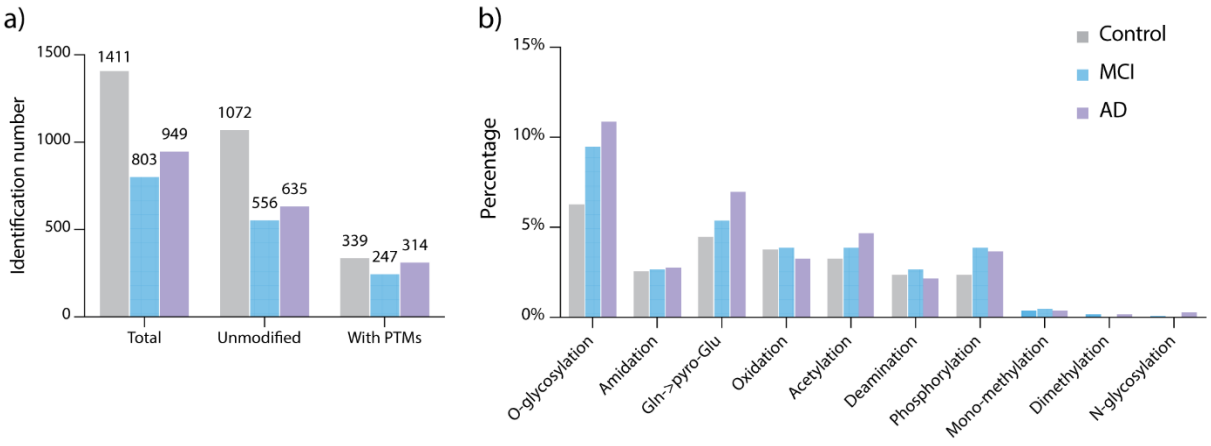


Figure 7. Endogenous peptidome comparison among control, MCI, and AD. a) The number of endogenous peptides identified with/without PTMs. b) The comparison of the percentage of endogenous peptides with different PTMs identified in control, MCI, and AD.

Supplemental Information

Supplemental Results and Discussion

Site-specific O-glycopeptide characterization enabled by EThcD. Among the various analytical techniques developed to advance the MS-based intact O-glycopeptide analysis, one focus is the fragmentation technique. The commonly used HCD fragmentation generates B/Y ions and abundant oxonium ions for glycan identification and b/y ions for peptide sequence information, but it does not provide much information for glycosite localization. On the other hand, 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. To take advantage of both modes, the "hybrid" dissociation method electron-transfer and higher-energy collision dissociation (EThcD), introduced by Heck and co-workers[1] has shown great potential for O-glycopeptide analysis. EThcD can produce rich fragment ion information for glycan moiety, peptide backbone and glycosylation site identification in one spectrum, which enables site-specific O-glycopeptide analysis.[2–6]

As an example, EThcD spectrum of the identified O-glycopeptide VHENENIGTTEPGEHQEAK is shown in **Figure S2a**. The O-glycan composition of HexNAcHexNeuAc, a sialylated Tf antigen, was assigned through the 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), 274.05 (NeuAc-18), 292.11 (NeuAc), HexNAcHex (366.14), and HexNAcHexNeuAc (657.23). The peptide sequence VHENENIGTTEPGEHQEAK were deduced based on the rich backbone fragments b/y ions and c/z ions. The c2−c10 ions enabled unambiguous

localization of the glycan at 9th threonine other than 10th threonine, which agreed well with previous reports.[7,8] Similarly, the O-glycosite of an endogenous O-glycopeptide AAVGTSAAPVPSDNH with 3 potential O-glycosites (5th T, 6th S, 12th S), was also unambiguously localized at the 6th serine residue using the c3-c7 ions (**Figure S2b)**. This O-glycopeptide originates from Apolipoprotein E, and the O-glycosite at 6th serine has been tentatively identified and reported as a novel O-glycosite in the previous study.[7] Although the author narrowed down the O-glycosite to the 5th threonine and 6th serine, the precise location could not be assigned due to a lack of informative fragments generated by electron-capture dissociation (ECD). Furthermore, limited by the enrichment method used in their study, the sialylation information of the O-glycan was lost. The EThcD spectrum in the current study not only helped us to precisely pinpoint this novel O-glycosite, but also allowed explicit assignment of the O-glycan composition with sialic acid structure fully preserved. In addition, labile phosphorylation was preserved as well benefitting from EThcD. As shown in **Figure S2c**, along with other b/y, c/z ions and 98 Da neutral loss, the detection of c4 and c5 ions helped us unambiguously localize the phosphorylation site at the 5th serine residue of an endogenous phosphorylated peptide VDPKSKEEDKH.

Selected examples of O-glycoproteins. In a previous CSF O-glycoproteome study, Halim et al. discussed the presence of O-glycosylation happened at Thr residues of the N-glycosylation Asn-X-Ser/Thr consensus motif.[8] This simultaneous occurrence of N- and O-glycosylation on this motif indicated that a preformed N-glycan structure might not necessarily prevent the polypeptide GalNAc transferase from interacting with its substrate. In our data set, we also found several O-

glycopeptides containing this special O-glycosylation site. For example, the GNLTGAPGQR peptide from endothelin B receptor-like protein 2 (ETBR2) was O-glycosylated at the 4th Thr residue, and the 2nd asparagine (Asn) underwent deamination **(Table S4)**. The deamidation of Asn residue was a sign of previous N-glycosylation since enzyme PNGase F was used in our study to first remove N-glycans. This finding was supported by the sequence analysis that this Asn residue was in the Asn-X-Ser/Thr N-glycosylation consensus motif and its annotation as a N-glycosylation site in Uniprot. Our previous CSF N-glycoproteome study also confidently identified the 2nd Asn residue of the GNLTGAPGQR peptide as N-glycosite.[9] Our results not only agreed with previous literature,[8] but also enabled identification of the intact O-glycan compositions as $HexNAc_{(2)}Hex_{(2)}$ and $HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}NeuAc_{(1)}$. A second example was the O-glycopeptide LPTTVLNATAK from protein YIPF3, with O-glycosylation at the 9th Thr and deamidation at 7th Asn. Both previous reports and our N-glycoproteome study confirmed 7th Asn as a true N-glycosite.[8,9] Two sialylated 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 has not been reported before were the two O-glycopeptides ELPGVCNETMMALWEECK and LANLTQGEDQYYLR from Apolipoprotein J (also known as clusterin) **(Table S4).** The deamidation at Asn residue in the two peptide sequences were proven to be N-glycosites.[8–11] These O-glycosites were first reported in our study due to their glycosylation type. Unlike sialylated ones on other two proteins, the O-glycan detected here were nonsialylated forms $HexNAc_{(3)}Fuc_{(1)}$ and $HexNAc_{(6)}Hex_{(5)}Fuc_{(3)}$. Such glycoforms are likely to be missed using sialic acid-specific enrichment method, while the

boronic acid enrichment method used in our study successfully captured nonsialylated O-glycoforms.

Experimental Section

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), and 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 subject information can be found in **Table S1**. All MCI and AD participants were diagnosed via applicable clinical criteria in standardized and multidisciplinary consensus conferences.[12,13] 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.[14,15]

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 Cell Culture. 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 °C.

Protein Extraction and Sigestion of PANC1 Cells. Protein extraction and trypsin digestion was performed based on previously reported filter-aided sample preparation (FASP) protocol[16] with some modifications. Cell pellets were lysed by sonication in lysis buffer (4% SDS, 100 mM Tris/Base pH 8.0). Bicinchoninic acid assay (BCA assay) was conducted to determine the protein concentration and 200 μg proteins were aliquoted. Disulfide bond reduction was performed with DTT at final concentration of 0.1 M at 95 ^{o}C for 3 min. 200 μL UA buffer (8 M UA in 100 mM Tris/Base) was added and sample solution was then 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 centrifuged again. 100 μL of IAA buffer (0.05 M IAA in UA buffer) was added onto the filter, followed by gentle swirl and incubation in darkness for 20 min. The filter was centrifuged at 14, 000g for 10 min. 100 μL more of UA buffer was added onto the filter and centrifuged. This step

was repeated for two more times to completely remove SDS. 100 μL of ABC buffer (50 mM) was added onto the filter and centrifuged at 14, 000g for 15 min. This step was repeated for two more times. All the centrifugation was done at 20 ^{o}C. Then 10 μL of trypsin and 40 μL of ABC buffer was added onto the filter, which was then incubated at 37 ^{o}C 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 centrifuged at 14, 000g for 10min for two times. TFA was added to make final concentration at 0.25%. Samples were desalted using a SepPak C18 SPE cartridge (Waters, Milford, MA).

CSF Sample Processing. The CSF sample was separated into a peptide fraction and a protein fraction using 10 kDa molecular weight cutoff (MWCO) following the previous protocol.[17,18] The peptide fraction was injected for LC-MS/MS analysis after desalting with SepPak C18 SPE cartridge (Waters, Milford, MA). 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 at room temperature. Samples were diluted with 50 mM Tris buffer to make urea concentration below 1 M. Trypsin was added at a 1:50 (w/w) ratio and incubated for 18 h at 37 °C. Digestion was quenched by adding TFA to a final concentration of 0.3%. Finally, the samples were desalted on a C18 SepPak cartridge and dried under vacuum. Tryptic peptides were incubated with PNGase F to remove N-glycans to avoid interference with O-glycosylated peptide detection. Salts and released N-glycans were removed by C18 SepPak cartridge desalting.

Boronic Acid Enrichment. Boronic acid enrichment was conducted according to a previously reported protocol with slight modifications.[13] 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 ºC for 2 h with both ends of the cartridge sealed. Unbound peptides were washed away with 1 mL anhydrous ACN for 3 times. The O-glycopeptides were eluted twice through incubation in 600μL 0.1% TFA at 37℃ for 1 h. The enriched O-glycopeptides were dried down under vacuum and stored under –80 °C before analysis.

High-pH Fractionation. Enriched O-glycopeptides were fractionated using a C18 reversed-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), 35–0% (50–54 min), 60–70% (54–58 min), and 70–100 % (58–59 min). Seven fractions were collected from 4 min to 62 min and 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% FA 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 acquisition (DDA) mode. An

MS1 scan was acquired from *m/z* 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 2.1 (PD 2.1). For O-glycopeptide analysis, 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) and deamidation of asparagine and glutamine (+0.98402 Da) as "rare" modification. O-glycosylation was set as "common" modification. Two rare modification and one common modification were allowed. Human O-glycan database embedded in Byonic containing 70 entries were used. For FDR control, Byonic default settings were applied that cut 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 generated more proteins. After that, Byonic estimated the spectrum-level FDR of the remaining PSMs to the reported proteins which were typically in the range 0-5%. Only those O-glycopeptides with PSMs FDR ≤ 1%, Byonic score > 50 and Delta Mod Score > 40 were reported. For endogenous peptides, three consecutive searches were conducted as shown in **Figure**

S1. Searches were performed with a precursor mass error tolerance of 10 ppm and a fragment mass error tolerance of 0.01 Da, and unspecific digestion. The 1st search used the whole human proteome database. Total "rare" modification was set to 1, and dynamic modifications included oxidation of methionine residues (+15.99492 Da), amidation at peptide C-terminus (-0.984016), acetylation at peptide N-terminus and lysine, serine (+42.010565), and Gln to pyro-Glu conversion (-17.026549). For 2nd search, a focused protein database was built on the protein precursors identified in the first search at 1% FDR together with proteins reported in literature.[19,20] The 2nd search used the same dynamic modification settings as 1st search, except for adding N-glycosylation and O-glycosylation as "common" modification, with total common modification set as 1. The human glycan database containing 182 N-glycans and 70 O-glycans were used. The 2nd search yielded a list of identified N-glycans and O-glycans, and these glycans were used to build a focused glycan database. For the 3rd search, focused protein database and focused glycan database were used. Dimethylation at peptide N-terminus, lysine, arginine (+28.0313), deamidation at asparagine (+0.984016), methylation at peptide N-terminus, lysine, arginine, glutamine (14.01565), phosphorylation at serine, tyrosine, threonine (+79.966331) were set as dynamic "rare" modifications and N-glycosylation and O-glycosylation were set as "common" modification, with total common modification set as 1. Only peptides with PSM FDR < 1%, Byonic score > 50 and PEP 2D < 0.01 were reported. For endogenous peptides with PTMs, a Delta Mod Score > 40 was required. All O-glycosylated endogenous peptide spectra were manually inspected. The peptide sequence analysis of glycosylation site-containing peptides was conducted by on-line tool Weblogo 3 (http://weblogo.threeplusone.com).[21]

References:

(1) Frese, C. K.; Altelaar, A. F. M.; Toorn, H. van den; 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. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

(2) Marino, F.; Bern, M.; Mommen, G. P.; Leney, A. C.; Brink, J. A. van G. den; Bonvin, A. M.; Becker, C.; Els, C. A. van; Heck, A. J. Extended O-GlcNAc on HLA Class-I-Bound Peptides. *J. Am. Soc. Mass Spectrom.* **2015**, *137*, 10922–10925.

(3) 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.

(4) 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.* **2018**, *427*, 35–42.

(5) 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.

(6) Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R. Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J. Proteome Res.* **2020**, *19*, 3286–3301.

(7) 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.

(8) Halim, A.; Rüetschi, U.; Larson, G.; 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.

(9) Chen, Z.; Yu, Q.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific Analysis of N-Glycoproteome in Human Cerebrospinal Fluid and Glycosylation Landscape Changes in Alzheimer's Disease. *Mol. Cell Proteomics* **2021**, *20*, 100081.

(10) 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*, 1493–1503.

(11) 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*, 2070–2080.

(12) 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.

(13) 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*, 2462–2475.

(14) Apweiler, R.; Hermjakob, H.; Sharon, N. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473*, 4–8.

(15) 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*, 8260–8287.

(16) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat Methods* **2009**, *6* (5), 359–362.

(17) 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*, 1327–1332.

(18) 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. *Proteom. - Clin. Appl* **2016**, *10* (12), 1225–1241.

(19) Hölttä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J. Peptidome Analysis of Cerebrospinal Fluid by LC-MALDI MS. *PLoS One* **2012**, *7*, e42555.

(20) 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*, 386–399.

(21) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190.

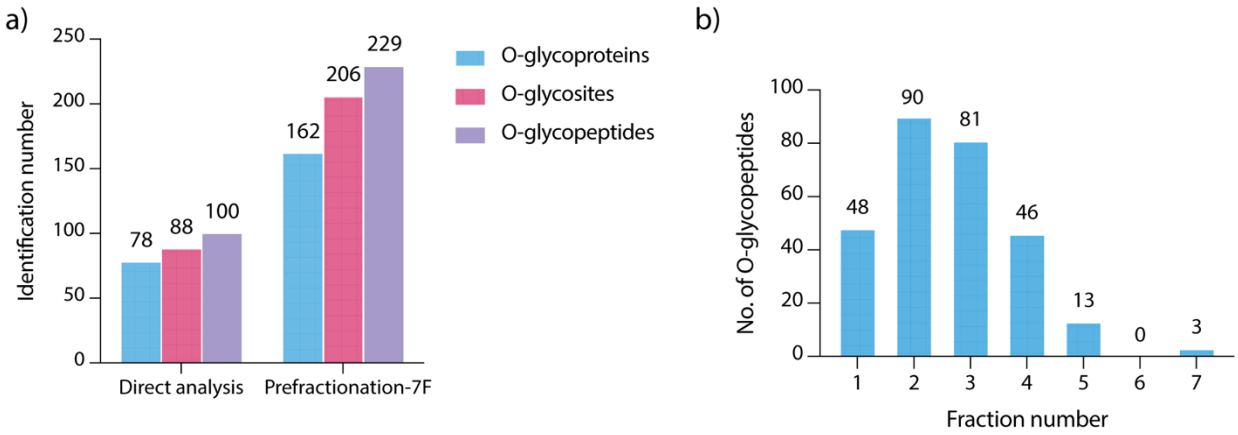


Figure S1. Boronic acid enrichment optimization. a) Comparison of O-glycoproteome coverage between direct analysis and prefractionation with 7 HpH fractions. b) Distribution of identified O-glycopeptides in 7 HpH fractions.

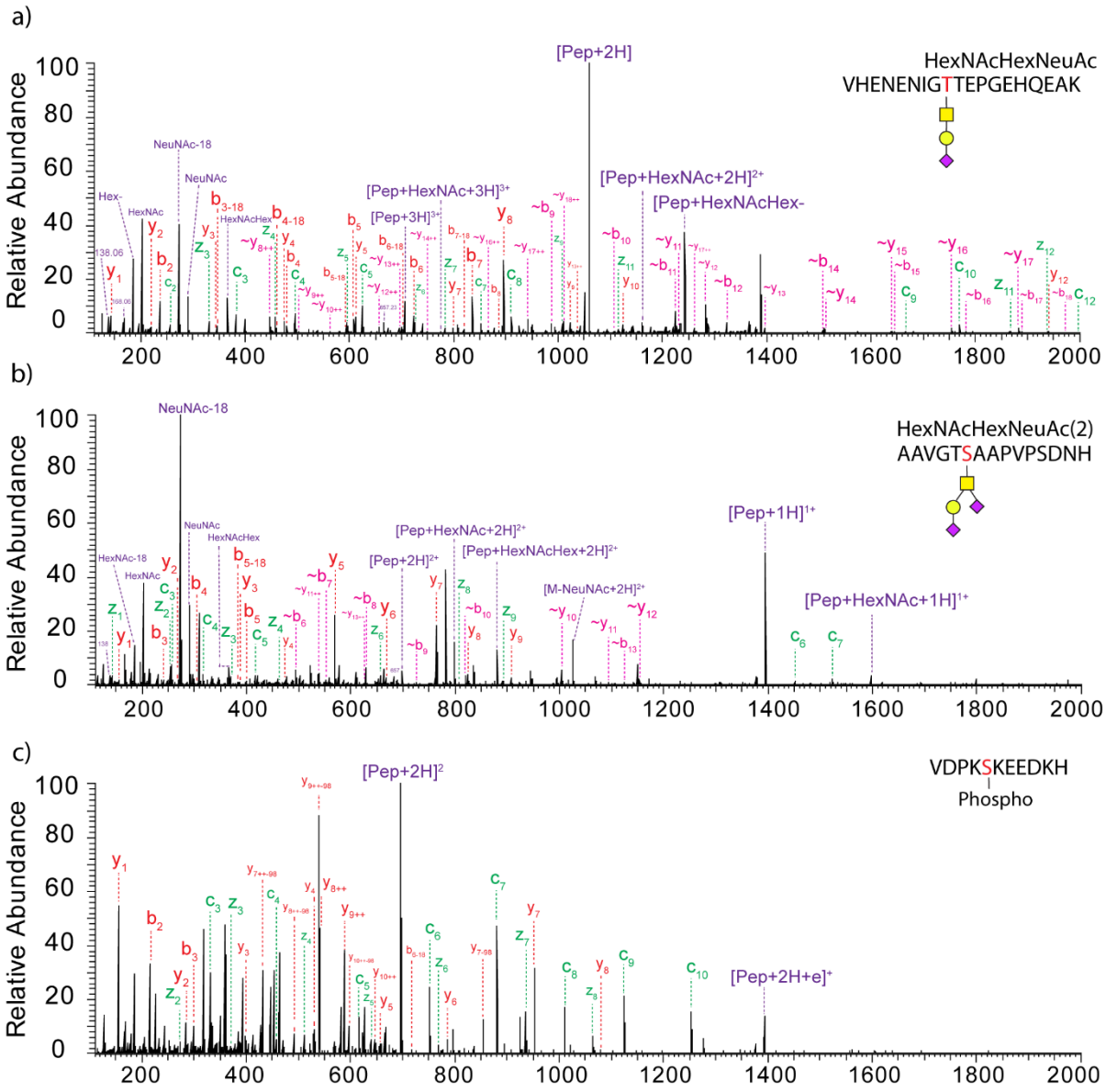


Figure S2. Representative EThcD fragmentation spectra of a) O-glycopeptide, b) endogenous O-glycopeptide, and c) endogenous phosphorylated peptide.

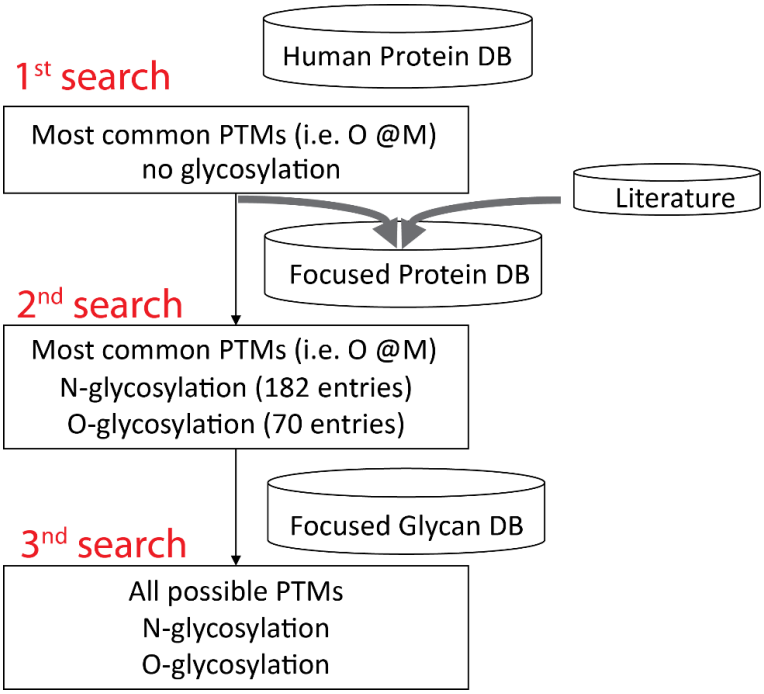


Figure S3. The three-step searching strategy for comprehensive PTMs analysis of CSF endogenous peptides.

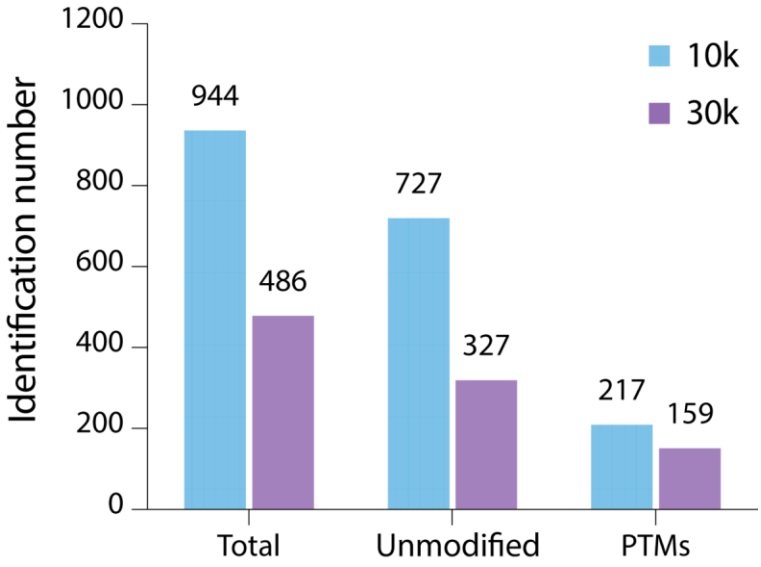


Figure S4. Comparison of endogenous peptide identification numbers between 10 kDa MWCO-based protocol and 30 kDa MWCO-based protocol.

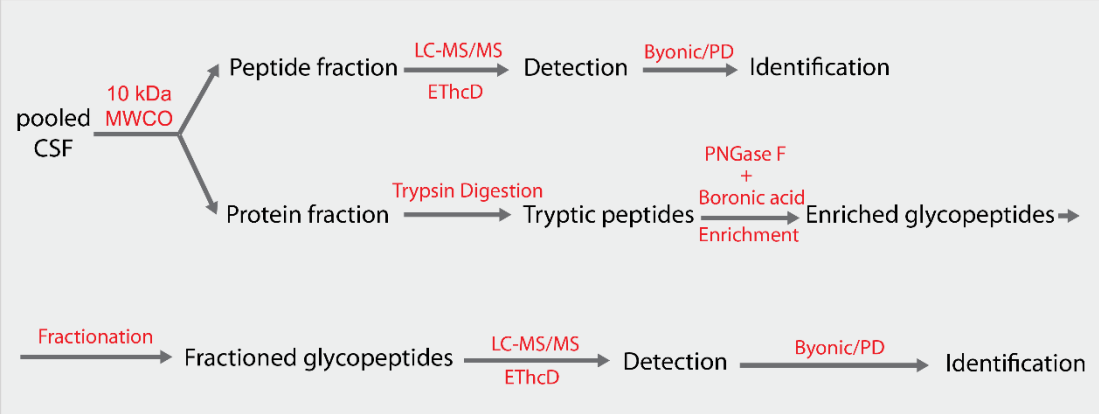


Figure S5. The overall workflow for O-glycoproteome and endogenous peptide analysis.

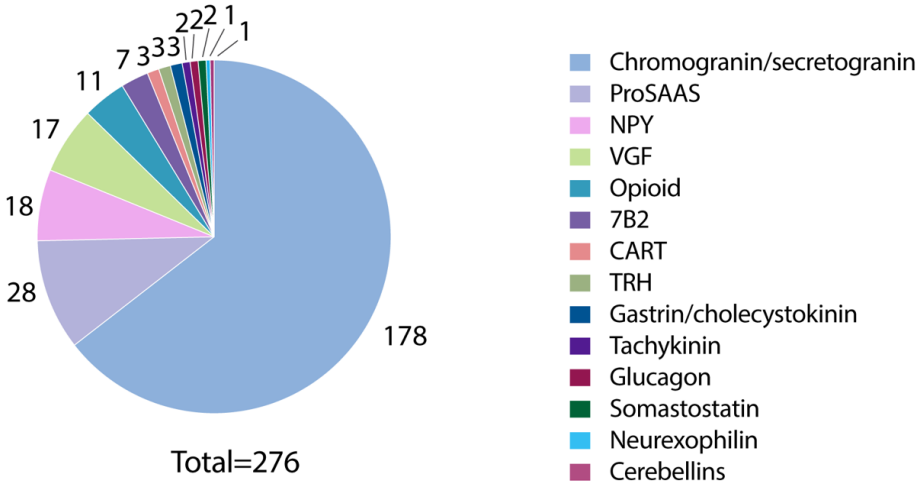


Figure S6. Different families of neuropeptides in CSF from healthy individuals.

Table S1. Subject 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%)

Table S2. O-glycopeptides identified using different enrichment strategies.

https://pubs.acs.org/doi/suppl/10.1021/acschembio.1c00932/suppl_file/cb1c00932_si_003.xlsx

Table S3. O-glycoproteome characterization in control, MCI and AD.

https://pubs.acs.org/doi/suppl/10.1021/acschembio.1c00932/suppl_file/cb1c00932_si_004.xlsx

Table S4. Selected O-glycoproteins in control, MCI and AD.

https://pubs.acs.org/doi/suppl/10.1021/acschembio.1c00932/suppl_file/cb1c00932_si_005.xlsx

Table S5. Endogenous peptides identified using 10kDa MWCO and 30kDa MWCO.

https://pubs.acs.org/doi/suppl/10.1021/acschembio.1c00932/suppl_file/cb1c00932_si_006.xlsx

Table S6. O-glycosylated neuropeptides identified in CSF.

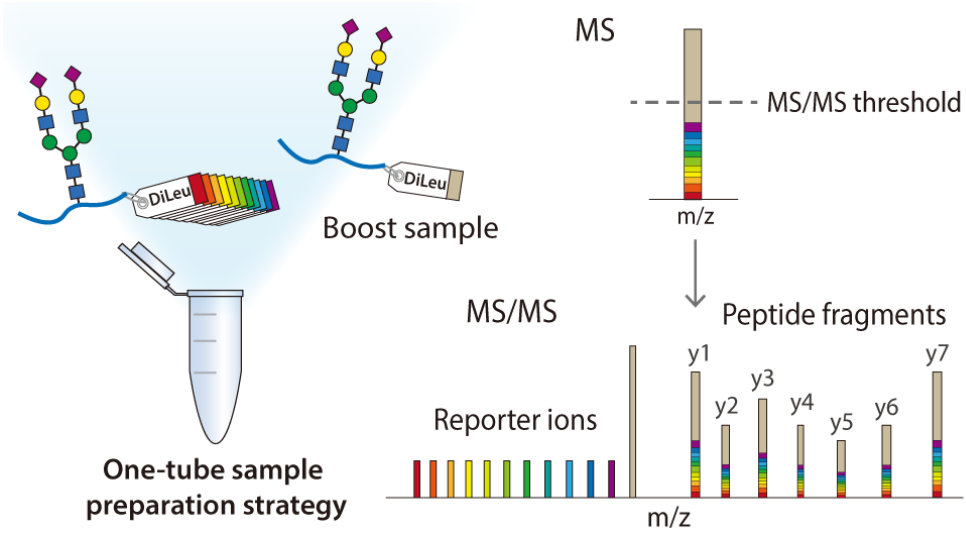
https://pubs.acs.org/doi/suppl/10.1021/acschembio.1c00932/suppl_file/cb1c00932_si_007.xlsx

Table S7. Endogenous peptides identified in control, MCI and AD.

https://pubs.acs.org/doi/suppl/10.1021/acschembio.1c00932/suppl_file/cb1c00932_si_008.xlsx

Chapter 6

Boost-DiLeu: Enhanced Isobaric *N*,*N*-Dimethyl Leucine Tagging Strategy for Comprehensive Quantitative Glycoproteomic Analysis



Adapted from: **Wang, D.**[#]; Ma, M.[#]; Huang, J.; Gu, T.-J.; Cui, Y.; Li, M.; Wang, Z.; Zetterberg, H.; Li, L. Boost-DiLeu: Enhanced Isobaric *N*,*N*-Dimethyl Leucine Tagging Strategy for a Comprehensive Quantitative Glycoproteomic Analysis. *Anal. Chem.* **2022**, 94 (34), 11773–11782. ([#]co-first authors) Author contribution: study was designed by D.W, M.M., and J.H under the supervision of L.L.; experiment was performed by D.W.; DiLeu tags were synthesized by T.G., M.L., and Z.W.; data was analyzed by D.W and M.M.; samples were provided by M.M. and H.Z.; manuscript was written by D.W., and edited by M.M., J.H., T.G., Y.C., M.L., and L.L.

Abstract

Intact glycopeptide analysis has been of great interest because it can elucidate glycosylation site information and glycan structural composition at the same time. However, mass spectrometry (MS)-based glycoproteomic analysis is hindered by the low abundance and poor ionization efficiency of glycopeptides. Relatively large amounts of starting materials are needed for the enrichment, and this makes the identification and quantification of intact glycopeptides from samples with limited quantity more challenging. To overcome these limitations, we developed an improved isobaric labeling strategy with an additional boosting channel to enhance *N*,*N*-dimethyl leucine (DiLeu) tagging-based quantitative glycoproteomic analysis, termed as Boost-DiLeu. With the integration of one-tube sample processing workflow and high-pH fractionation, 3514 quantifiable *N*-glycopeptides were identified from 30 μg HeLa cell tryptic digests with reliable quantification performance. Furthermore, this strategy was applied to human cerebrospinal fluid (CSF) samples to differentiate *N*-glycosylation profiles between Alzheimer's disease (AD) patients and non-AD donors. The results revealed processes and pathways affected by dysregulated *N*-glycosylation in AD, including platelet degranulation, cell adhesion, and extracellular matrix, which highlighted the involvement of *N*-glycosylation aberrations in AD pathogenesis. Moreover, weighted gene co-expression network analysis (WGCNA) showed 9 modules of glycopeptides, two of which were associated with the AD phenotype. Our results demonstrated the feasibility of using this strategy for comprehensive glycoproteomic analysis of size-limited clinical samples. Taken together, we developed and optimized a strategy for the enhanced comprehensive quantitative intact glycopeptide analysis with DiLeu labeling, showing significant promise for

identifying novel therapeutic targets or biomarkers in biological systems with limited sample quantity.

Introduction

Glycosylation is one of the most common post-translational modifications (PTMs), which is involved in many physiological processes including cell signaling, host-pathogen interaction, and immune response.[1–3] Studies have shown that aberrant glycosylation plays a key role in the pathological processes during disease progressions, such as neurodegenerative diseases, diabetes, and cancers.[4–7] Intact glycopeptide analysis has been of great interest recently because it retains information on glycosylation sites while elucidating peptide sequences and glycan structures, which enables the investigation of functional effects of heterogeneity across glycoproteome.[8] However, akin to many other PTMs, mass spectrometry (MS)-based glycoproteomic analysis has been hindered by the low abundance of glycopeptides among complex protein digests and their poor ionization efficiency.[9,10]

To overcome these limitations, various enrichment strategies have been developed to separate glycopeptides from other non-glycosylated peptides in biological samples before MS analysis.[4] For example, hydrophilic interaction liquid chromatography (HILIC) has been extensively used to enrich glycopeptides based on their increased hydrophilicity introduced by glycan moieties.[11–13] However, even utilizing these highly efficient enrichment methods, an in-depth glycoproteomic analysis still requires a relatively large amount of starting sample. This drawback particularly limits glycoproteomic analysis of precious clinical samples such as pathological tissues and

cerebrospinal fluid (CSF), which cannot be expanded in vitro and are often limited in quantities for research.[14]

On the other hand, insufficient MS signal intensity from low-abundance peptides usually leads to low proteome coverage in the LC-MS/MS analysis because the poor-quality MS/MS spectra hardly generate confident peptide identifications. Isobaric labeling can be used to enhance MS detection sensitivity in the data-dependent acquisition (DDA) mode analysis since the signal intensities of precursor ions in the full scan are combined from all labeling channels of the same species.[13] Abundant peptide backbone fragments and enhanced signal intensities in MS/MS facilitate identification while achieving multiplex quantification at the same time. Recent studies have taken advantage of isobaric labeling and introduced the concept of a "boosting" (or "carrier") channel, in which a large amount of content-relevant samples is labeled by one channel of the isobaric tags and combined with those size-limited samples labeled by the rest channels. The combined signal intensity of a given peptide can be greatly amplified, thus enhancing the detection of low-abundance peptides, as well as mitigating the impact of sample loss on low-volume samples during sample preparation.[15,16] Such a strategy has been applied to single-cell proteomics,[17,18] quantification of phosphopeptides,[15] phosphotyrosine-containing peptides,[16] deglycosylated peptides of secreted glycoproteins,[19] stable isotope labeling using amino acids in cell culture (SILAC)-labeled peptides,[20] and low-abundant proteins in thermal proteome profiling (TPP).[21] These studies showed promising features of the boosting approach to increase the proteome coverage and enable quantification of low-abundance proteins/peptides simultaneously, which

would be especially useful for the PTM analysis of size-limited samples. Nonetheless, such a strategy has not been applied to intact glycopeptide analysis yet.

Our group previously developed *N*,*N*-dimethyl leucine (DiLeu) isobaric tags for quantitative proteomics analysis, where the amine-reactive group reacted with N-termini and lysine residues of peptides similar to tandem mass tags (TMTs)[22] and isobaric tags for relative and absolute quantification (iTRAQ)[23]. Compared to these commercially available tags, DiLeu is much more cost-effective and can be synthesized in-house readily at high yields.[24] It was first designed as a 4-plex set[25] and then has been expanded to 12- and 21-plex by incorporating neutron encoding (NeuCode) strategy, which utilized the differences in mass between each isotope of every element, arising from their differences in nuclear binding energy.[24,26–28] In the 12-plex DiLeu setting, reporter ions are grouped into four regions ranging from *m/z* 115 to *m/z* 118, and the mass difference of adjacent channel is as small as 6 mDa within each region.[24] DiLeu tag is well-suited for developing a boosting strategy for intact glycopeptides since it has been successfully applied to intact glycoproteomic quantification using cell lines and tissue samples in our previous studies.[11,29] More importantly, it is less likely to be affected by the isotopic impurity "leakage" problem reported for 10-plex TMT due to the use of the NeuCode strategy.[15]

In this study, we developed a 12-plex DiLeu tag-based boosting strategy (Boost-DiLeu) to perform a comprehensive quantitative *N*-glycoproteomic analysis of size-limited samples (**Figure 1**). To further recover glycopeptides from samples of small quantity, we adapted a one-tube processing strategy[14] for DiLeu labeling, leveraging the benefits of an acid-cleavable detergent, RapiGest SF, to circumvent the necessities of sample desalting. The strategy was well established

by carefully optimizing the sample-handling processes, boosting-to-study channel (B/S) ratios, and instrumental parameters for data acquisition including automatic gain control (AGC) and ion injection time. Coupling with HILIC enrichment and high-pH (HpH) reversed-phase liquid chromatography (RPLC) fractionation, large-scale global mapping of *N*-glycoproteome was achieved from a small amount of HeLa cell digests. To further demonstrate the feasibility of the Boost-DiLeu strategy for size-limited clinical sample analysis, we quantified alterations of the *N*-glycoproteomes in CSF between Alzheimer's disease (AD) patients and non-AD donors. Taken together, the Boost-DiLeu strategy not only increased the glycoproteome coverage but also enabled accurate and robust quantification, which shed light on future applications to site-specific quantitative glycoproteomic studies involving samples of limited availability.

Experimental Section

Supplementary Experimental Procedures. Details about chemicals and materials, cell culture, sample preparation with the conventional protocol, offline HpH fractionation, global proteomic analysis, and expanded data analysis are provided in the Supporting Information.

HeLa Cell Sample Preparation. In the one-tube sample processing workflow, HeLa cell pellets were solubilized in 0.1% (w/v) RapiGest prepared in 100 mM TEAB with 1% (v/v) protease inhibitor and phosphatase inhibitor. The mixture was then incubated at 60 ^{o}C for 30 min and sonicated at 4 ^{o}C for 1 min. Lysates were centrifuged at 21000 g at 4 ^{o}C for 5 min and the supernatant was collected. Protein BCA assay, reduction, alkylation, and trypsin digestion were performed in the same way as mentioned in the conventional method. After digestion, peptides were dried down *in vacuo* without acidification and desalting.

CSF Sample Preparation. Detailed information on CSF samples is provided in Supplemental Information and **Table S1**. The study was approved by the regional ethics committee at the University of Gothenburg. After a BCA protein assay, 30 µg of proteins was aliquoted for each study channel. 900 µg of proteins pooled from both AD patients and non-AD donors CSF samples at a ratio of 3:5 was used as the boosting channel. The aliquots were dried down *in vacuo* and treated with the one-tube sample preparation approach, as described above.

DiLeu Labeling. DiLeu tags synthesis and labeling were performed as previously reported.[24,29] DiLeu tags were activated in anhydrous DMF with DMTMM and NMM at 0.6× molar ratios to tags. The mixture was vortexed at room temperature for 1h and the supernatant was then added to each sample for labeling. After vortexing at room temperature for 2 h, the reaction was quenched by adding 5% NH_2OH to a final concentration of 0.25%. Each batch of labeled peptides was pooled together and dried down *in vacuo.*

SAX-HILIC Enrichment. Enrichment of DiLeu-labeled glycopeptides was performed with in-house-packed strong anion exchange (SAX)-HILIC SPE tips following previous publications with minor modification.[29,30] Briefly, 3 mg of cotton wool was plugged into a 200 µL empty TopTip, which was placed on a 2 mL microcentrifuge tube with an adapter unit. After activation in 1% TFA for 15 min, the SAX material was transferred into the cotton-packed TopTip at a beads-to-peptide ratio of 30:1. The solvent was removed at 200 g for 2 min. The stationary phase was conditioned with 300 µL of 1% TFA and loading buffer (80% ACN/1% TFA) three times. The DiLeu-labeled sample was dissolved in 300 µL of loading buffer and loaded onto the SAX-cotton tip. The tip was centrifuged at 200 g for 2 min and flow-through was reloaded to the SAX-cotton

four more times to ensure complete retention. The SAX-cotton tip was washed with 300 µL of loading buffer six times, and then eluted with 150µL of 0.1% FA solution three times. Samples were dried down *in vacuo* before direct MS analysis or HpH fractionation.

LC-MS/MS Analysis. Lyophilized samples reconstituted in 0.1% FA were loaded onto a 15 cm length, 75 µm i.d. in-house-packed Bridged Ethylene Hybrid C18 (1.7 µm, 130 Å, Waters) column and analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex UltiMate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Two technical replicates were run for each fraction. For the intact *N*-glycopeptide analysis, binary buffers (A, 0.1% FA; B, 90% ACN, 0.1%FA) were used in LC and the linear gradient was from 0 to 30% B for 80 min. MS survey scans of peptides were acquired from 400 to 2000 *m/z* at a resolution of 120 K, using an AGC target setting of 4E5 and a maximum injection time of 100 ms. For the MS2 scan, a duty cycle of 3s was set in top speed mode. Only spectra with a charge state among 2-7 were selected for fragmentation by stepped higher-energy collision dissociation (HCD) with a normalized collision energy (NCE) of 30 ± 3%. The MS2 spectra were acquired with a resolution of 60 K, a lower mass limit of *m/z* 110, and a dynamic exclusion of 12 s with 10 ppm mass tolerance. Different AGC settings and maximum injection times were tested for method optimization. All data were acquired in profile mode.

Data Analysis. Raw files were searched against the UniProt Homo sapiens reviewed database (August 2020, 20311 sequences) using the Byonic search engine (version 2.9.38, Protein Metrics Inc) embedded within Proteome Discoverer 2.1 (PD 2.1, Thermo Fisher Scientific). Trypsin was selected as the enzyme and two maximum missed cleavages were allowed. Searches were

performed with a precursor mass tolerance of 15 ppm and a fragment mass tolerance of 0.03 Da. Static modifications were specified as carbamidomethylation (+57.02146 Da) on cysteine residues and 12-plex DiLeu (+145.12801 Da) on peptide N-terminus and lysine residues. Dynamic modifications consisted of the oxidation of methionine residues (+15.99492 Da), deamidation (+0.984016 Da) of asparagine and glutamine residues, and *N*-glycosylation. Oxidation and deamidation were set as "rare" modification, and *N*-glycosylation was set as "common" modification. Glycan modifications were searched against a glycan database expanded from the Byonic embedded 182 human *N*-glycan database to include mannose-6-phosphate (M6P) glycans consisting of HexNAc (2-4) Hex (3-9) Phospho (1-2) modifications. *N*-glycopeptides were filtered at a 1% peptide FDR, Byonic score >150 and log|Prob| >1. Glycopeptides were exclusively categorized into six glycosylation type categories based on the glycan composition identified: (1) M6P (containing M6P glycan); (2) both sialylated and fucosylated; (3) sialylated only; (4) fucosylated only; (5) complex/hybrid (> 2 HexNAc); (6) oligomannose (2 HexNAc and ≥ 5 Hex), and (7) paucimannose (1-2 HexNAc and < 5 Hex). Quantification was performed in PD 2.1 with a reporter ion integration tolerance of 10 ppm for the most confident centroid. Reporter ion intensities were exported, and isotopic interference correction was performed with a Python script according to the previously described equations.[24] The reporter ion intensity of each channel was normalized by the median to correct systematic biases of 12-plex DiLeu tags via Perseus.[31] Student's t-test was performed, where significant change was defined by a p-value less than 0.05 and fold-change over 1.5.[31] The MS data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the accession code PXD029269.[32]

Results and Discussion

One-Tube Sample Processing Workflow for DiLeu-Labeled Glycoproteomics. Recently, Wu et al introduced a strategy termed Nanogram TMT Processing in One Tube (NanoTPOT) to achieve desalt-free isobaric labeling using TMT tags in a single tube.[14] An acid-cleavable detergent, RapiGest SF, was utilized to eliminate the need for commonly used SDS lysis buffer and urea that are detrimental to MS and tag labeling. Here we adapted and optimized this strategy to conduct desalting-free DiLeu labeling. To assess the performance of this optimized one-tube DiLeu labeling sample processing, the same amounts of HeLa cell pellets were processed both in the conventional method and this method as described in the Experimental Section. SDS-PAGE result suggested that the trypsin digestion was complete in both methods (**Figure S1**). In terms of proteome coverage, both workflows identified comparable numbers of labeled peptides in global proteome analysis (**Figure S2A**), and 80.1% of proteins identified in one-tube processing workflow were also found in the conventional method (**Figure S2B**). For DiLeu-labeled glycoproteomics, the one-tube sample processing method outperformed the conventional one in all aspects, including identification of glycopeptide-spectrum matches (GPSMs), unique glycopeptides and corresponding glycoproteins (**Figure 2A**). Among all identified glycoproteins, 57.1% were found in both methods, while the one-tube processing method recovered 13.8% (35 out of 254) more unique glycoproteins than the conventional one (**Figure 2B, Table S2**). The excellent performance of the one-tube sample processing method can be attributed to two factors: (1) RapiGest is a preferred reagent for handling membrane proteins and the protein precipitation step can be omitted since there is no need to remove RapiGest after extraction.[33] (2) Avoiding

desalting before labeling effectively reduces sample loss, especially for hydrophilic glycopeptides. Taken together, the one-tube sample processing strategy is suitable for DiLeu-labeled glycoproteomic analysis.

Evaluation of Boost-DiLeu Strategy. In the Boost-DiLeu strategy, the analysis of small-quantity samples labeled with study channels was enhanced by utilizing a boosting channel that contained a much larger amount of labeled glycopeptides from analogous biological samples. After being pooled together, isobaric labeled glycopeptides from each channel appeared as a single precursor ion at the MS1 level, and the existence of boosting channel greatly increased the intensity of the precursor ions and facilitated triggering MS2 fragmentation. To validate the design of the Boost-DiLeu strategy and evaluate the effects of different boosting-channel-to-study-channel (B/S) ratios of the sample amount on glycoproteomic coverage and quantification, a series of comparisons were performed using HeLa cell tryptic digests prepared with the one-tube sample processing workflow. The first three channels (115a,115b, and 116a) in 12-plex DiLeu were set as study channels and loaded with 10 μg tryptic peptides respectively. The last channel, DiLeu 118d, was selected as the boosting channel to generate different B/S ratios at 30×, 50×, and 100× (**Figure 3A**). A control group was prepared at the same time that only consisted of the first three study channels. Each group of samples was pooled together after labeling and enriched through HILIC before LC-MS/MS analysis.

DiLeu 118d was selected as the boosting channel because the reporter ion in this channel had no interference with other channels. Its -1 isotopic peak (^{2}H to ^{1}H) caused by isotopic impurity was at *m/z* 117.14656, which was 3 mDa away from the DiLeu 12-plex 117c reporter ion at *m/z*

117.14363. Such mass difference can be resolved using high-resolution capability during LC-MS/MS analysis. **Figure 3B** displayed the reporter ion signal intensity distribution at a 30× B/S ratio. A similar reporter ion intensity distribution in study channels were observed in all groups and none of the empty channels had unusually high reporter ion signal, indicating that 118d channel successfully served as a boosting channel and avoided the isotopic interference problem reported in 10-plex TMT tags,[15] which ensured high quantification accuracy and maximum multiplexing capacity of DiLeu tags.

Figure 3C demonstrats the effectiveness of adding a boosting channel, as the quantifiable glycopeptides in study channels, which stand for glycopeptides without missing reporter ion intensities, increased three times at B/S 30× compared to the control group. As expected, the identified number of GPSMs went up from 2132 to 2370 as the B/S ratios increased. However, the unique glycopeptide number and the quantifiable glycopeptides dropped as the B/S ratios increased. Simultaneously, an increased median coefficient of variations (CV) from 13.41% to 28.19% (**Figure 3D**) was observed, which showed a similar trend reported in other studies.[15,18] Since the total number of ions entering the Orbitrap was controlled in each scan, ions from boosting channel were likely to occupy more space than those in study channels. At a higher B/S ratio, the signal intensity of reporter ions from study channels showed a slightly decreasing trend (**Figure S3**), which could be a result of reporter ion suppression. This negatively affected the quantification performance, including more missing values found in 50× and 100× groups and increased CVs. The same comparison was also made at three different B/S ratios using samples prepared with the conventional method (**Figure S4**). Despite the lower identification number, a similar trend of fewer

quantifiable glycopeptides and a higher median of CVs was observed with the increase of B/S ratios, suggesting that 30× boosting ratio was optimal in the aspect of identification and quantification accuracy.

Optimization of Instrument Parameters. When performing MS/MS analysis, the number of precursor ions reaching to the orbitrap analyzer is controlled by the AGC setting and maximum injection time. These parameters are essential to balance the detection sensitivity and MS2 scan rate. For global proteomic analysis, AGC is usually set in the range of 5E4 to 1E5 to improve the overall coverage.[18] However, such a setting may not be well suited for PTM analysis, especially in the case where a large portion of boosting samples are mixed with study samples. Ions from boosting samples can fill in the analyzer quickly, thus impairing the detection of signals from study channels. To understand the influence of AGC settings on glycoproteome coverage and quantification quality, four different AGC settings including 5E3, 3E4, 5E4, and 5E5 were compared with DiLeu-labeled HeLa cell tryptic digests with a 30× boosting channel added. The maximum injection time was fixed at 200 ms. In general, a higher AGC setting allows more ions to accumulate for MS/MS analysis at expense of MS2 scan rate and longer duty cycle time, thus affecting proteome coverage.[18,34] Such a trend was observed in **Figure 4A**, where AGC 3E4 achieved the highest identification numbers of GPSMs, glycopeptides, and quantifiable glycopeptides while higher AGC values led to decreasing identification numbers. As reflected in **Figure 4B**, the median injection time increased from 18 ms to 200 ms when increasing AGC. The lowest AGC setting 5E3 resulted in significantly poorer glycoproteome coverage, indicating the necessity of allowing enough ions to enter the orbitrap for high-quality MS/MS analysis. On the

other hand, accumulating more ions improves quantification performance since more ions from study channels are allowed to enter the mass analyzer.[18,35] Though the reporter ion intensity decreased slightly at higher AGC settings due to potential space charge effects[36] or preference of fragmentation on glycosidic bonds (**Figure 4D**), the median of CV distribution decreased from 23.63% to 12.97% (**Figure 4C**).

Besides the AGC settings, maximum injection time is another important parameter to control the number of ions for MS2 analysis. To determine an optimal setting, two different maximum injection time at 200 ms and 300 ms were evaluated with the same B/S ratio of 30× and AGC setting of 5E4. As shown in **Figure S5A**, increased injection time elongated duty cycle time, thus resulting in lower identification numbers. In terms of quantification performance, the distribution of reporter ion signal intensity (**Figure S5B**) and CVs (**Figure S5C**) were almost similar in both cases, despite that 200 ms injection time had a slightly lower median of CV of 16.61% than 17.71% of 300 ms injection time setting.

These results clearly demonstrated a trade-off between glycoproteome coverage and quantification performance in different AGC settings. Higher AGC improved quantification quality at expense of the number of identified glycopeptides. Maximum injection time also had a great impact on glycoproteome coverage, while the influence on quantification performance was not as significant as AGC settings. To achieve a balance between glycoproteome coverage and quantification performance, AGC 3E4 and maximum injection time 200 ms were employed in the following experiments for glycoproteomics analysis.

Global Glycoproteome Mapping in HeLa Cell Line. HpH provides an additional dimension of separation from conventional low-pH RPLC due to its high orthogonality and compatibility with direct MS analysis.[37] Since the conventional HILIC fractionation based on decreasing ACN concentration showed less efficiency for glycopeptide fractionation, HpH fractionation was implemented to further boost the glycoproteome coverage in this study.[38] Though HpH fractionation has been adopted in several PTM analyses to bolster the proteome coverage, additional separation requires a sufficient amount of sample to compensate for the potential sample loss during the process, and this is usually unsuitable for samples of low quantity.[15,39–41] With the Boost-DiLeu strategy, a larger amount of boosting sample can serve as a "carrier" sample that mitigates the impact of sample loss.[42] After HILIC enrichment and HpH fractionation, a total of 4119 glycopeptides corresponding to 277 proteins were identified from 30 µg HeLa tryptic digest peptides labeled in the first three study channels, among which 3514 were quantifiable (**Figure 5A**, **Table S3**). Glycoproteome coverage was expanded by approximately 2.5-fold compared to the analyses without HpH fractionation. **Figure 5B** displays the overlap of glycopeptides identified in four fractions, and 29.62%, 15.36%, 17.56%, and 23.64% of glycopeptides were unique in each fraction, indicating the high separation efficiency provided by HpH fractionation. Such comprehensive intact glycoproteome coverage also enabled the global profiling of site-specific micro-heterogeneity of HeLa cell glycosylation, as visualized in **Figure 5C**. This glycoprotein-glycan network diagram mapped which glycans (outer nodes, 151 total) modified which proteins (inner bar, 277 total). The glycoproteins were sorted based on the different numbers of glycosites, and all nodes and edges were colored by the corresponding glycosylation type. Of note, more than

half of the glycoproteins had only one glycosylation site, but 79.1% of glycosites have more than one glycan. In addition, several glycosylation patterns were observed, such as the prevalence of oligomannose glycosylation (44.11% of glycopeptides). Interestingly, 34 out of 151 glycans were found to be M6P glycans (a representative MS/MS spectrum shown in **Figure S6).** Due to the negatively charged phosphate group and extremely hydrophilic property, the detection of M6P glycopeptides usually suffers from interference by other *N*-glycopeptides.[43] With HpH fractionation, M6P glycopeptides were better separated from other glycopeptides so the detection sensitivity was enhanced. Along with improved glycoproteome coverage, the Boost-DiLeu strategy also presented good quantification performance after HpH fractionation. **Figure 5D** shows that the median intensities of study channels were close to 1:1:1 without any isotopic interference from the DiLeu 118d channel. A CV median of 16.75% (**Figure 5E**) and the average Pearson correlation over 0.97 (**Figure 5F**) among three biological replicates demonstrated the robustness and accuracy of quantification of this strategy.

Site-Specific Quantitative Glycoproteomic Analysis of Human CSF Samples. Through direct interchanges with the extracellular fluid of the central nervous system (CNS), CSF can directly reflect pathological changes in the CNS, providing the opportunity for biomarker discovery in neurological diseases.[44–47] However, the analysis of CSF samples is challenging due to the limited availability and the relatively low total protein concentration (0.2 - 0.8 mg/mL).[48] To address these challenges, the Boost-DiLeu strategy was adopted to profile glycoproteomic changes in CSF between AD patients and non-AD donors at B/S ratio of 30×. In total, 1321 intact *N*-glycopeptides were identified, and 1172 of them were quantifiable (with at least three valid reporter ion intensities

in one group), mapping to 164 glycoproteins and 158 different glycans (**Figure S7A**, **Table S4**). **Figure S7B** shows the glycoprotein-glycan network of all quantifiable glycopeptides. Nearly 62% of the glycoproteins were observed with only one glycosite, yet 47% of the glycosites were modified with more than one glycan. In the meantime, 5% of glycoproteins were found to have more than 4 glycosylation sites, suggesting a high degree of heterogeneity in CSF glycosylation. Among all quantifiable glycopeptides, 49% and 22% of them were sialylated and fucosylated, and the prevalence of these two glycosylation types was also in line with our previously reported label-free study of *N*-glycopeptides in CSF samples.[39]

Compared to non-AD controls, 18 glycopeptides were found significantly changed, with 9 upregulated and 9 downregulated by student's *t*-test (p-value<0.05) with a fold-change over 1.5 (**Figure 6A**, **Figure S8A**). The corresponding 14 glycoproteins were not observed to display the alteration of abundances in parallel global proteomics analysis, suggesting that the dysregulation of glycopeptides was likely from the PTM level (**Figure S8B**). Among these glycopeptides, 13 of them were sialylated and 7 were fucosylated, which further emphasized the significance of the two types of glycosylation on the pathogenesis of AD.[5,49] GO analysis showed that these dysregulated glycoproteins were significantly enriched in platelet degranulation, reverse cholesterol transport, cell adhesion, and complement activation, which have been reported to be associated with neuronal degeneration in AD (**Figure S8D**).[50–53] A majority of dysregulated glycoproteins in AD were localized in the extracellular region, in accordance with the fact that CSF is in direct contact with the extracellular space of the brain and many *N*-glycoproteins are secreted proteins (**Figure S8E**).[54,55] The protein-protein interaction network indicated that these proteins were mainly related

to platelet degranulation and response to elevated platelet cytosolic Ca^{2+} (**Figure S8C**). Platelet degranulation was reported to play a pivotal role in platelet-mediated amyloid-β (Aβ) oligomerization in AD, which is believed to be greatly relevant to disease progression.[56,57] Alterations in calcium homeostasis have been found to be critically implicated in brain aging and the neuropathology of AD, and the involvement of glycoproteins with response to elevated platelet cytosolic Ca^{2+} may provide deeper insight into the role of the cytosolic calcium level during the progression of AD.[58,59] Among these glycoproteins, four upregulated sialyl-glycopeptides were observed at two sites of prostaglandin-H2 D-isomerase (PTHDS, Uniprot accession: P41222), a protein that is involved in multiple CNS functions and acts as a chaperone for preventing the formation of neurotoxic agents such as Aβ fibrils.[60] This finding suggests that aberrant sialylation on this protein might play a specific role in AD progression through accelerated Aβ aggregation.[61]

To further study individual *N*-glycoproteome organization in CSF and its alterations in AD, WGCNA was conducted to construct an *N*-glycoproteomic network. Nine modules with similar expression patterns were grouped via the average linkage hierarchical clustering (**Figure 6B**). Module-trait association analysis was performed to determine the correlation relationships between each module eigenglycopeptide and AD-relevant phenotypic traits and other clinical or sample characteristics (**Figure S9A**). Our analysis identified one positively correlated module (pink module) and one negatively correlated module (red module) that showed significant association with AD status, amyloid-β pathology, total-tau (T-tau), and phospho-tau 181 (P-tau 181) level. Subsequently, GO enrichment analysis was performed on the two disease-associated network modules (**Figure 6C, 6D**) which uncovered that these two modules were closely

associated with brain development and extracellular function. In addition, the top 30 most connected hub glycopeptides in the two modules are plotted in **Figure S9B** and **S9C**.

Conclusions

Herein, we successfully developed a robust and highly sensitive strategy, Boost-DiLeu, for enhanced isobaric labeled quantitative intact glycoproteomic analysis. For the first time, we incorporated the one-tube sample processing workflow into DiLeu labeling experiments, which not only simplified experimental steps but also greatly reduced sample loss and increased proteome coverage. During labeling, we utilized a boosting channel consisting of relatively large amounts of biological samples analogous to samples of interest to enhance the MS signal, and this further increased glycoproteome coverage and ensured quantification performance. A boost/study ratio of 30× was found to be optimal in both identification and quantification accuracy. MS2 parameters including AGC settings and maximum injection times were optimized, and a trade-off between glycoproteome coverage and quantification accuracy was observed. As a compromise, 3E4 AGC and 200 ms injection time were suggested as a starting point. Additional HpH fractionation after conventional HILIC enrichment provided complementary orthogonality in separation, which further boosted glycoproteome coverage. As a proof-of-principle experiment, the Boost-DiLeu strategy was applied to the site-specific *N*-glycoproteome analysis of human CSF samples from AD patients and non-AD donors. Overall, 1172 quantifiable intact glycopeptides were identified from 164 glycoproteins, and 18 glycopeptides were found to be dysregulated. WGCNA revealed that two modules of glycopeptides significantly correlated with AD. These results demonstrated the feasibility of using this strategy for glycoproteomic analysis of size-limited clinical samples,

which would be essential for future biomarker discovery and the development of personalized medicine. Notably, while we developed the Boost-DiLeu strategy with the 12-plex DiLeu tag set in this study, a similar strategy may also be adapted to the 21-plex DiLeu tags for higher multiplex capacity without apparent isotopic interference from the boosting channel.[28] Future implementation of this strategy could also be extended to other DiLeu-labeled PTM analyses of size-limited biological samples, such as phosphorylation and citrullination in clinical specimens.

Acknowledgements

This work was supported, in part, by the National Institutes of Health Grants RF1AG052324, U01CA231081, and R01 DK071801 (to L.L.). The Orbitrap instruments were purchased through the support of an NIH Shared Instrument Grant (NIH-NCRR S10RR029531 to L.L.) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. L.L. would like to acknowledge NIH grants R21AG065728, and S10OD025084, as well as funding support from a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. H.Z. is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#681712), Swedish State Support for Clinical Research (#ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C and #ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor,

Hjärnfonden, Sweden (#FO2019-0228), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at UCL.

Reference

(1) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126*, 855–867.

(2) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.

(3) Lin, B.; Qing, X.; Liao, J.; Zhuo, K. Role of Protein Glycosylation in Host-Pathogen Interaction. *Cells* **2020**, *9* (4), 1022.

(4) Chen, Z.; Huang, J.; Li, L. Recent Advances in Mass Spectrometry (MS)-Based Glycoproteomics in Complex Biological Samples. *Trends Analyt Chem.* **2018**, *118* (Cell 126 2006), 880–892.

(5) Fang, P.; Xie, J.; Sang, S.; Zhang, L.; Liu, M.; Yang, L.; Xu, Y.; Yan, G.; Yao, J.; Gao, X.; Qian, W.; Wang, Z.; Zhang, Y.; Yang, P.; Shen, H. Multilayered N-Glycoproteome Profiling Reveals Highly Heterogeneous and Dysregulated Protein N-Glycosylation Related to Alzheimer's Disease. *Anal. Chem.* **2019**, *92* (1), 867–874.

(6) Hu, Y.; Pan, J.; Shah, P.; Ao, M.; Thomas, S. N.; Liu, Y.; Chen, L.; Schnaubelt, M.; Clark, D. J.; Rodriguez, H.; Boja, E. S.; Hiltke, T.; Kinsinger, C. R.; Rodland, K. D.; Li, Q. K.; Qian, J.; Zhang, Z.; Chan, D. W.; Zhang, H.; Consortium, C. P. T. A.; Pandey, A.; Paulovich, A.; Hoofnagle, A.; Zhang, B.; Mani, D. R.; Liebler, D. C.; Ransohoff, D. F.; Fenyo, D.; Tabb, D. L.; Levine, D. A.; Kuhn, E.; White, F. M.; Whiteley, G. A.; Zhu, H.; Shih, I.-M.; Bavarva, J.; McDermott, J. E.; Whiteaker, J.; Ketchum, K. A.; Clauser, K. R.; Ruggles, K.; Elburn, K.; Ding, L.; Hannick, L.; Zimmerman, L. J.; Watson, M.; Thiagarajan, M.; Ellis, M. J. C.; Oberti, M.; Mesri, M.; Sanders, M. E.; Borucki, M.; Gillette, M. A.; Snyder, M.; Edwards, N. J.; Vatanian, N.; Rudnick, P. A.; McGarvey, P. B.; Mertins, P.; Townsend, R. R.; Thangudu, R. R.; Smith, R. D.; Rivers, R. C.; Slebos, R. J. C.; Payne, S. H.; Davies, S. R.; Cai, S.; Stein, S. E.; Carr, S. A.; Skates, S. J.; Madhavan, S.; Liu, T.; Chen, X.; Zhao, Y.; Wang, Y.; Shi, Z. Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Rep.* **2020**, *33* (3), 108276.

(7) Tang, L.; Chen, X.; Zhang, X.; Guo, Y.; Su, J.; Zhang, J.; Peng, C.; Chen, X. N-Glycosylation in Progression of Skin Cancer. *Med. Oncol.* **2019**, *36* (6), 50.

(8) Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Capturing Site-Specific Heterogeneity with Large-Scale N-Glycoproteome Analysis. *Nat. Commun.* **2019**, *10* (1), 1–13.

(9) Nwosu, C. C.; Strum, J. S.; An, H. J.; Lebrilla, C. B. Enhanced Detection and Identification of Glycopeptides in Negative Ion Mode Mass Spectrometry. *Anal. Chem.* **2010**, *82* (23), 9654–9662.

(10) Hart-Smith, G.; Raftery, M. J. Detection and Characterization of Low Abundance Glycopeptides Via Higher-Energy C-Trap Dissociation and Orbitrap Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 124–140.

(11) 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.

(12) Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal. Chem.* **2010**, *82* (13), 5598–5609.

(13) Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X. Recent Advances in Hydrophilic Interaction Liquid Interaction Chromatography Materials for Glycopeptide Enrichment and Glycan Separation. *Trends Anal. Chem* **2020**, *124*, 115570.

(14) Wu, R.; Pai, A.; Liu, L.; Xing, S.; Lu, Y. NanoTPOT: Enhanced Sample Preparation for Quantitative Nanoproteomic Analysis. *Anal. Chem.* **2020**, *92* (9), 6235–6240.

(15) Yi, L.; Tsai, C.-F.; Dirice, E.; Swensen, A. C.; Chen, J.; Shi, T.; Gritsenko, M. A.; Chu, R. K.; Piehowski, P. D.; Smith, R. D.; Rodland, K. D.; Atkinson, M. A.; Mathews, C. E.; Kulkarni, R. N.; Liu, T.; Qian, W.-J. Boosting to Amplify Signal with Isobaric Labeling (BASIL) Strategy for Comprehensive Quantitative Phosphoproteomic Characterization of Small Populations of Cells. *Anal. Chem.* **2019**, *91* (9), 5794–5801.

(16) Slavov, N. Single-Cell Protein Analysis by Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2021**, *60*, 1–9.

(17) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: Mass Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity during Cell Differentiation. *Genome Biol.* **2018**, *19* (1), 161.

(18) Tsai, C.-F.; Zhao, R.; Williams, S. M.; Moore, R. J.; Schultz, K.; Chrisler, W. B.; Pasa-Tolic, L.; Rodland, K. D.; Smith, R. D.; Shi, T.; Zhu, Y.; Liu, T. An Improved Boosting to Amplify Signal with Isobaric Labeling (IBASIL) Strategy for Precise Quantitative Single-Cell Proteomics. *Mol. Cell Proteomics* **2020**, *19* (5), 828–838.

(19) Suttapitugsakul, S.; Tong, M.; Sun, F.; Wu, R. Enhancing Comprehensive Analysis of Secreted Glycoproteins from Cultured Cells without Serum Starvation. *Anal. Chem.* **2021**.

(20) Klann, K.; Tascher, G.; Münch, C. Functional Translatome Proteomics Reveal Converging and Dose-Dependent Regulation by MTORC1 and EIF2α. *Mol. Cell* **2020**, *77* (4), 913-925.e4.

(21) Justice, S. A. P.; McCracken, N. A.; Victorino, J. F.; Qi, G. D.; Wijeratne, A. B.; Mosley, A. L. Boosting Detection of Low-Abundance Proteins in Thermal Proteome Profiling Experiments by Addition of an Isobaric Trigger Channel to TMT Multiplexes. *Anal. Chem.* **2021**, *93* (18), 7000–7010.

(22) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Mohammed, A.; 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.

(23) Ross, P.; Huang, Y.; Marchese, J.; 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. Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169.

(24) Frost, D. C.; Greer, T.; Li, L. High-Resolution Enabled 12-Plex DiLeu Isobaric Tags for Quantitative Proteomics. *Anal. Chem.* **2015**, *87* (3), 1646–1654.

(25) 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* (7), 2817–2825.

(26) Sleno, L. The Use of Mass Defect in Modern Mass Spectrometry. *J. Mass Spectrom.* **2012**, *47* (2), 226–236.

(27) Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. Neutron-Encoded Mass Signatures for Multiplexed Proteome Quantification. *Nat. Methods* **2013**, *10* (4), 332–334.

(28) Frost, D. C.; Feng, Y.; Li, L. 21-Plex DiLeu Isobaric Tags for High-Throughput Quantitative Proteomics. *Anal. Chem.* **2020**, *92* (12), 8228–8234.

(29) Dieterich, I. A.; Cui, Y.; Braun, M. M.; Lawton, A. J.; Robinson, N. H.; Peotter, J. L.; Yu, Q.; Casler, J. C.; Glick, B. S.; Audhya, A.; Denu, J. M.; Li, L.; Puglielli, L. Acetyl-CoA Flux from the Cytosol to the ER Regulates Engagement and Quality of the Secretory Pathway. *Sci. Rep.* **2021**, *11* (1), 2013.

(30) Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L. Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2019**, 1–11.

(31) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13* (9), 731–740.

(32) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yılmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (Database issue), D442–D450.

(33) Wu, F.; Sun, D.; Wang, N.; Gong, Y.; Li, L. Comparison of Surfactant-Assisted Shotgun Methods Using Acid-Labile Surfactants and Sodium Dodecyl Sulfate for Membrane Proteome Analysis. *Anal. Chim. Acta* **2011**, *698* (1–2), 36–43.

(34) Randall, S. M.; Cardasis, H. L.; Muddiman, D. C. Factorial Experimental Designs Elucidate Significant Variables Affecting Data Acquisition on a Quadrupole Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (10), 1501–1512.

(35) Specht, H.; Slavov, N. Optimizing Accuracy and Depth of Protein Quantification in Experiments Using Isobaric Carriers. *J. Proteome Res.* **2020**, *20* (1), 880–887.

(36) Dabrowski, R.; Ripa, R.; Latza, C.; Annibal, A.; Antebi, A. Optimization of Mass Spectrometry Settings for Steroidomic Analysis in Young and Old Killifish. *Anal. Bioanal. Chem.* **2020**, *412* (17), 4089–4099.

(37) Yang, F.; Shen, Y.; Camp, D. G.; Smith, R. D. High-PH Reversed-Phase Chromatography with Fraction Concatenation for 2D Proteomic Analysis. *Expert Rev. Proteom* **2014**, *9* (2), 129–134.

(38) Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

(39) Chen, Z.; Yu, Q.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific Analysis of N-Glycoproteome in Human Cerebrospinal Fluid and Glycosylation Landscape Changes in Alzheimer's Disease. *Mol. Cell Proteomics* **2021**, *20*, 100081.

(40) Fang, P.; Ji, Y.; Silbern, I.; Doebele, C.; Ninov, M.; Lenz, C.; Oellerich, T.; Pan, K.-T.; Urlaub, H. A Streamlined Pipeline for Multiplexed Quantitative Site-Specific N-Glycoproteomics. *Nat. Commun.* **2020**, *11* (1), 5268.

(41) Song, C.; Ye, M.; Han, G.; Jiang, X.; Wang, F.; Yu, Z.; Chen, R.; Zou, H. Reversed-Phase-Reversed-Phase Liquid Chromatography Approach with High Orthogonality for Multidimensional Separation of Phosphopeptides. *Anal. Chem.* **2010**, *82* (1), 53–56.

(42) Cheung, T. K.; Lee, C.-Y.; Bayer, F. P.; McCoy, A.; Kuster, B.; Rose, C. M. Defining the Carrier Proteome Limit for Single-Cell Proteomics. *Nat. Methods* **2020**, 1–8.

(43) Huang, J.; Dong, J.; Shi, X.; Chen, Z.; Cui, Y.; Liu, X.; Ye, M.; Li, L. Dual-Functional Titanium(IV) Immobilized Metal Affinity Chromatography Approach for Enabling Large-Scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. *Anal. Chem.* **2019**, *91* (18), 11589–11597.

(44) Segal, M. B. Extracellular and Cerebrospinal Fluids. *J. Inherit. Metab. Dis.* **1993**, *16* (4), 617–638.

(45) Abdi, F.; Quinn, J. F.; Jankovic, J.; McIntosh, M.; Leverenz, J. B.; Peskind, E.; Nixon, R.; Nutt, J.; Chung, K.; Zabetian, C.; Samii, A.; Lin, M.; Hattan, S.; Pan, C.; Wang, Y.; Jin, J.; Zhu, D.; Li, G. J.; Liu, Y.; Waichunas, D.; Montine, T. J.; Zhang, J. 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.

(46) Zhang, J. Proteomics of Human Cerebrospinal Fluid – the Good, the Bad, and the Ugly. *Proteom. - Clin. Appl* **2007**, *1* (8), 805–819.

(47) Chen, Z.; Wang, D.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Yu, Q.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients. *Acs Chem Biol* **2021**.

(48) Shores, K. S.; Knapp, D. R. Assessment Approach for Evaluating High Abundance Protein Depletion Methods for Cerebrospinal Fluid (CSF) Proteomic Analysis. *J. Proteome Res.* **2007**, *6* (9), 3739–3751.

(49) Yang, K.; Yang, Z.; Chen, X.; Li, W. The Significance of Sialylation on the Pathogenesis of Alzheimer's Disease. *Brain Res. Bull.* **2021**, *173*, 116–123.

(50) Evin, G.; Li, Q.-X. Platelets and Alzheimer's Disease: Potential of APP as a Biomarker. *World J. Psychiatry* **2012**, *2* (6), 102–113.

(51) Knebl, J.; DeFazio, P.; Clearfield, M. B.; Little, L.; McConathy, W. J.; Pherson, R. M.; Lacko, A. G. Plasma Lipids and Cholesterol Esterification in Alzheimer's Disease. *Mech. Ageing Dev.* **1994**, *73* (1), 69–77.

(52) Zhong, X.; Wang, J.; Carlsson, C.; Okonkwo, O.; Zetterberg, H.; Li, L. A Strategy for Discovery and Verification of Candidate Biomarkers in Cerebrospinal Fluid of Preclinical Alzheimer's Disease. *Front. Mol. Neurosci.* **2019**, *11*, 483.

(53) McGeer, P. L.; McGeer, E. G. The Possible Role of Complement Activation in Alzheimer Disease. *Trends Mol. Med.* **2002**, *8* (11), 519–523.

(54) 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.

(55) Zhong, X.; Yu, Q.; Ma, F.; Frost, D. C.; Lu, L.; Chen, Z.; Zetterberg, H.; Carlsson, C.; Okonkwo, O.; Li, L. HOTMAQ: A Multiplexed Absolute Quantification Method for Targeted Proteomics. *Anal. Chem.* **2019**, *91* (3), 2112–2119.

(56) Donner, L.; Fälker, K.; Gremer, L.; Klinker, S.; Pagani, G.; Ljungberg, L. U.; Lothmann, K.; Rizzi, F.; Schaller, M.; Gohlke, H.; Willbold, D.; Grenegard, M.; Elvers, M. Platelets Contribute to Amyloid-β Aggregation in Cerebral Vessels through Integrin AIIbβ3–Induced Outside-in Signaling and Clusterin Release. *Sci. Signal.* **2016**, *9* (429), ra52–ra52.

(57) Sackmann, C.; Hallbeck, M. Oligomeric Amyloid-β Induces Early and Widespread Changes to the Proteome in Human IPSC-Derived Neurons. *Sci. Rep.* **2020**, *10* (1), 6538.

(58) PASCALE, A.; ETCHEBERRIGARAY, R. CALCIUM ALTERATIONS IN ALZHEIMER'S DISEASE: PATHOPHYSIOLOGY, MODELS AND THERAPEUTIC OPPORTUNITIES. *Pharmacol. Res.* **1999**, *39* (2), 81–88.

(59) Řípová, D.; Platilová, V.; Strunecká, A.; Jirák, R.; Höschl, C. Cytosolic Calcium Alterations in Platelets of Patients with Early Stages of Alzheimer's Disease. *Neurobiol. Aging* **2000**, *21* (5), 729–734.

(60) Chen, C. P. C.; Huang, Y.-C.; Chang, C.-N.; Chen, J.-L.; Hsu, C.-C.; Lin, W.-Y. Changes of Cerebrospinal Fluid Protein Concentrations and Gait Patterns in Geriatric Normal Pressure Hydrocephalus Patients after Ventriculoperitoneal Shunting Surgery. *Exp. Gerontol.* **2018**, *106*, 109–115.

(61) Domenico, F. D.; Pupo, G.; Giraldo, E.; Badìa, M.-C.; Monllor, P.; Lloret, A.; Schininà, M. E.; Giorgi, A.; Cini, C.; Tramutola, A.; Butterfield, D. A.; Viña, J.; Perluigi, M. Oxidative Signature of Cerebrospinal Fluid from Mild Cognitive Impairment and Alzheimer Disease Patients. *Free Radic. Biol. Med* **2016**, *91*, 1–9.

Figures

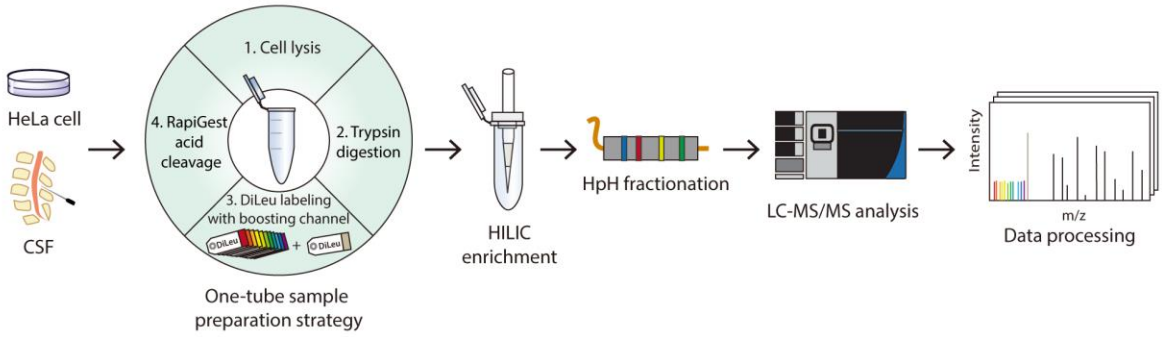


Figure 1. Workflow of Boost-DiLeu strategy. Proteins from biological samples were extracted, enzymatically digested, and labeled through a one-tube sample preparation workflow. DiLeu 118d was used as a boosting channel. Samples were pooled after labeling for HILIC enrichment and HpH fractionation, followed by the LC-MS/MS analysis.

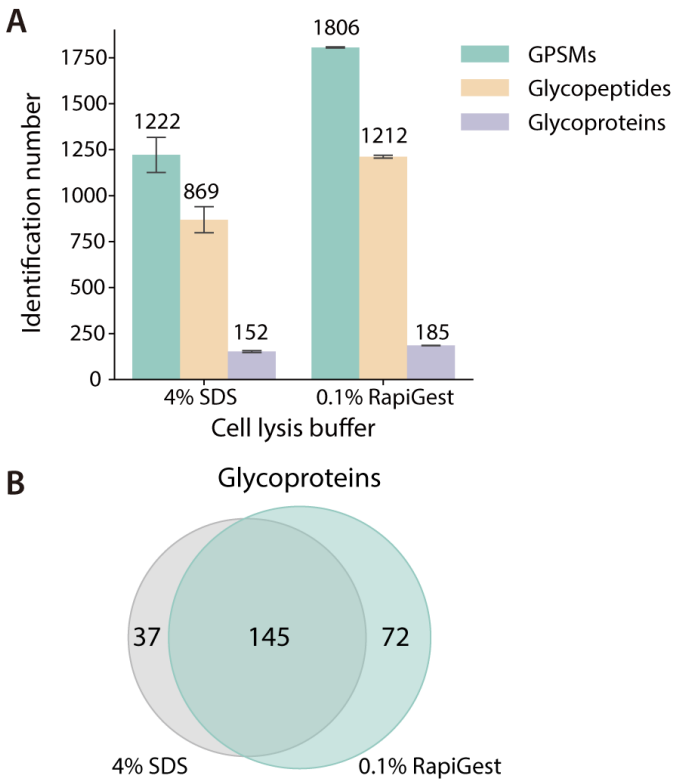


Figure 2. Comparison of sample preparation using conventional and one-tube sample processing. (A) Identification number in glycoproteomic analysis. (B) Overlap of glycoproteins by the two methods, as shown using the Venn diagram.

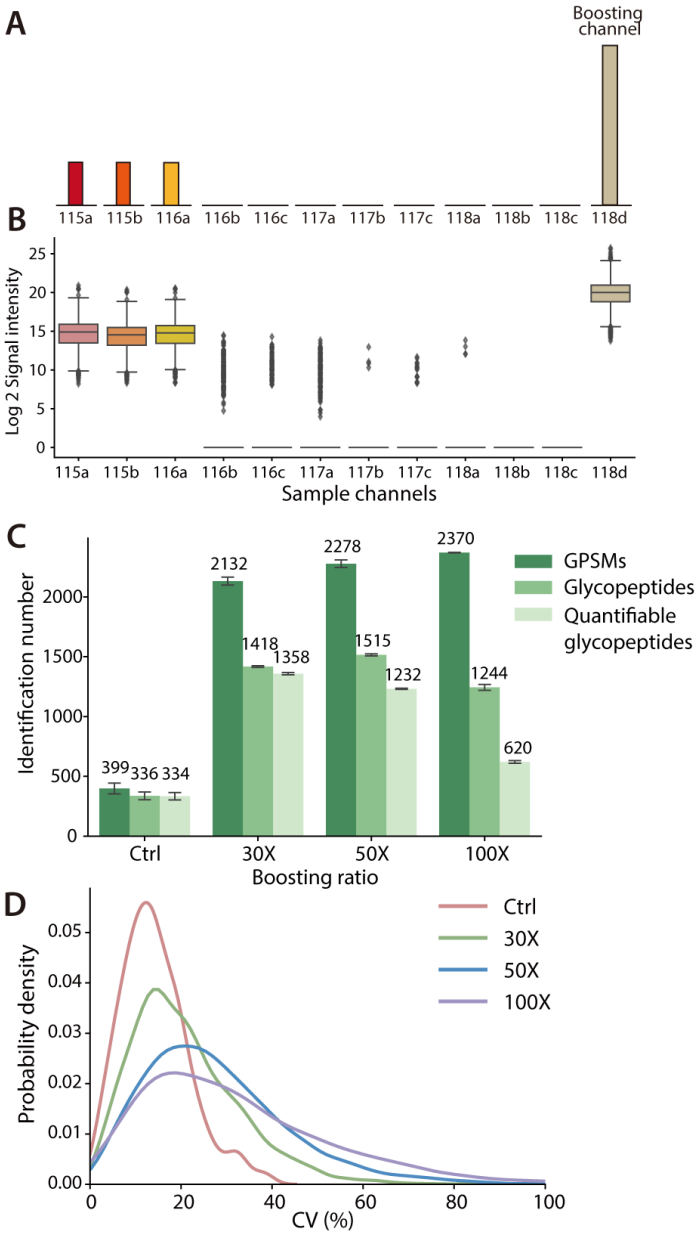


Figure 3. Comparison of different B/S ratios. (A) Experimental design: first three channels were labeled as study channels and DiLeu 118d channel was used as a boosting channel. (B) Reporter ion signal intensities at a 30× B/S ratio. (C) Identification number of GPSMs, total glycopeptides and quantifiable glycopeptides at different B/S ratios. (D) CV distribution of quantifiable glycopeptides.

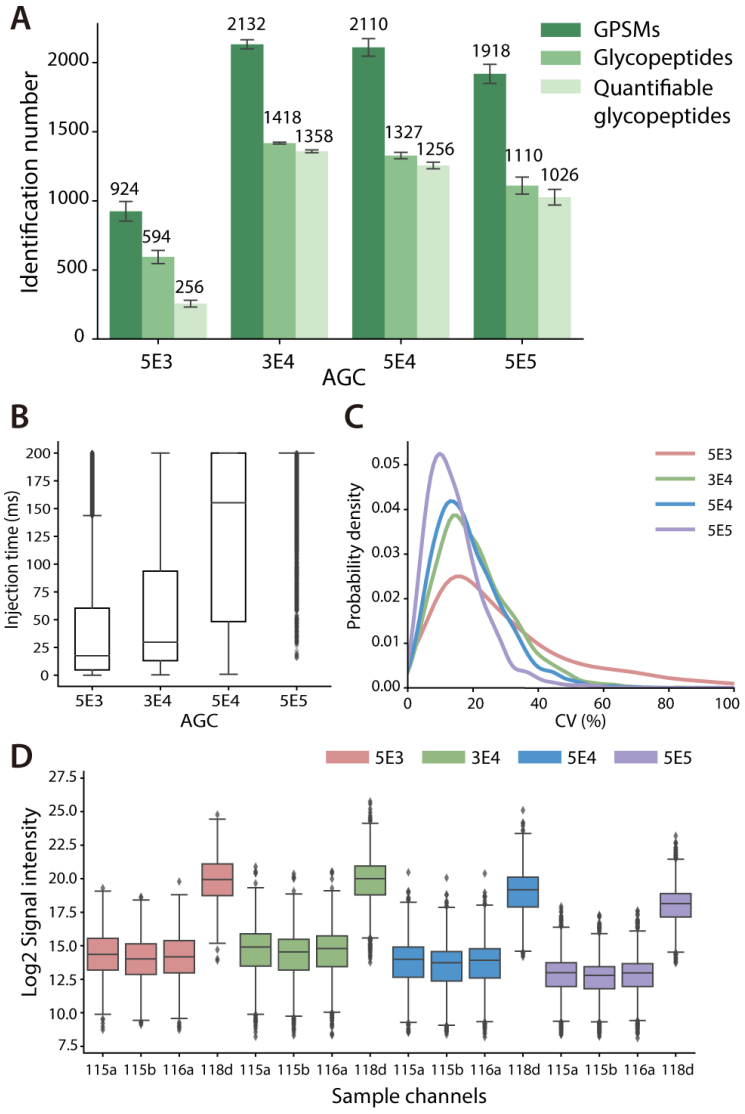


Figure 4. Comparison of different AGC settings. (A) Identification number of GPSMs, total glycopeptides, and quantifiable glycopeptides with different AGCs. (B) Distribution of the actual ion injection time. (C) CV distribution of quantifiable glycopeptides. (D) Distribution of reporter ion signal intensities.

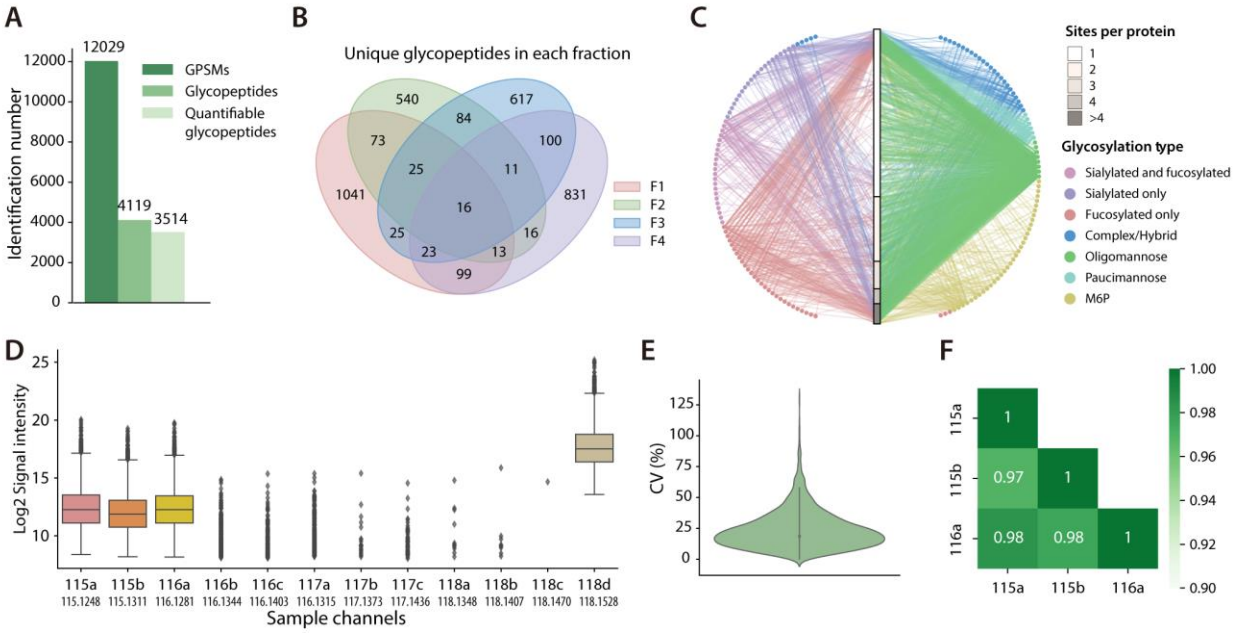


Figure 5. Global glycoproteome mapping in HeLa cell line. (A) Identification number of GPSMs, total glycopeptides and quantifiable glycopeptides. (B) Overlap of unique glycopeptides identified in each fraction. (C) A glycoprotein-glycan network diagram that maps which glycans (outer nodes) modify which proteins (inner bar). Glycoproteins are sorted by the number of glycosites. Glycan nodes and their linkage to the proteins are colored according to the glycosylation type. (D) Distribution of reporter ion signal intensities across 12 channels. (E) CV distribution of quantifiable glycopeptides. (F) Pearson correlation of reporter ion intensities between three study channels.

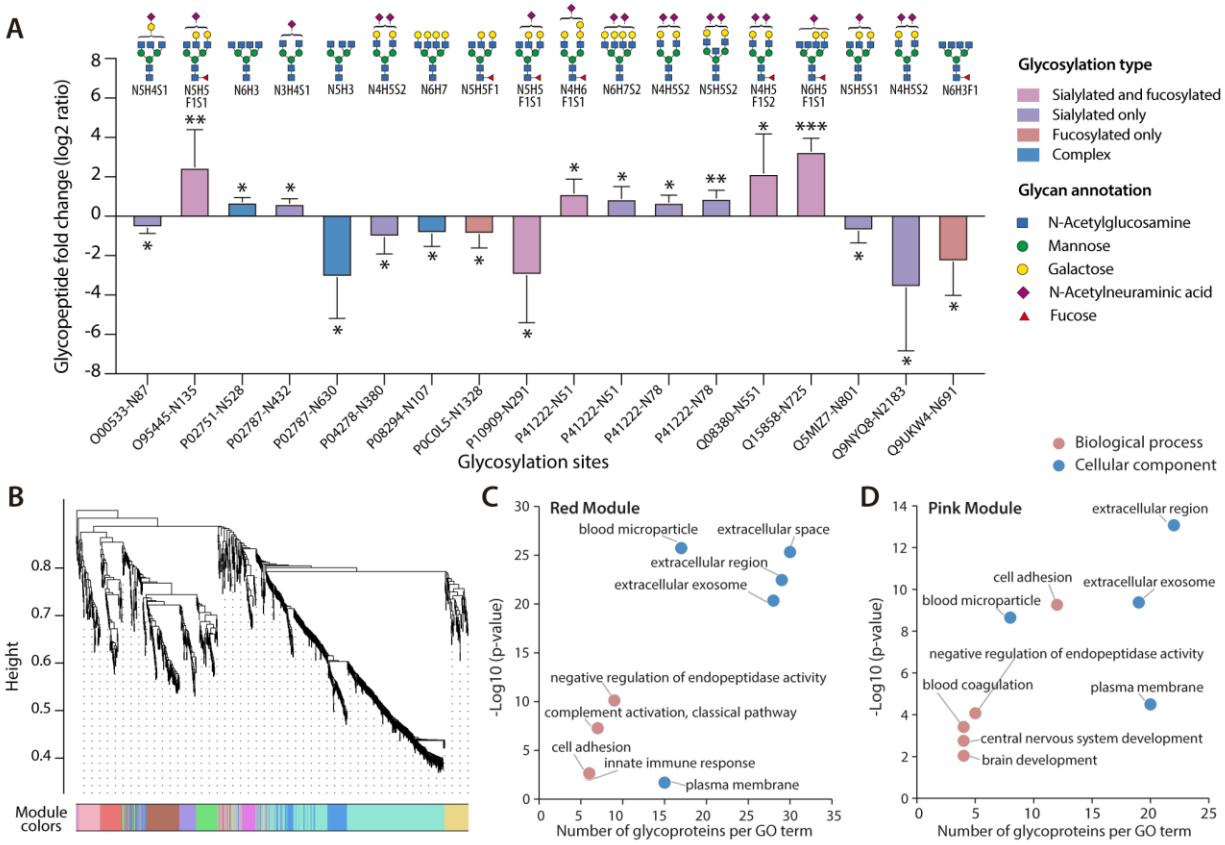


Figure 6. Site-specific quantitative glycoproteomic analysis of human CSF samples. (A) 18 aberrant *N*-glycopeptides in AD CSF samples with site-specific information (*p < 0.05, **p < 0.01, and ***p < 0.001). Each glycan structure only represents one possible glycan candidate due to the lack of structural information (N: HexNAc, H: Hex, F: fucose, S: NeuAc). (B) WGCNA cluster dendrogram of 1172 glycopeptides. Nine modules with similar expression patterns were grouped via the average linkage hierarchical clustering. (C) GO analysis of the red module glycopeptides clustered in WGCNA dendrogram. (D) GO analysis of the pink module glycopeptides clustered in WGCNA dendrogram.

Supplemental Information

Experimental section

Chemicals and Materials. Dithiothreitol (DTT), sequencing grade trypsin were from Promega (Madison, WI). Optimal LC/MS grade solvents, Tris base, urea, and sodium chloride were from Fisher Scientific (Pittsburgh, PA). Trifluoroacetic acid (TFA), iodoacetamide (IAA), triethylammonium bicarbonate (TEAB), sodium dodecyl sulfate (SDS), *N*,*N*-dimethylformamide (DMF), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium tetrafluoroborate (DMTMM), and dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St Louis, MO). N-methylmorpholine (NMM) was purchased from TCI America (Tokyo, Japan). Oasis HLB 1 cc (10 mg) extraction cartridges and RapiGest SF (RapiGest) were purchased from Waters Corporation (Milford, MA). Hydroxylamine solution was purchased from Alfa Aesar (Ward Hill, MA). Strong anion exchange (SAX) bulk material PolySAX LP (12 µm, 300 Å) and strong cation exchange (SCX) spin tips were purchased from PolyLC (Columbia, MD). Empty TopTips were from Glygen Corp (Columbia, MD). Protease inhibitor cocktail tablets and phosphatase inhibitor cocktail tablets were from Roche (Mannheim, Germany).

Cell Culture. HeLa cells were cultured in DMEM (Hyclone) supplemented with 10% fetal bovine serum (Gibco), 1% penicillin/streptomycin (SigmaAldrich), and incubated at 37 °C in a humidified chamber with 5% CO_2. Cells were harvested by washing with PBS buffer (Gibco, pH 7.4)

HeLa Cell Sample Preparation. HeLa cell pellet was lysed in extraction buffer consisting of 50 mM Tris base (pH 7.4), 4% SDS, 65 mM DTT, 175 mM NaCl, and 1% (v/v) protease and phosphatase inhibitor cocktail. Lysates were sonicated with a probe sonicator in an ice water bath at 50% power with pulse 5s on 5s off for 12 cycles. Cell lysates were centrifuged at 18000 g for 5

min, and the supernatant was collected and poured into ice-cold precipitation buffer (acetone: ethanol: acetic acid=50: 50: 0.1). The ratio of lysate to precipitation was 1:5. The precipitation went overnight at -20^{o}C. Protein pellets were collected through centrifugation at 18000 g for 10 min and washed twice with ice-cold precipitation buffer. The pellets were dried in the fume hood for 10 min and redissolved in 8M urea/50 mM TEAB buffer (pH 8.0). Protein concentrations were measured by a BCA assay kit (Thermo Fisher Scientific, San Jose, CA). Protein extracts were reduced with 5 mM DTT at 37 ^{o}C for 2h, followed by alkylation with 15 mM IAA at room temperature at dark for 30 min. Alkylation was quenched by adding 5 mM DTT for another 10 min. The urea buffer was diluted to 1.6 M with 50 mM TEAB buffer. The proteins were first digested with trypsin at a protein to enzyme ratio at 100:1 and incubated at 37 ^{o}C. After 12 h, the same amount of trypsin was added again and incubated at 37 ^{o}C for another 4 h to make a final protein to enzyme ratio of 50:1. The digestion was quenched by adding TFA to a final concentration of 1%, and peptides were then desalted by Oasis HLB cartridges. Peptides were dried down *in vacuo*.

CSF Sample Information. CSF samples were collected by lumbar puncture from five male AD patients and five male non-AD donors. The samples were from patients who sought medical advice because of cognitive impairment. Patients were designated as AD and non-AD according to CSF biomarker levels using cutoffs that are >90% specific for AD[1]: total-tau (T-tau) >350 pg/mL, phospho-tau 181 (P-tau181) >60 pg/mL and Aβ42 <530 pg/mL (2 out 3 positive). None of the biochemically normal subjects fulfilled these criteria. Detailed subjects' information is provided

in Table S1. The study was approved by the regional ethics committee at the University of Gothenburg. Protein concentration was determined by a BCA protein assay.

Offline HpH Fractionation. HpH fractionation was performed on a Waters Alliance e2695 HPLC using a 150 mm x 2.1 mm, 5mm, 100 Å, C18 column (Phenomenex) operating at 0.2 mL/min. Mobile phase A was 10 mM ammonium formate in water (pH 10) and mobile phase B was 10 mM ammonium formate in 90% acetonitrile (ACN) (pH 10). Separation was performed with the following gradient: 1% B (0–5 min), 1–40% B (5–50 min), 40–60% B (50–54 min), 60–70% B (54–58 min), and 70–100% B (58–59 min). Fractions were collected every 4 minutes and 14 fractions were collected in total. Nonadjacent fractions were concatenated into 4 tubes for MS analysis.

CSF Global Proteomic Analysis. 10 µg proteins were aliquoted for each study channel and 300 µg proteins pooled from both AD patients and non-AD donors CSF samples at a ratio of 3:5 were used as boosting channel. The aliquots were dried down in vacuo and treated with the one-tube sample preparation approach. After DiLeu labeling, the samples were pooled and went through SCX spin-tip clean-up according to manufacturer's protocols and HpH fractionation. Samples were analyzed on the Orbitrap Q-Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to a Dionex UPLC system. Peptides were separated on BEH C18 column using a linear gradient from 0% to 30% ACN (0.1% FA). Top 20 precursors were selected for MS/MS data acquisition. Survey scans of peptide precursors from m/z 300 to 1500 were performed at a resolving power of 120 K and AGC target of 1E6 with a maximum injection time of 100 ms. The precursors were fragmented by HCD with NCE of 30 and ± 3% collision energy.

Tandem MS acquisition was at resolving power of 60 K, a lower mass limit of 110 m/z, dynamic exclusion of 45 s, AGC of 1E5 and maximum injection time of 200 ms.

Expanded Data Analysis. For global proteomic analysis, raw files were searched by the same database and parameters as the glycoproteomic dataset by using Byonic and PD 2.1, except that *N*-glycosylation was not included as a dynamic modification. Proteins were filtered at a 1% protein FDR. Gene ontology analysis was conducted using Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8.[2] Protein-protein interaction network was generated by Metascape and Cytoscape.[3,4] WGCNA algorithm was used for network analysis.[5] A total of 1172 glycopeptides (Table S4) were screened and no outlier samples were excluded. The power of $\beta = 6$ (scale-free $R^2 = 0.9$) was selected as the soft-thresholding parameter to ensure a scale-free network.

Reference

(1) Hansson, O.; Zetterberg, H.; Buchhave, P.; Londos, E.; Blennow, K.; Minthon, L. Association between CSF Biomarkers and Incipient Alzheimer's Disease in Patients with Mild Cognitive Impairment: A Follow-up Study. *Lancet Neurol.* **2006**, *5* (3), 228–234.

(2) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* **2009**, *4* (1), 44–57.

(3) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* **2019**, *10* (1), 1523.

(4) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504.

(5) McKenzie, A. T.; Moyon, S.; Wang, M.; Katsyv, I.; Song, W.-M.; Zhou, X.; Dammer, E. B.; Duong, D. M.; Aaker, J.; Zhao, Y.; Beckmann, N.; Wang, P.; Zhu, J.; Lah, J. J.; Seyfried, N. T.; Levey, A. I.; Katsel, P.; Haroutunian, V.; Schadt, E. E.; Popko, B.; Casaccia, P.; Zhang, B. Multiscale Network Modeling of Oligodendrocytes Reveals Molecular Components of Myelin Dysregulation in Alzheimer's Disease. *Mol. Neurodegener.* **2017**, *12* (1), 82.

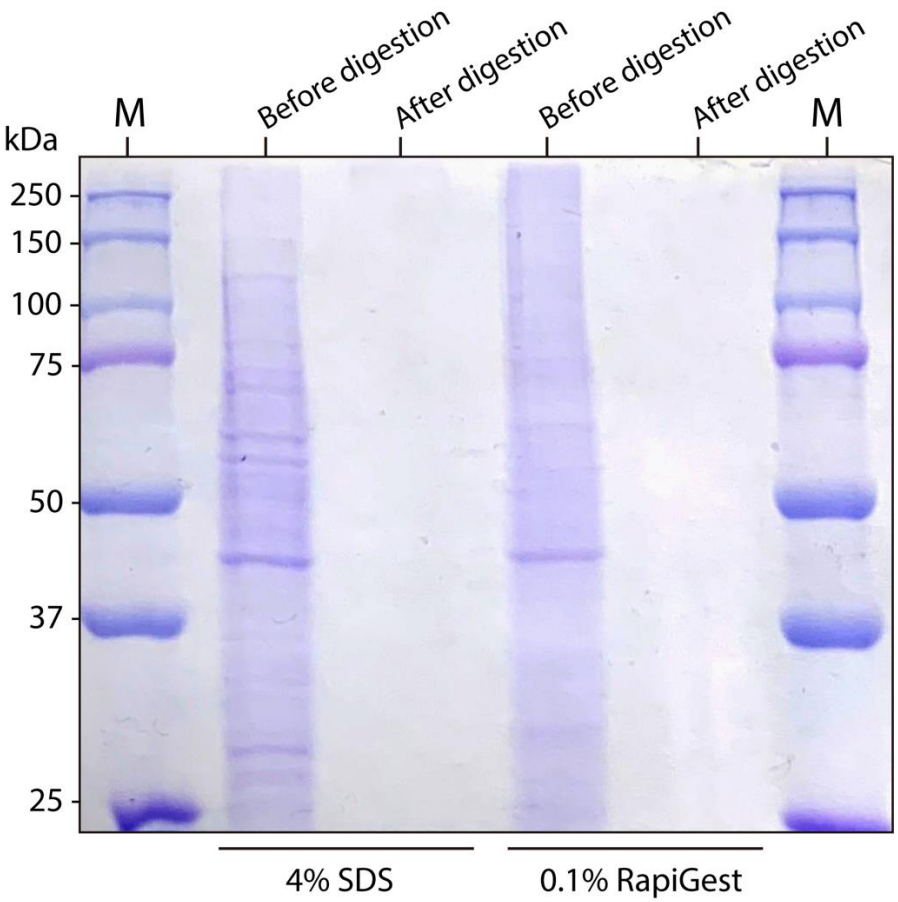


Figure S1. SDS-PAGE of proteins before and after trypsin digestion in two sample preparation workflows.

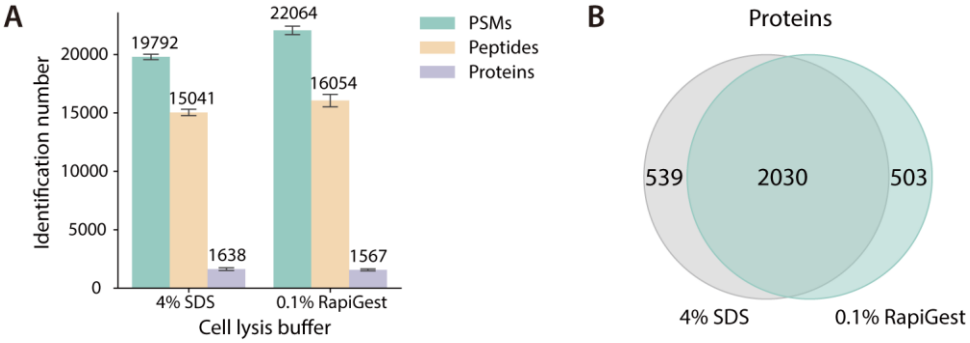


Figure S2. Comparison of sample preparation using conventional and one-tube sample processing workflow. (A) Identification number in proteomic analysis. (B) Venn diagram showing the overlap of identified proteins.

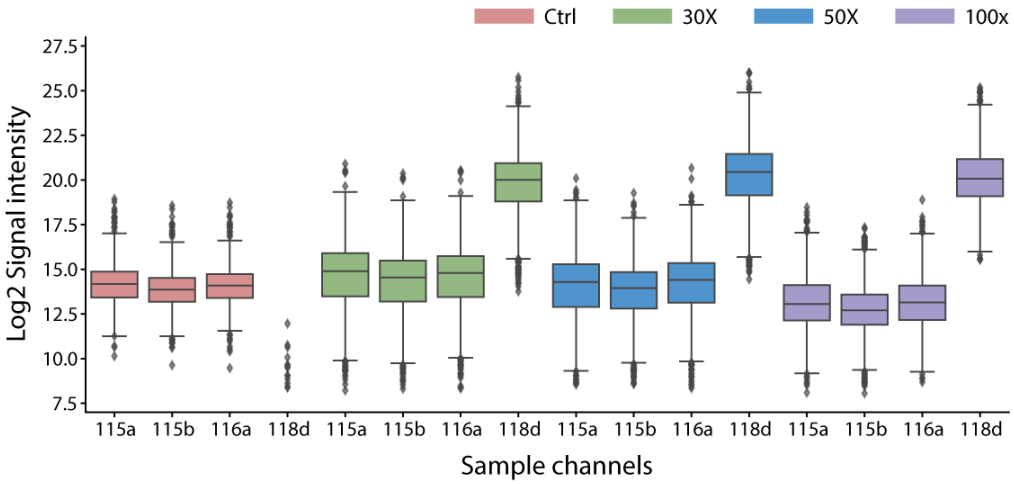


Figure S3. Distribution of reporter ion signal intensities at different B/S ratios.

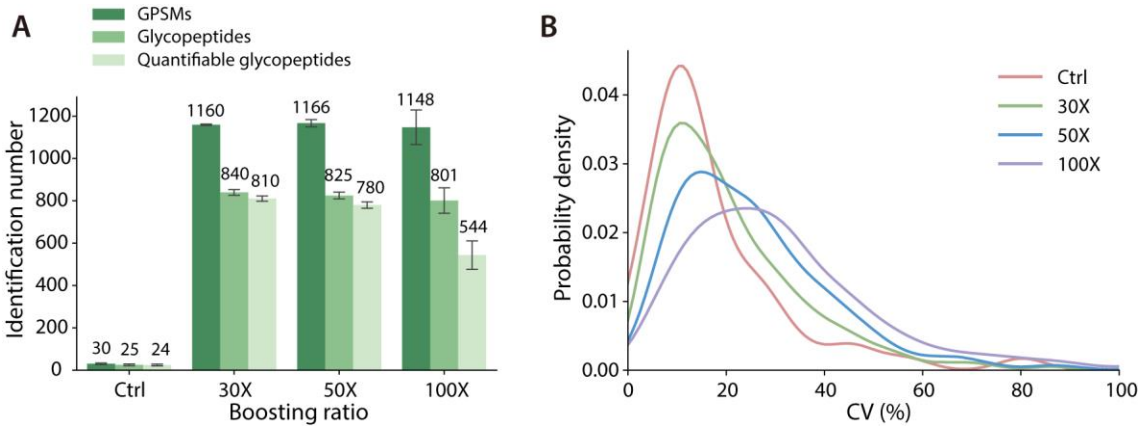


Figure S4. Comparison of different B/S ratios with the sample prepared in the conventional workflow using 4% SDS lysis buffer. (A) Identification number of GPSMs, total glycopeptides and quantifiable glycopeptides at different boosting ratios. (B) CV distribution of quantifiable glycopeptides.

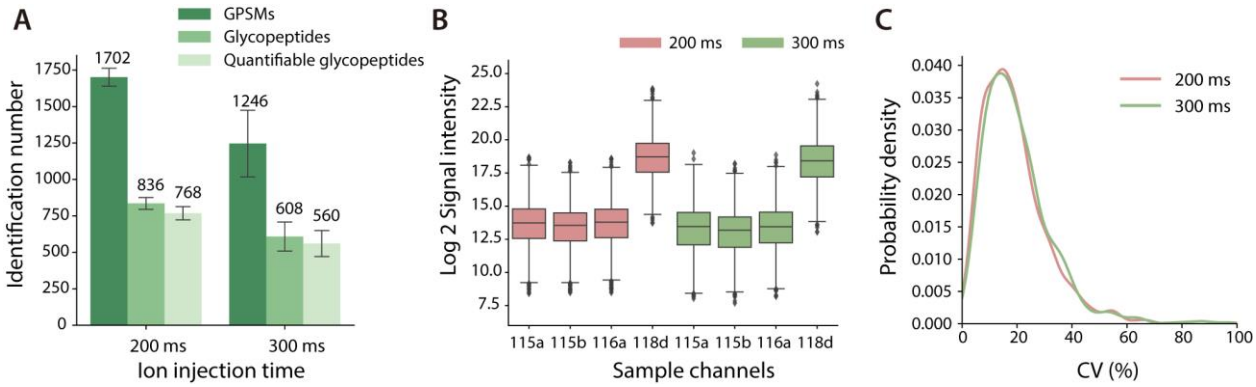


Figure S5. Comparison of different ion injection time (IT) settings. (A) Identification number of GPSMs, total glycopeptides and quantifiable glycopeptides with different IT. (B) Distribution of reporter ion signal intensities. (C) CV distribution of quantifiable glycopeptides.

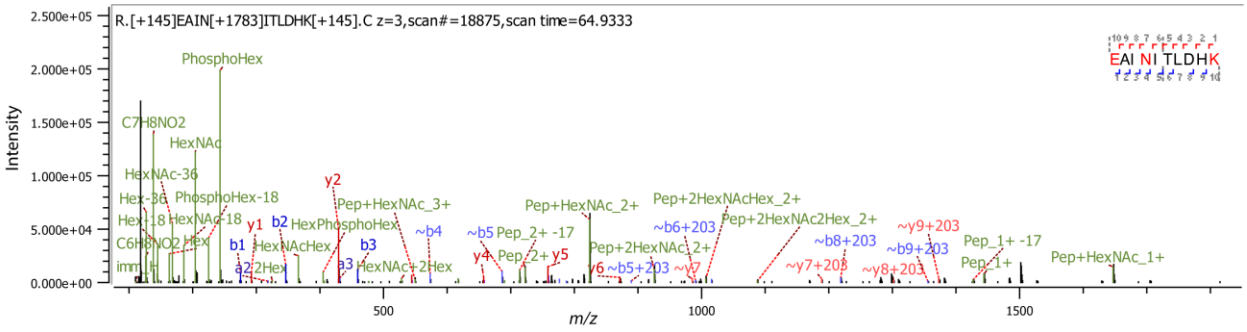


Figure S6. Representative MS/MS spectrum of an M6P glycopeptide in HeLa Cell digests. The peptide is identified as EAINITLDHK with HexNAc(2)Hex(8)Phospho(1) on N4. The existence of the M6P glycan is evidenced by the phospho-oxonium ions PhosphoHex at m/z 243, PhosphoHex -H_2O at m/z 225, and HexPhosphoHex at m/z 405.

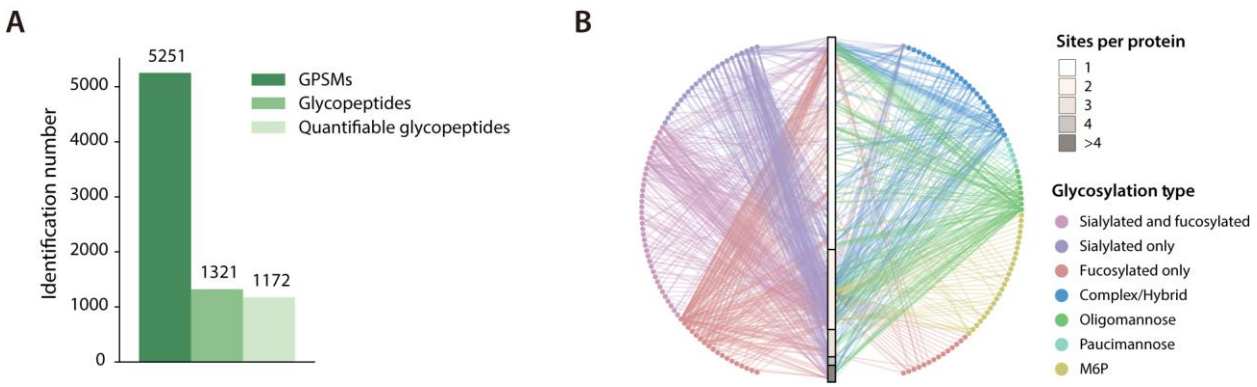


Figure S7. Site-specific glycoproteomic mapping of human CSF samples. (A) Identification number of GPSMs, total glycopeptides and quantifiable glycopeptides. (B) A glycoprotein-glycan network maps which glycans (outer nodes) modify which proteins (inner bar). Glycoproteins are sorted by the number of glycosites.

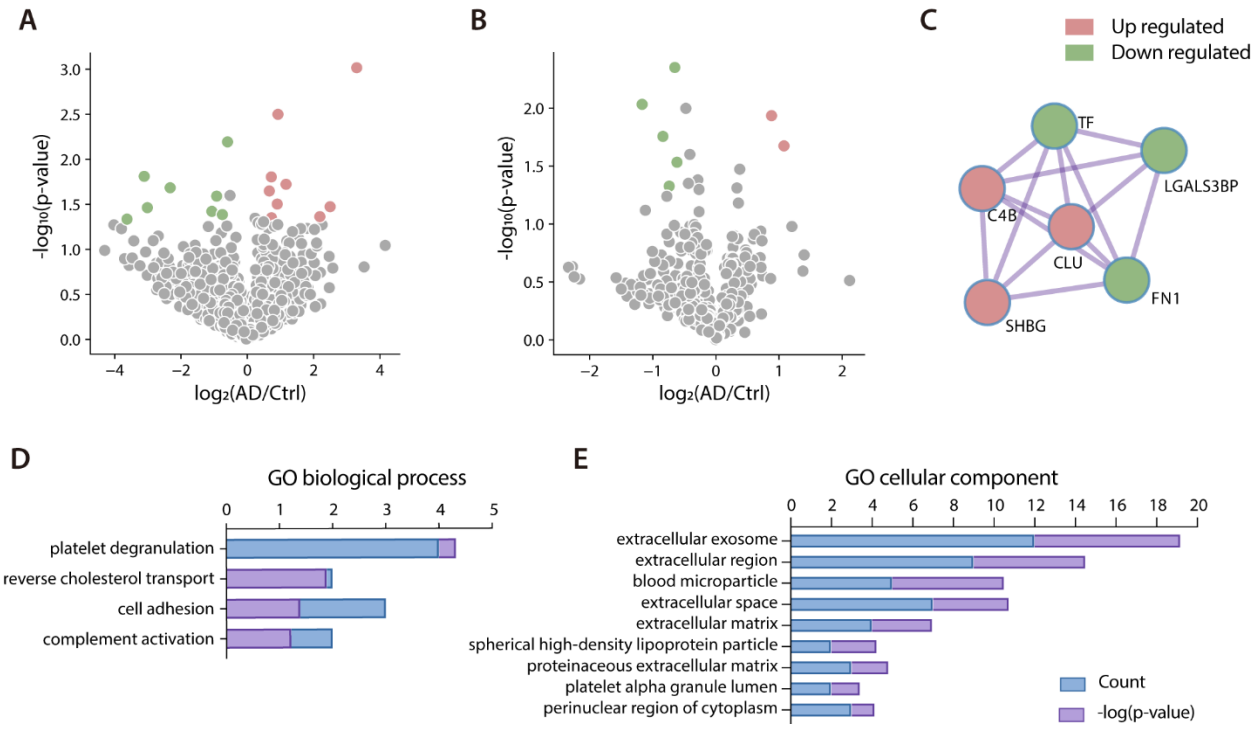


Figure S8. Site-specific quantitative glycoproteomic analysis of human CSF samples. (A) Volcano plot displays dysregulated intact glycopeptides. Student's t-test, $p < 0.05$, fold-change > 1.5. (B) Volcano plot displays dysregulated proteins in a parallel proteomic analysis. Student's t-test, $p < 0.05$, |fold-change| > 1.5. (C) Protein-protein interaction network of dysregulated glycoproteins in AD. (D) GO biological process analysis of dysregulated glycoproteins in AD. (E) GO cellular component analysis of dysregulated glycoproteins in AD.

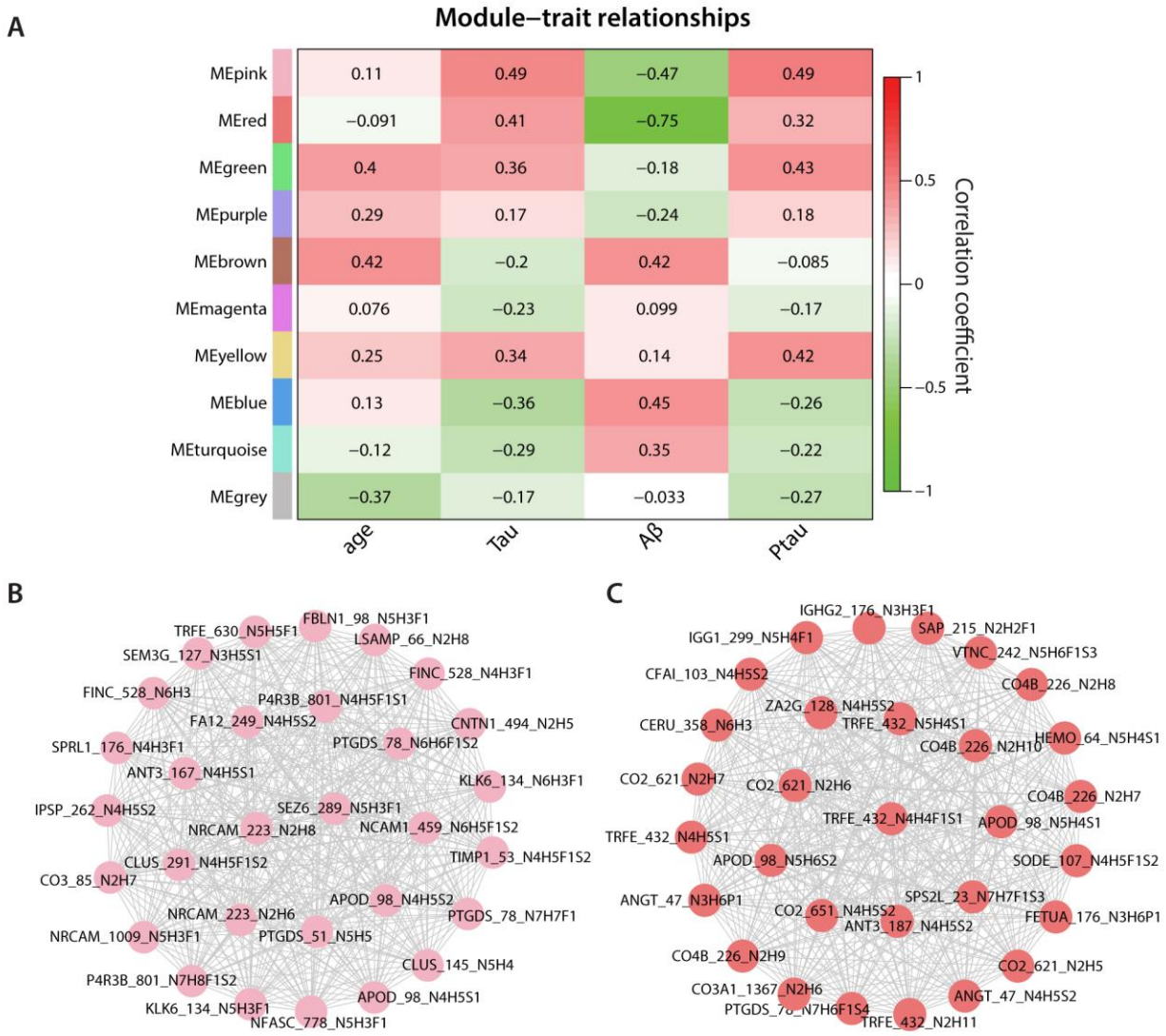


Figure S9. Site-specific quantitative glycoproteomic analysis of human CSF samples. (A) Identification of disease-relevant glycopeptide modules associated with AD phenotypic traits. Module-trait relationships were determined by biweight midcorrelation between module eigenglycopeptide expression and the indicated clinical or neuropathological feature. Correlation coefficients are indicated on the top of each cell. (B) N-glycopeptide co-regulation network plots of AD-associated pink module with top 30 hub glycopeptides. (C) N-glycopeptide co-regulation network plots of AD-associated red module with top 30 hub glycopeptides.

Table S1. CSF subject information.

LID-nr	age	gender	T-tau	Aβ	P-tau 181	AD stage	DiLeu labeling channel
3382	84	m	667	429	93	AD	115a
3625	70	m	699	205	82	AD	116a
3699	75	m	777	382	100	AD	117a
3751	73	m	625	564	76	AD	117c
3600	77	m	781	274	94	AD	118c
3451	73	m	241	996	34	non-AD	116b
3496	75	m	300	1250	48	non-AD	116c
3509	67	m	359	744	44	non-AD	117b
3628	75	m	382	830	58	non-AD	118a
3594	72	m	152	656	21	non-AD	118b

Table S2. Glycoproteins from two sample preparation methods.

https://pubs.acs.org/doi/suppl/10.1021/acs.analchem.2c01773/suppl_file/ac2c01773_si_003.xlsx

Table S3. Quantifiable glycopeptides in HeLa cells.

https://pubs.acs.org/doi/suppl/10.1021/acs.analchem.2c01773/suppl_file/ac2c01773_si_004.xlsx

Table S4. Quantifiable glycopeptides in CSF samples.

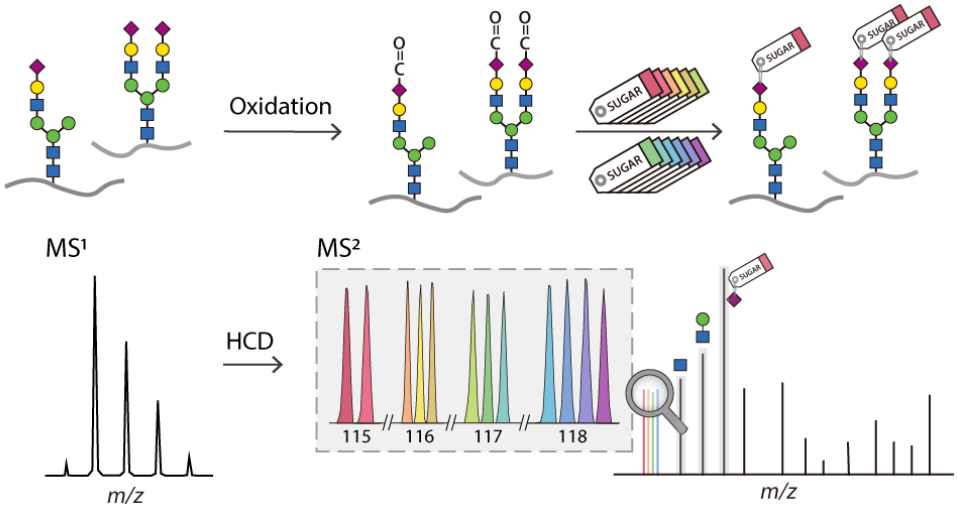
https://pubs.acs.org/doi/suppl/10.1021/acs.analchem.2c01773/suppl_file/ac2c01773_si_005.xlsx

Table S5. GO analysis of red and pink modules from WGCNA analysis.

https://pubs.acs.org/doi/suppl/10.1021/acs.analchem.2c01773/suppl_file/ac2c01773_si_006.xlsx

Chapter 7

High-throughput Quantification of Intact Sialylated Glycopeptides Enabled by 12-plex SUGAR Isobaric Tags



Adapted from: **Wang, D.**[#]; Li, M.[#]; Wang, Z.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, L. High-throughput Quantification of Intact Sialylated Glycopeptides Enabled by 12-plex SUGAR Isobaric Tags. (To be submitted; [#]co-first authors) Author contribution: study was designed by D.W. under the supervision of L.L.; experiment was performed by D.W.; SUGAR tags were synthesized by M.L. and Z.W.; data was analyzed by D.W; manuscript was written by D.W. and M.L., and edited by J.H., T.G., Y.C., and L.L.

Abstract

Protein sialylation plays a crucial role in many biological processes. The high acidity and number of hydroxyl groups in sialic acids have a profound influence on surrounding species. Altered expression of sialylated glycoproteins has been associated with many diseases. However, direct mass spectrometry (MS) analysis of intact sialylated glycopeptides (SGPs) remains a great challenge due to their micro- and macro-heterogeneities, low abundances, and poor ionization efficiencies stemming from their negative charges. Furthermore, quantification strategies targeting at SGP are also limited. Here we present a high-throughput quantification strategy for intact SGPs that combines mild periodate oxidation and 12-plex isobaric multiplex labeling reagents for carbonyl-containing compound (SUGAR) tag labeling. The oxidation reaction exclusively introduces an aldehyde group to the sialic acid, which can react directly with the SUGAR tag. The chemical labeling strategy achieves almost complete reaction efficiency and enhances the detection of labeled SGPs. We validated our method using both peptide standards and complex tissue samples, demonstrating a promising glycoproteome coverage and excellent quantification accuracy. Moreover, we applied this strategy to investigate sialylation alterations in different brain regions using Alzheimer's disease (AD) mouse models. Our results revealed dysregulated sialylated glycoforms in both cortex and hippocampus regions, with a more distinct down regulation trend observed for hippocampus region. Overall, this study provides a robust and highly efficient approach for profiling and quantifying site-specific sialylation changes in complex biological samples, which is readily applicable to various biological applications.

Introduction

As one of the most common post-translational modifications (PTMs), glycosylation is present in most human proteins and plays a vital role in various biological processes such as cellular recognition, immune signaling, and host-pathogen interactions.[1] Of the various glycosylation forms, sialylation is unique where the end of glycans are decorated by the negatively charged sialic acids (SAs). SAs are a class of nine-carbon monosaccharides, with 5-N-acetylneuraminic acid (Neu5Ac) and 5-N-glycolylneuraminic acid (Neu5Gc) being widely distributed on membrane and secreted proteins in vertebrates.[2] The high acidity and numerous hydroxyl groups in SAs have a profound effect on surrounding species, resulting in partially charge-mediated promotion or inhibition of binding between proteins and cells, and high viscosity of highly sialylated surfaces.[3–5] Aberrant expression of SAs can induce pathological changes in various diseases, such as inflammatory diseases and cancers.[6,7] Therefore, in-depth analysis of protein sialylation is crucial for gaining insights into its myriad of functions.

Sialylation, like other types of glycosylation, exhibits both macro- and microheterogeneity, due to its occurrence on multiple glycosites and the presence of multiple glycan structures on the same site.[8] Previous studies have primarily focused on either glycan structures or glycosites, providing only average results of glycans from different glycosites or glycosite occupancies from total glycoforms, thereby missing the full picture of sialylation.[9,10] In contrast, intact sialylated glycopeptides (SGPs) has become a great research target because it contains the information of both glycan structure and glycosite, which enables site-specific analysis. Mass spectrometry (MS)-based proteomics has greatly advanced the characterization of glycoproteins.[11] However, the

analysis of intact SGPs still faces several obstacles, resulting from the low stoichiometry after digestion, and the reduced ionization efficiency due to the negatively charged SAs.[4,12] Moreover, SAs are labile and can be lost during sample preparation or in-source fragmentation.[9] Hence, the direct analysis of intact SGPs remains an analytical challenge.

In recent years, tremendous efforts have been made to enhance MS detection of SGPs, particularly through the development of highly efficient enrichment methods prior to LC-MS analysis.[13] While the traditional enrichment methods including lectin-based affinity chromatography and hydrophilic interaction liquid chromatography (HILIC) are still prevalent,[13,14] many novel materials have been developed recently for solid-phase extraction of SGPs based on their exclusive interactions.[15–20] In addition to physical adsorption, the *cis*-diol groups within the SAs structure offer chemical selectivity. By selective oxidation and derivatization, SGPs can be differentiated from the background matrix. This chemical strategy, called “reverse-glycoblotting”, has shown outstanding specificity.[21–23]

In addition to enrichment and qualitative profiling, quantitative characterization of SGPs takes the in-depth analysis a step further. Targeted MS using multiple reaction monitoring (MRM) was employed to investigate SGPs in mouse serum long time ago,[24] whereas a more sophisticated design utilizing mass-defect chemical labeling to guide targeted characterization of SGPs in MS1 level was reported recently.[25] Qin, H. et al achieved site-specific quantification of SGPs and identified significant changes in certain glycoforms between hepatocellular carcinoma and control samples.[26] Our group previously developed a novel chemical labeling strategy to enhance the ionization efficiency and fragmentation performance of glycopeptides, which is particularly

beneficial for the analysis of SGPs during electron-transfer/higher-energy collision dissociation (EThcD), as the charge states of the analytes are elevated and much more abundant c/z ions are collected.[27] Although these endeavors have made significant contributions to quantitative investigations of SGPs, they are mostly limited by sample handing capacity and analysis throughput.

Isobaric labeling is a popular method for achieving high-throughput analysis in MS, as it allows for the labeling of the same analytes from different samples with different channels of isobaric tags before combining them for LC-MS/MS. Labeled analytes have the same nominal mass in full MS but will produce distinct reporter ions in MS2 scan which contains relative abundance information.[28] Our group previously developed isobaric multiplex reagents for carbonyl-containing compound (SUGAR) tag for released glycan analysis.[29,30] It was designed as a 4-plex set initially and soon extended three times larger, capable of processing 12 samples at one time.[31] SUGAR tag employs reductive amination chemistry targeting the reducing end of glycans and possesses high labeling efficiency. The ionization and fragmentation of glycans were also enhanced after SUGAR conjugation.[30] We reasoned that these outstanding features could be further exploited and would be beneficial to quantitative SGP analysis.

In this study, we integrated selective chemical oxidation, 12-plex SUGAR tag isobaric labeling, and HILIC enrichment to develop a novel high-throughput quantitative method for intact SGPs analysis. Specifically, we used mild periodate oxidation to selectively convert the *cis*-diols of SAs on SGPs into aldehyde groups, which could be labeled by 12-plex SUGAR tags subsequently. The labeled samples were then equally pooled before HILIC enrichment, followed

by LC-MS/MS analysis. Our method was first validated using an SGP standard and further evaluated for quantitative accuracy and proteome coverage in analyzing complex samples such as bovine fetuin digests and mouse tissues. Finally, we applied our method to quantify SGPs in two brain regions of wild-type (WT) and Alzheimer's disease (AD) mouse models, revealing dysregulated sialylated glycoforms in both cortex and hippocampus regions, with more distinct differences observed for hippocampus region. Overall, our method offers a new and efficient way of quantitatively characterizing intact SGPs, which can aid in better understanding their roles in complex biological systems.

Experimental Section

Chemicals and Materials. Sequencing grade trypsin were from Promega (Madison, WI). Optimal LC/MS grade solvents, Tris base, urea, formic acid (FA), acetic acid (AA), ammonium acetate, ammonium bicarbonate (ABC), calcium chloride ($CaCl_2$) and sodium chloride (NaCl) were from Fisher Scientific (Pittsburgh, PA). Bovine fetuin, trifluoroacetic acid (TFA), Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), iodoacetamide (IAA), and sodium cyanoborohydride ($NaBH_3CN$) were purchased from Sigma-Aldrich (St Louis, MO). Sodium periodate ($NaIO_4$) was obtained from Thermo Scientific Chemicals (Waltham, MA). Sialylglycopeptide (SGP) standard was purchased from TCI (Portland, OR). HILIC bulk material PolyHYDROXYETHYL (12 µm, 300 Å) was purchased from PolyLC (Columbia, MD). Empty TopTips were from Glygen Corp (Columbia, MD). Protease inhibitor cocktail tablets and phosphatase inhibitor cocktail tablets were from Roche (Mannheim, Germany). Microcon 30kDa molecular weight cut-off (MWCO) filter units were purchased from MilliporeSigma (Burlington, MA).

Mouse tissue sample preparation. The mouse heart tissue used in this study was collected from male adult C57BL/6 mice. The mouse brain tissue was collected from APP/PS1 double transgenic mice and their WT littermates. Tissues were lysed in buffer consisting of 8M urea, 50 mM Tris (pH 8), 5 mM $CaCl_2$, 20mM NaCl, and 1 EDTA-free Roche protease inhibitor tablet and 1 Roche PhosSTOP phosphatase inhibitor tablet with a probe sonicator in ice water bath at 50% power with pulse 5s on 5s off for 12 cycles. Tissue homogenates were centrifuged at 14000 ×g for 5 min, after which the supernatant was collected. Protein concentrations were measured by a BCA assay kit (Thermo Fisher Scientific, San Jose, CA).

Protein digestion. Protein digestion was performed based on filter-aided sample preparation (FASP) protocol with some modifications.[32] For standard proteins, 1 mg protein was dissolved in 500 µL of 50 mM ABC buffer and heat denatured at 95^{o}C for 10 min. Proteins were reduced in 10 mM TCEP at room temperature for 30 min, followed by alkylation with 20 mM IAA in the dark for 30 min. The protein solution was transferred to a 30k MWCO filter unit and centrifuged at 14000 ×g for 10 min. The filter was washed three times with 250 µL of 50 mM ABC followed by centrifugation at 14000 ×g for 20 min each time. The filter was then transferred to a new collection tube, and the sample was resuspended with 250 µL of 50 mM ABC and digested with trypsin at a protein-to-enzyme ratio of 50:1 at 37 ^{o}C for 18 h. The peptide digests were collected by centrifugation at 14000 ×g for 10 min, and the process was repeated once with an additional 50 µL of water. The resulting samples were dried down *in vacuo* and stored at -80^{o}C for future use.

Sialic acid oxidation and Sugar Labeling. SUGAR tags were synthesized as described in previous publication.[31] Peptides were dissolved in 100 mM ammonium acetate (pH 5.5). For 12-

plex experiment, 0.1 µg of SGP standard was spiked into 100 µg peptide sample as an internal standard to minimize variation during sample preparation. Sialic acids were selectively oxidized by adding $NaIO_4$ to a final concentration of 5 mM and incubating on ice in the dark for 30 min. SUGAR tags were dissolved in MeOH containing 1% FA at a concentration of 10 mg/mL and directly added to the peptide sample at a tag-to-peptide ratio of 10:1. The mixture was vortexed at room temperature for 10 min, followed by addition of $NaBH_3CN$ to a final concentration of 50 mM. The mixture was vortexed at room temperature for another 30 min to allow for reductive amination, after which the reaction was quenched by adding AA to a final concentration of 5% (v/v). Labeled peptides was pooled together and dried down *in vacuo.*

HILIC Enrichment. Enrichment of unlabeled or SUGAR-labeled glycopeptides was performed with in-house packed HILIC SPE tips following previous publications with minor modification.[33] Briefly, a 200 µL empty TopTip was plugged with 3 mg cotton wool and placed on a 2 mL microcentrifuge tube with an adapter unit. The HILIC material was activated in 1% TFA for 15 min and then transferred into the cotton-packed TopTip at a beads-to-peptide ratio of 30:1. After removing the solvent at 200 ×g for 2 min, the stationary phase was conditioned with 300 µL 1% TFA and loading buffer (80% ACN/1% TFA) for three cycles. The peptides were dissolved in 300 µL loading buffer and loaded onto the HILIC-cotton tip. The tip was centrifuged at 200 ×g for 2 min, and the flow-through was reloaded to the HILIC-cotton five more times to ensure complete retention. The HILIC-cotton tip was washed six times with 300 µL loading buffer, and then eluted with 150 µL 0.1% FA solution three times. Samples were dried down *in vacuo* before direct MS analysis or HpH fractionation (details provided in **Supplemental information**).

LC-MS/MS Analysis. An in-house fabricated nano-C18 column (15 cm length, 75 μm i.d.,) packed with Bridged Ethylene Hybrid C18 material (1.7 μm, 130 Å, Waters) was used for sample separation. Details of MS analysis of SGP standard and bovine fetuin samples are provided in **Supplemental information.** For mouse tissue samples, lyophilized peptides were reconstituted in 0.1% FA and analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled with a Dionex UltiMate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). The LC was performed with binary buffers consisting of buffer A (0.1% FA) and buffer B (90% ACN, 0.1% FA), and a linear gradient from 0% to 30% buffer B was used over 80 minutes. MS survey scans were acquired over the range of 400 to 2000 *m/z* at a resolution of 60K, with an AGC target setting of 4E5 and a maximum injection time of 50 ms. For MS2 scan, a duty cycle of 3s was set in top speed mode. Only spectra with a charge state among 2-7 were selected for fragmentation by stepped higher-energy collision dissociation (HCD) with normalized collision energy (NCE) of 30 ± 5%. The MS2 spectra were acquired with a resolution of 60 K, a lower mass limit of *m/z* 110, dynamic exclusion of 12 s with 10 ppm mass tolerance, an isolation width of 1 Da, an AGC target of 5E4, and an IT of 200 ms.

Data Analysis. Raw files were searched against the bovine fetuin protein sequence or the UniProt Mus musculus reviewed database (January 2021, 17033 sequences) using Byonic search engine (version 4.5.2, Protein Metrics Inc) embedded within Proteome Discoverer 2.5 (PD 2.5, Thermo Fisher Scientific). The precursor mass tolerance was set at 10 ppm, and the fragment mass tolerance was set at 0.02 Da. Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.02 Da. Static modifications were specified as

carbamidomethylation (+57.02146 Da) on cysteine residues. Dynamic modifications included oxidation of methionine residues (+15.99492 Da, rare 2), and N-glycosylation (common 1). For the bovine fetuin samples, glycan modifications were searched against a glycan database in **Table S1** as reported in previous literature.[25,34] For mouse tissue samples, the glycan modifications were searched against the Byonic embedded 309 mammalian N-glycan database. The sialic acid containing glycans were modified with an additional mass of +152.1426 or +156.1586 per sialic acid for nonisotopic (NI) or 12-plex SUGAR tag-labeled samples. The Custom Peaks function in Byonic was utilized to include the labeled oxonium ions and reporter ions (**Table S2**) in database search and enhance the scoring of SUGAR-labeled glycopeptides.[35] Dynamic modifications on peptide N-terminal Ser (NI SGUAR: +183.13720 Da, or 12-plex SUGAR: +187.1532, rare 1) , threonine (NI SGUAR: +169.12155 Da, or 12-plex SUGAR: +173.13755, rare 1) and oxidation on Cys (rare 2) were set to include potential side reactions during oxidation and labeling.[36] The N-glycopeptides identified were filtered at 1% peptide FDR and Byonic score $\geq$ 250. Further manual inspection was performed to ensure the existence of the signature oxonium ions. Quantification was performed in PD 2.5 with a reporter ion integration tolerance of 10 ppm for the most confident centroid. Reporter ion intensity of each channel was normalized by the intensity of the spike-in SGP standard to mitigate the variation in sample handling or systematic biases of 12-plex SUGAR tags using a Python script. Student's t-test was performed by Perseus,[37] where significant change was defined by a p-value less than 0.05 and fold-change over 1.2.

Results and Discussion

SGP Oxidation and SUGAR Tag Labeling. The SUGAR tags were originally developed for N-glycan labeling and quantification,[29,31] but can be adapted for proteome-level quantification if an aldehyde group can be generated on a peptide to provide an active reaction site with the hydrazide group. Previous studies have shown that the *cis*-diol group of the sialic acid can be selectively oxidized, and an aldehyde group is formed at the C7 position using periodate under mild conditions.[21,25,30,38] The resulting aldehyde group can directly react with SUGAR tag to enable subsequent high-throughput relative quantification. The strategy was first tested with an SGP standard. As shown in **Figure 2A** and **2B**, the two sialic acids on the SGP standard were completely oxidized after incubating in 5mM $NaIO_4$ on ice for 30 min, resulting in a mass shift of -124 Da in the MALDI-TOF MS spectra. Next, NI SUGAR tag solution was directly added to the peptide sample and reacted at room temperature for 10 min. **Figure 2C** shows that the oxidized SGP successfully reacted with the SUGAR tag, as reflected by the mass shift of +424 Da. In this step, the hydrazide formed a hydrazone bond with the aldehyde group as a reversible conjugation. To preserve the labeled peptide during the subsequent sample clean-up and MS data acquisition, it is necessary to convert hydrazone bond into hydrazine bond through irreversible reductive amination. In the original N-glycan SUGAR labeling workflow, a high concentration of the reducing agent, 1M $NaBH_3CN$ in 7:3 (v/v) dimethyl sulfoxide: acetic acid, was used and reacted at 70^{o}C for 2h to facilitate the yield of glycan reducing end and achieve complete irreversible labeling.[29] We first tried the same condition on the SGP labeling, however our result suggested that the condition was too harsh for SGP labeling, as no signal of intact SGP was observed on MALDI-MS after the reaction and clean-up. We then decreased the reaction temperature and reduced the concentration

of $NaBH_3CN$, as we found that a mild concentration is sufficient to reduce the hydrazone bond and there is no need to yield the glycan reducing end in this workflow. Ultimately, we achieved the complete reductive amination reaction and preserved the labile glycosidic bond in SGP by adding $NaBH_3CN$ directly to the original peptide solution to a final concentration of 50 mM and reacted at room temperature for 30 min. The product was spotted on MALDI plate right after the reaction, and showed a mass increase of 4 Da, indicating the successful addition of four H atoms (**Figure 2D**). Finally, the labeled peptide was cleaned up by HILIC to remove the excess salts and tags. The MALDI spectrum clearly showed that the labeled SGP was well-recovered from the reaction system and exhibited near complete labeling efficiency (**Figure 2E**).

LC-MS/MS Performance of Labeled SGP. After validating the completeness of the labeling strategy with MALDI-MS, we assessed the LC-MS/MS performance of the labeled SGP. **Figure 3A** illustrates the chemical structure of the NI SUGAR-labeled sialic acid on the peptide and the resulting product ions upon HCD fragmentation. **Figure 3B** is a representative MS/MS spectrum of the NI SUGAR labeled SGP, which shows a high intensity of the reporter ion at *m/z* 114.13. Besides the commonly observed oxonium ions at *m/z* 204.09 and 366.14, the labeled SGP also generated the sialic acid oxonium ions carrying the SUGAR tag, which had a significantly enhanced signal intensity at *m/z* 444.25. These signature ions provide evidence of successful labeling on the sialic acid and can serve as diagnostic ions for validating the identification of glycopeptides when analyzing complex biological samples. SGP standards were also labeled with 12-plex SUGAR tags and mixed at 1:1:1:1 ratio. As shown in **Figure 3C**, equal ratio of 12 reporter ions were observed from *m/z* 115 to 118, all of which were well-resolved at a MS2 resolution of

70K on the Q Exactive mass spectrometer. A high intensity of sialic acid oxonium ion was also detected at *m/z* 448.26, which indicates the consistency of the generation of signature ions.

Furthermore, a charge state enhancement was observed for labeled SGP compared to the unlabeled SGP. While over 70% of PSMs of the unlabeled SGP were identified with charge state of 3+, 41.6% and 43.5% of PSMs of the labeled SGP were found to have a charge state of 5+ and 6+, respectively (**Figure S2**). The higher charge state might be a result of the increased hydrophobicity provided by the SUGAR tag structure. This increase in ionization efficiency is beneficial for SGP analysis since intact glycopeptides have considerably greater *m/z* distributions compared to unmodified peptides, and low-charge density precursors ($z \leq 3$) can be challenging to dissociate.[39] The extra negative charge on sialic acid further suppresses their detection in positive mode MS, and the increase of charge state helps to mitigate the suppression.[40] This feature may also be useful when performing EThcD fragmentation for SGP characterization, as higher charge states prevent the non-dissociative electron transfer (ETnoD) phenomenon.[41,42]

Next, the chromatographic property of the labeled SGP was evaluated to examine whether the incorporation of isotopologues in different channels would potentially cause retention time shifts and affect quantification accuracy. Unlike our previously reported DiLeu tags,[43] half of the 12-plex SUGAR tags incorporated deuterium atoms within the structure to eliminate the need for O^{18} exchange. However, this puts a risk of causing retention time shift, also referred to deuterium effect.[44,45] We divided the 12 channels into 5 groups based on the number of deuterium ions and evaluated the chromatography retention time alignment of the labeled SGP on RPLC. According to the extracted ion chromatograms (EIC), no retention time shifts were observed (**Figure S3**)

which confirmed the accuracy of SUGAR tag-based isobaric quantification. This can be explained by the fact that the deuterium atoms were distributed around the polar glycan structure, while the hydrophobic peptide backbone played a major role in interacting with the C18 stationary phase, thus minimizing the effect of deuterium shift.

Evaluation of the Labeling Strategy with Bovine Fetuin Digests. To further test this labeling strategy with protein digests and establish a database search workflow, we selected bovine fetuin as a model protein due to its reported sialylation at multiple sites.[23,25,34,46] The protein standard was denatured at high temperature and processed with the FASP method to obtain tryptic digested peptides. With our optimized sample oxidation and labeling workflow, all reactions can be directly conducted in a single tube without the need of extra drying down or desalting process before the final HILIC enrichment. As the LC-MS/MS data was acquired from a more complex sample, it is necessary to build an efficient workflow to interpret the data with current software tools. Our initial attempt with the regular Byonic search and default database settings resulted in a low number of identified SGP PSMs, as Byonic failed to recognize the SUGAR labeled sialic acid oxonium ions. To overcome this issue, we used the new feature introduced by Byonic in 2021, Custom Peaks, to include information about signature ions during database search.[35] Specifically, we included the reporter ion peak and the sialic acid oxonium ion peaks (*m/z* 444.2458 for NI SUGAR and *m/z* 448.2619 for 12-plex SUGAR) via the "CustomPeaks" command. This allowed Byonic to use a median predicted peak intensity for new peaks added, thus improving the scoring. With this function, the PSM number of NI SUGAR labeled bovine fetuin digests increased from 244 to 478. **Figure S4** shows one MS/MS spectrum that was identified by using the conventional approach

and by using the new setting with custom peaks. The conventional method failed to recognize the reporter ion at *m/z* 114.13 and the sialic acid oxonium ions at *m/z* 444.25 and 364.23, whereas all these ions were well-annotated in the current method. The score also dramatically increased from 503.4 to 835.9.

After setting up the Byonic search method, we compared the SUGAR labeled datasets with bovine digests analyzed using the conventional label-free approach. Overall, the labeled samples showed a comparable coverage with the label-free samples in terms of the total N-glycopeptides and N-glycoforms (**Figure 4A, Table S3**). 18 glycoforms overlapped in three experiments (**Figure 4B**), covering all reported N-glycosylation sites on protein P12763 and Q58D62 in Uniprot. This demonstrates the efficacy of the labeling strategy and good recovery of labeled glycopeptides from peptide mixtures. In the 12-plex experiment, samples were mixed at 1:1:1:1 ratio, and the observed ratios were all well-matched to the theoretical ratios, with less than 7% of deviations across all channels (**Figure 4C**).

Streamlined Workflow of Multiplex SGP Quantification from Complex Biological Samples. With the successful implementation of the labeling and database search strategy on glycoprotein standards, we aimed to streamline the full workflow for samples with higher complexity. We extracted proteins from mouse heart tissues and performed the same oxidation and labeling in one tube after FASP digestion. We also analyzed the same amount of peptide aliquots as a label-free control in parallel experiments. As an average of two technical replicates, 253 intact SGP labeled with SUGAR tags were identified with high confidence, corresponding to 230 unique glycoforms (**Figure 5A**). This identification accounted for 83.8% of the identification number in the label-free

dataset. Given the fact that multiple steps of chemical derivatization were conducted, and label-based quantification strategies were reported to have considerably lower proteome coverage compared to label-free approaches,[47,48] the slight decrease in identification is acceptable. [49,50]Still, it can be offset by the high-throughput feature of this labeling strategy, as additional fractionation or multiple injections can be performed while still saving the instrument time compared to analyzing 12 samples individually. Additionally, the label-free and labeled datasets showed a decent overlap of identified glycosites and glycan compositions (**Figure S5**). In the meantime, the labeling strategy improved the mapping of multi-sialylated glycopeptides, as evidenced by the discovery of more such peptides in the labeled dataset (**Figure 5B**). The enhanced ionization efficiency and oxonium ion signal intensities likely contributed to this improvement. Furthermore, our modified database search strategy also contributed to the SGP identification. As both NeuAc and NeuGc may be present in mammalian tissue samples, it can be challenging to differentiate the glycan compositions that have the same mass, e.g., Hex(1)NeuAc(1) = Fuc(1)NeuGc(1).[49] By using the conventional searching parameters, Byonic not only scored lower for a spectrum with good quality but also misassigned the isobaric glycan composition due to a lack of sialic acid-specific oxonium ion information (**Figure S6A**). However, after applying the Custom Peak function, the signature ions were well-annotated, resulting in a much higher scoring. With the confident assignment of glycan compositions, we observed similar percentages of SGPs containing only NeuAc in the two datasets and a slight increase of glycans containing both NeuAc and NeuGc, likely due to the enhanced detection of multiply sialylated species (**Figures 5C, 5D**). In general, more NeuGc-containing SGPs were found in mouse heart, which was consistent with the previous

literature,[50] indicating that our method did not show any bias towards the sialylation type in biological samples.

To evaluate the quantification performance in complex biological samples, we prepared 12-plex labeled samples mixed at 1:1:1:1 and 1:2:5:10 ratios and performed the quantitative analysis. To minimize potential variation during sample preparation and bias in data acquisition, the same amount of SGP standard was spiked into the mouse heart digests as an internal standard. The peptide intensities were normalized by the intensity of the SGP standard in each channel during data analysis. The resulting ratio boxplots showed good quantification performance, with median ratios measured within 8% of the expected values and an average coefficient of variation of 19.1% across all channels at the two ratios (**Figures 6A, 6B)**. Altogether, the results indicated a highly effective, accurate and streamlined workflow for multiplex intact SGP quantification.

Site-specific SGP Quantification in APP/PS1 Mouse Model. Alzheimer's disease (AD) is the most common cause of dementia and has affected more than 50 million patients worldwide,[51] yet its underlying pathogenic mechanisms remain elusive, and a cure is not available. [52,53] As a proof-of-principle study, we adopted our 12-plex SUGAR tag method to quantify sialylation changes in different mouse brain regions during AD progression. We extracted proteins from the cortex and hippocampus tissues of three WT and three APP/PS1 mice, labeled their tryptic digests, pooled the samples, and analyzed them simultaneously by LC-MS/MS. In total, 390 sialylated glycoforms were identified and quantified from one 12-plex experiment, mapping to 134 glycoproteins and 125 different glycans. 84.3% of the glycoproteins were observed with only one sialylated glycosite, while 39.6% of the glycosites were modified with more than one sialic acid-containing glycan. Of

note, 9.1% of the sialylated sites were found to have more than 5 glycans attached, suggesting a high degree of heterogeneity in mouse brain sialylation. Near half of identified glycans contain more than one sialic acid, with 14.4% and 4.8% of them possessing 3 and 4 sialic acid residues respectively, indicating the importance to develop a highly sensitive method to detect and quantify these SGP species (**Figure S7A**). Unlike the mouse heart tissue, the mouse brain was dominated by NeuAc-containing SGPs in accordance with previous findings (**Figure S7B**).[54,55]

Quantitative results revealed dysregulated sialylated glycoforms in both cortex and hippocampus regions by student's *t*-test (*p*-value<0.05) with a fold-change over 1.2 (**Table S4**). More pronounced difference was observed for the SGPs from the hippocampus region than the cortex region, as 45 unique glycoforms were found to be downregulated in hippocampus while only 2 were significantly changed in cortex (**Figure 7A**, **Figure S8A**). The hierarchical clustering also indicated larger intergroup differences than intragroup variations in hippocampus compared to cortex (**Figure 7B**, **Figure S8B**). This can be explained by the fact that hippocampus is the crucial region for learning and memory and is among the first structures affected by AD.[56–60] The decreasing trend of sialylation were also observed in other AD-related studies,[55,61,62] and might be a result of decreasing sialyltransferase levels as the disease progresses.[63,64] To further understand the functions of dysregulated proteins in hippocampus, GO analysis was performed and a network of enriched biological processes was constructed using Metascape (**Figure 7B)**.[65] The results revealed that these proteins were associated with several important signaling and developmental processes, including transmission across chemical synapses, axon development, cell-cell adhesion, and regulation of neuron apoptotic process, which have been discussed to have potential

correlation with the neuronal degeneration in AD.[66–70] While further studies are necessary to decipher the mechanism of the dysregulation of sialylation in AD hippocampus, the results demonstrated the feasibility of our method for high-throughput and site-specific quantification of sialylation from different disease models in a single experiment.

Conclusions

In summary, our study presents a high-throughput quantification strategy for intact SGP using mild periodate oxidation and 12-plex SUGAR isobaric tag labeling. The optimized experimental conditions allow for all reactions to be performed in one tube, greatly simplifying the workflow and minimizing sample loss. The strategy was validated using both peptide standards and complex biological samples, demonstrating promising glycoproteome coverage and quantification accuracy. Furthermore, we successfully applied our method to explore sialylation changes between AD and WT mouse model, providing new insights into brain region-specific PTM expression in disease progression. While the current study focuses on N-linked sialylation, this method can be adapted for O-linked SGP profiling in the future. Overall, the high-throughput capability of our method provides a valuable tool for large-scale and in-depth quantitative analysis of SGPs in complex biological systems.

Acknowledgements

This work was supported, in part, by the National Institutes of Health Grants RF1AG052324, R01 AG078794, and R01 DK071801 (to L.L.). The Orbitrap instruments were purchased through the support of an NIH Shared Instrument Grant (NIH-NCRR S10RR029531 to L.L.) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate

Education with funding from the Wisconsin Alumni Research Foundation. L.L. would like to acknowledge NIH grants R21AG065728, and S10OD025084, as well as funding support from a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Reference

(1) Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology.*; Cold Spring Harber Laboratory Press, New York; **1998**.

(2) Griffin, M. E.; Hsieh-Wilson, L. C. Tools for Mammalian Glycoscience Research. *Cell* **2022**, *185* (15), 2657–2677.

(3) Schauer, R.; Kamerling, J. P. Exploration of the Sialic Acid World. *Adv. Carbohyd. Chem. Biochem.* **2018**, *75*, 1–213.

(4) Haan, N. de; Yang, S.; Cipollo, J.; Wuhrer, M. Glycomics Studies Using Sialic Acid Derivatization and Mass Spectrometry. *Nat. Rev. Chem.* **2020**, *4* (5), 229–242.

(5) Puigdellívol, M.; Allendorf, D. H.; Brown, G. C. Sialylation and Galectin-3 in Microglia-Mediated Neuroinflammation and Neurodegeneration. *Front Cell Neurosci.* **2020**, *14*, 162.

(6) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.

(7) Pagan, J. D.; Kitaoka, M.; Anthony, R. M. Engineered Sialylation of Pathogenic Antibodies In Vivo Attenuates Autoimmune Disease. *Cell* **2018**, *172* (3), 564-577.e13.

(8) Fang, P.; Ji, Y.; Oellerich, T.; Urlaub, H.; Pan, K.-T. Strategies for Proteome-Wide Quantification of Glycosylation Macro- and Micro-Heterogeneity. *Int. J. Mol. Sci.* **2022**, *23* (3), 1609.

(9) Nie, H.; Li, Y.; Sun, X.-L. Recent Advances in Sialic Acid-Focused Glycomics. *J. Proteomics* **2012**, *75* (11), 3098–3112.

(10) Zhang, Z.; Sun, Z.; Zhu, J.; Liu, J.; Huang, G.; Ye, M.; Zou, H. High-Throughput Determination of the Site-Specific N-Sialoglycan Occupancy Rates by Differential Oxidation of Glycoproteins Followed with Quantitative Glycoproteomics Analysis. *Anal. Chem.* **2014**, *86* (19), 9830–9837.

(11) Lu, H.; Zhang, Y.; Yang, P. Advancements in Mass Spectrometry-Based Glycoproteomics and Glycomics. *Natl. Sci. Rev.* **2016**, *3* (3), 345–364.

(12) Gil, G.-C.; Iliff, B.; Cerny, R.; Velander, W. H.; Cott, K. E. V. High Throughput Quantification of N-Glycans Using One-Pot Sialic Acid Modification and Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2010**, *82* (15), 6613–6620.

(13) Zhu, F.; Clemmer, D. E.; Trinidad, J. C. Characterization of Lectin Binding Affinities via Direct LC-MS Profiling: Implications for Glycopeptide Enrichment and Separation Strategies. *Analyst* **2016**, *142* (1), 65–74.

(14) Chen, Y.-J.; Yen, T.-C.; Lin, Y.-H.; Chen, Y.-L.; Khoo, K.-H.; Chen, Y.-J. ZIC-CHILIC-Based StageTip for Simultaneous Glycopeptide Enrichment and Fractionation toward Large-Scale N-Sialoglycoproteomics. *Anal. Chem.* **2021**.

(15) Dong, X.; Qin, H.; Mao, J.; Yu, D.; Li, X.; Shen, A.; Yan, J.; Yu, L.; Guo, Z.; Ye, M.; Zou, H.; Liang, X. In-Depth Analysis of Glycoprotein Sialylation in Serum Using a Dual-Functional Material with Superior Hydrophilicity and Switchable Surface Charge. *Anal. Chem.* **2017**, *89* (7), 3966–3972.

(16) Zhang, M.; Liu, Y.; Zhang, D.; Chen, T.; Li, Z. Facile and Selective Enrichment of Intact Sialoglycopeptides Using Graphitic Carbon Nitride. *Anal. Chem.* **2017**, *89* (15), 8064–8069.

(17) Wan, H.; Zhang, X.; Chen, C.; Li, X.; Liang, X. Selective Enrichment of Sialylated Glycopeptides with Mesoporous Poly-Melamine-Formaldehyde (MPMF) Material. *Anal. Bioanal. Chem.* **2020**, *412* (7), 1497–1508.

(18) Sun, N.; Xiong, Y.; Qing, G.; Zhao, Y.; Li, X.; Liang, X. Selective Enrichment of Sialylated Glycopeptides with a d -Allose@SiO 2 Matrix. *Rsc Adv* **2018**, *8* (68), 38780–38786.

(19) Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

(20) Huang, J.; Wang, D.; Shipman, R. D.; Zhu, Z.; Liu, Y.; Li, L. Simultaneous Enrichment and Separation of Neutral and Sialyl Glycopeptides of SARS-CoV-2 Spike Protein Enabled by Dual-Functionalized Ti-IMAC Material. *Anal. Bioanal. Chem.* **2021**, 1–9.

(21) Kurogochi, M.; Amano, M.; Fumoto, M.; Takimoto, A.; Kondo, H.; Nishimura, S. Reverse Glycoblotting Allows Rapid-Enrichment Glycoproteomics of Biopharmaceuticals and Disease-Related Biomarkers. *Angew. Chem. Int. Ed.* **2007**, *46* (46), 8808–8813.

(22) 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.

(23) Li, M.; Huang, J.; Ma, M.; Shi, X.; Li, L. Selective Enrichment of Sialylglycopeptides Enabled by Click Chemistry and Dynamic Covalent Exchange. *Anal. Chem.* **2022**, *94* (18), 6681–6688.

(24) Kurogochi, M.; Matsushista, T.; Amano, M.; Furukawa, J.; Shinohara, Y.; Aoshima, M.; Nishimura, S.-I. Sialic Acid-Focused Quantitative Mouse Serum Glycoproteomics by Multiple Reaction Monitoring Assay*. *Mol Cell Proteomics* **2010**, *9* (11), 2354–2368.

(25) Di, Y.; Zhang, L.; Zhang, Y.; Zhao, H.; Yan, G.; Yao, J.; Zhang, S.; Lu, H. MdCDPM: A Mass Defect-Based Chemical-Directed Proteomics Method for Targeted Analysis of Intact Sialylglycopeptides. *Anal. Chem.* **2019**, *91* (15), 9986–9992.

(26) Qin, H.; Dong, X.; Mao, J.; Chen, Y.; Dong, M.; Wang, L.; Guo, Z.; Liang, X.; Ye, M. Highly Efficient Analysis of Glycoprotein Sialylation in Human Serum by Simultaneous Quantification of Glycosites and Site-Specific Glycoforms. *J. Proteome Res.* **2019**, *18* (9), 3439–3446.

(27) Li, M.; Zhong, X.; Feng, Y.; Li, L. Novel Isobaric Tagging Reagent Enabled Multiplex Quantitative Glycoproteomics via Electron-Transfer/Higher-Energy Collisional Dissociation (EThcD) Mass Spectrometry. *J. Am. Soc. Mass Spectr.* **2022**, *33* (10), 1874–1882.

(28) Rauniyar, N.; Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J. Proteome Res.* **2014**, *13* (12), 5293–5309.

(29) Feng, Y.; Chen, B.; Yu, Q.; Zhong, X.; Frost, D. C.; Ikonomidou, C.; Li, L. Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound (SUGAR) Tags: A Probe for Quantitative Glycomic Analysis. *Anal. Chem.* **2019**, *91* (4), 3141–3146.

(30) Feng, Y.; Li, M.; Lin, Y.; Chen, B.; Li, L. Multiplex Quantitative Glycomics Enabled by Periodate Oxidation and Triplex Mass Defect Isobaric Multiplex Reagents for Carbonyl-Containing Compound Tags. *Anal. Chem.* **2019**, *91* (18), 11932–11937.

(31) Li, M.; Feng, Y.; Ma, M.; Kapur, A.; Patankar, M.; Li, L. High-Throughput Quantitative Glycomics Enabled by 12-Plex Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound (SUGAR) Tags. *J. Proteome Res.* **2023**.

(32) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6* (5), 359–362.

(33) Wang, D.; Ma, M.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, M.; Wang, Z.; Zetterberg, H.; Li, L. Boost-DiLeu: Enhanced Isobaric N,N-Dimethyl Leucine Tagging Strategy for a Comprehensive Quantitative Glycoproteomic Analysis. *Anal. Chem.* **2022**, *94* (34), 11773–11782.

(34) Lin, Y.-H.; Franc, V.; Heck, A. J. R. Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* **2018**, *17* (8), 2861–2869.

(35) Roushan, A.; Wilson, G. M.; Kletter, D.; Sen, K. I.; Tang, W.; Kil, Y. J.; Carlson, E.; Bern, M. Peak Filtering, Peak Annotation, and Wildcard Search for Glycoproteomics. *Mol. Cell Proteomics* **2021**, *20*, 100011.

(36) Huang, J.; Qin, H.; Sun, Z.; Huang, G.; Mao, J.; Cheng, K.; Zhang, Z.; Wan, H.; Yao, Y.; Dong, J.; Zhu, J.; Wang, F.; Ye, M.; Zou, H. A Peptide N-Terminal Protection Strategy for Comprehensive Glycoproteome Analysis Using Hydrazide Chemistry Based Method. *Sci. Rep.* **2015**, *5* (1), 10164.

(37) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13* (9), 731–740.

(38) Zeng, Y.; Ramya, T. N. C.; Dirksen, A.; Dawson, P. E.; Paulson, J. C. High-Efficiency Labeling of Sialylated Glycoproteins on Living Cells. *Nat. Methods* **2009**, *6* (3), 207–209.

(39) Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Capturing Site-Specific Heterogeneity with Large-Scale N-Glycoproteome Analysis. *Nat. Commun.* **2019**, *10* (1), 1–13.

(40) Lin, C.; Haeuptle, M. A.; Aebi, M. Supercharging Reagent for Enhanced Liquid Chromatographic Separation and Charging of Sialylated and High-Molecular-Weight Glycopeptides for NanoHPLC–ESI-MS/MS Analysis. *Anal. Chem.* **2016**, *88* (17), 8484–8494.

(41) Frey, B. L.; Ladror, D. T.; Sondalle, S. B.; Krusemark, C. J.; Jue, A. L.; Coon, J. J.; Smith, L. M. Chemical Derivatization of Peptide Carboxyl Groups for Highly Efficient Electron Transfer Dissociation. *J. Am. Soc. Mass Spectr.* **2013**, *24* (11), 1710–1721.

(42) Yang, L.; Sun, Z.; Zhang, L.; Cai, Y.; Peng, Y.; Cao, T.; Zhang, Y.; Lu, H. Chemical Labeling for Fine Mapping of IgG N-Glycosylation by ETD-MS. *Chem. Sci.* **2019**, *10* (40), 9302–9307.

(43) Frost, D. C.; Greer, T.; Li, L. High-Resolution Enabled 12-Plex DiLeu Isobaric Tags for Quantitative Proteomics. *Anal. Chem.* **2015**, *87* (3), 1646–1654.

(44) Filer, C. N. Isotopic Fractionation of Organic Compounds in Chromatography. *J. Labelled Cpd. Radiopharm.* **1999**, *42* (2), 169–197.

(45) Boutilier, J. M.; Warden, H.; Doucette, A. A.; Wentzell, P. D. Chromatographic Behaviour of Peptides Following Dimethylation with H2/D2-Formaldehyde: Implications for Comparative Proteomics. *J. Chromatogr. B* **2012**, *908*, 59–66.

(46) Zhong, J.; Huang, Y.; Mechref, Y. Derivatization of Sialylated Glycopeptides (DOSG) Enabling Site-Specific Isomeric Profiling Using LC-MS/MS. *Anal. Chem.* **2021**, *93* (14), 5763–5772.

(47) Megger, D. A.; Pott, L. L.; Ahrens, M.; Padden, J.; Bracht, T.; Kuhlmann, K.; Eisenacher, M.; Meyer, H. E.; Sitek, B. Comparison of Label-Free and Label-Based Strategies for Proteome Analysis of Hepatoma Cell Lines. *Biochim. Biophys. Acta - Proteins Proteom.* **2014**, *1844* (5), 967–976.

(48) Stepath, M.; Zülch, B.; Maghnouj, A.; Schork, K.; Turewicz, M.; Eisenacher, M.; Hahn, S.; Sitek, B.; Bracht, T. Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi. *J. Proteome Res.* **2020**, *19* (2), 926–937.

(49) Nwosu, C. C.; Aldredge, D. L.; Lee, H.; Lerno, L. A.; Zivkovic, A. M.; German, J. B.; Lebrilla, C. B. Comparison of the Human and Bovine Milk N-Glycome via High-Performance Microfluidic Chip Liquid Chromatography and Tandem Mass Spectrometry. *J. Proteome Res.* **2012**, *11* (5), 2912–2924.

(50) 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.

(51) Breijyeh, Z.; Karaman, R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* **2020**, *25* (24), 5789.

(52) Canter, R. G.; Penney, J.; Tsai, L.-H. The Road to Restoring Neural Circuits for the Treatment of Alzheimer's Disease. *Nature* **2016**, *539* (7628), 187–196.

(53) Lane, C. A.; Hardy, J.; Schott, J. M. Alzheimer's Disease. *Eur. J. Neurol.* **2018**, *25* (1), 59–70.

(54) Gerardy-Schahn, R.; Delannoy, P.; Itzstein, M. von; Davies, L. R. L.; Varki, A. Why Is N-Glycolylneuraminic Acid Rare in the Vertebrate Brain? *Top. Curr. Chem.* **2013**, *366*, 31–54.

(55) Fang, P.; Xie, J.; Sang, S.; Zhang, L.; Liu, M.; Yang, L.; Xu, Y.; Yan, G.; Yao, J.; Gao, X.; Qian, W.; Wang, Z.; Zhang, Y.; Yang, P.; Shen, H. Multilayered N-Glycoproteome Profiling Reveals Highly Heterogeneous and Dysregulated Protein N-Glycosylation Related to Alzheimer's Disease. *Anal. Chem.* **2019**, *92* (1), 867–874.

(56) Chupin, M.; Gérardin, E.; Cuingnet, R.; Boutet, C.; Lemieux, L.; Lehéricy, S.; Benali, H.; Garnero, L.; Colliot, O.; Initiative, A. D. N. Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data from ADNI. *Hippocampus* **2009**, *19* (6), 579–587.

(57) Wang, X.; Michaelis, M. L.; Michaelis, E. K. Functional Genomics of Brain Aging and Alzheimer's Disease: Focus on Selective Neuronal Vulnerability. *Curr. Genomics* **2010**, *11* (8), 618–633.

(58) DeTure, M. A.; Dickson, D. W. The Neuropathological Diagnosis of Alzheimer's Disease. *Mol. Neurodegener.* **2019**, *14* (1), 32.

(59) Planche, V.; Manjon, J. V.; Mansencal, B.; Lanuza, E.; Tourdias, T.; Catheline, G.; Coupé, P. Structural Progression of Alzheimer's Disease over Decades: The MRI Staging Scheme. *Brain Commun.* **2022**, *4* (3), fcac109.

(60) Rao, Y. L.; Ganaraja, B.; Murlimanju, B. V.; Joy, T.; Krishnamurthy, A.; Agrawal, A. Hippocampus and Its Involvement in Alzheimer's Disease: A Review. *3 Biotech* **2022**, *12* (2), 55.

(61) Palmigiano, A.; Barone, R.; Sturiale, L.; Sanfilippo, C.; Bua, R. O.; Romeo, D. A.; Messina, A.; Capuana, M. L.; Maci, T.; Pira, F. L.; Zappia, M.; Garozzo, D. CSF N-Glycoproteomics for Early Diagnosis in Alzheimer's Disease. *J. Proteomics* **2016**, *131*, 29–37.

(62) Ding, S.; Cao, S.; Liu, Y.; Lian, Y.; Zhu, A.; Shi, G. Rational Design of a Stimuli-Responsive Polymer Electrode Interface Coupled with in Vivo Microdialysis for Measurement of Sialic Acid in Live Mouse Brain in Alzheimer's Disease. *ACS Sensors* **2017**, *2* (3), 394–400.

(63) Maguire, T. M.; Gillian, A. M.; O'Mahony, D.; Coughlan, C. M.; Dennihan, A.; Breen, K. C. A Decrease in Serum Sialyltransferase Levels in Alzheimer's Disease. *Neurobiol. Aging* **1994**, *15* (1), 99–102.

(64) Maguire, T. M.; Breen, K. C. A Decrease in Neural Sialyltransferase Activity in Alzheimer's Disease. *Dement. Geriatr. Cogn.* **1995**, *6* (4), 185–190.

(65) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* **2019**, *10* (1), 1523.

(66) Selkoe, D. J. Alzheimer's Disease Is a Synaptic Failure. *Science* **2002**, *298* (5594), 789–791.

(67) Cai, S.; Huang, K.; Yang, F.; Wang, X.; Wu, S.; Wang, Y.; Huang, L. Cortical Thickness Differences Are Associated With Chemical Synaptic Transmission Upregulated Genes in Degeneration of Mild Cognitive Impairment. *Front. Aging Neurosci.* **2021**, *13*, 745381.

(68) Vickers, J. C.; Dickson, T. C.; Adlard, P. A.; Saunders, H. L.; King, C. E.; McCormack, G. The Cause of Neuronal Degeneration in Alzheimer's Disease. *Prog. Neurobiol.* **2000**, *60* (2), 139–165.

(69) Schedin-Weiss, S.; Winblad, B.; Tjernberg, L. O. The Role of Protein Glycosylation in Alzheimer Disease. *FEBS J.* **2014**, *281*, 46–62.

(70) Yuan, J.; Yankner, B. A. Apoptosis in the Nervous System. *Nature* **2000**, *407* (6805), 802–809.

Figures

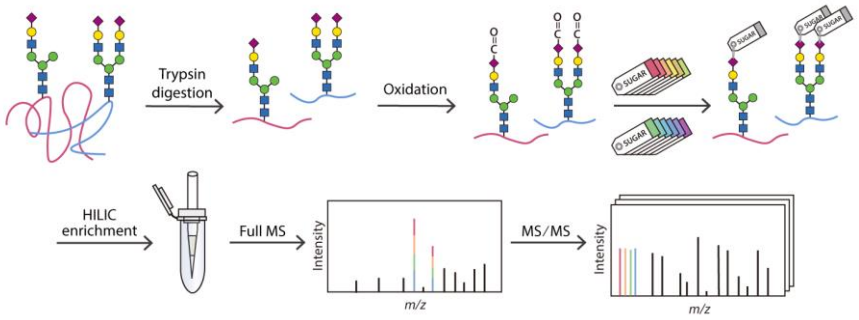


Figure 1. Workflow for multiplexed relative quantification of intact SGP from biological samples using SUGAR tags. Proteins are extracted and digested enzymatically. The resulting peptides are oxidized in a mild condition and labeled with SUGAR tags in one tube. Samples labeled with different channels are pooled after labeling for HILIC enrichment, followed by LC-MS/MS analysis.

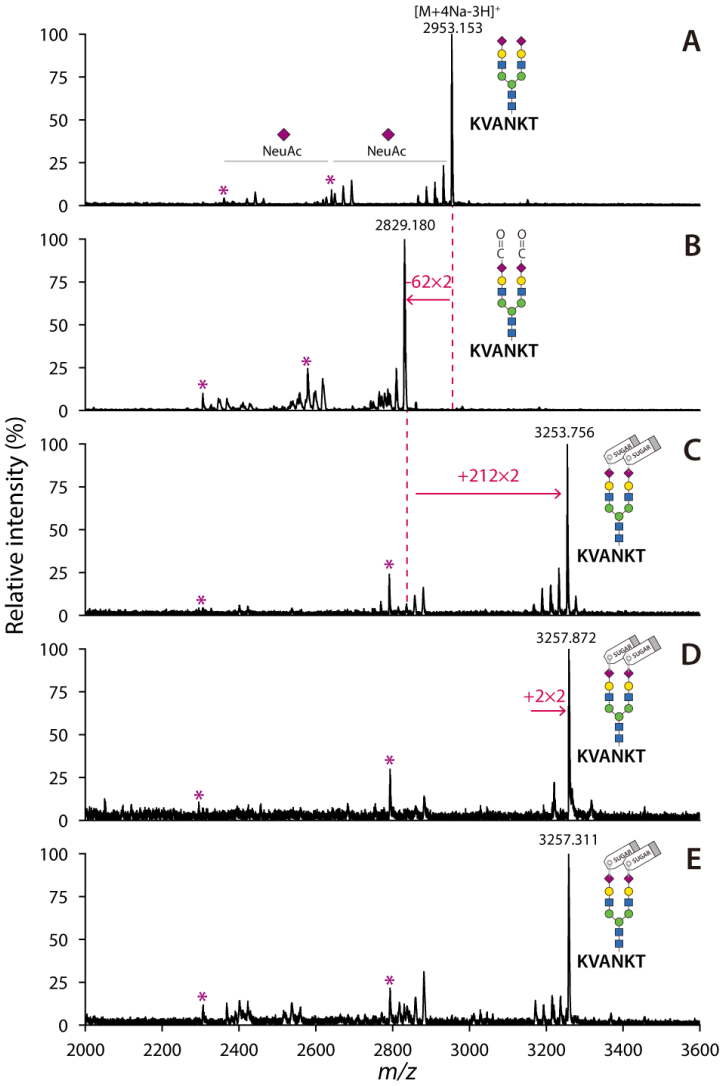


Figure 2. Test of reaction efficiency using an SGP standard. MALDI-MS spectrum of the SGP peptide standard (A) before reaction; (B) after periodate oxidation; (C) after NI SUGAR tag labeling; (D) after reductive amidation with $NaBH_3CN$; (E) after HILIC enrichment. "*" indicates neutral loss of one sialic acid residue.

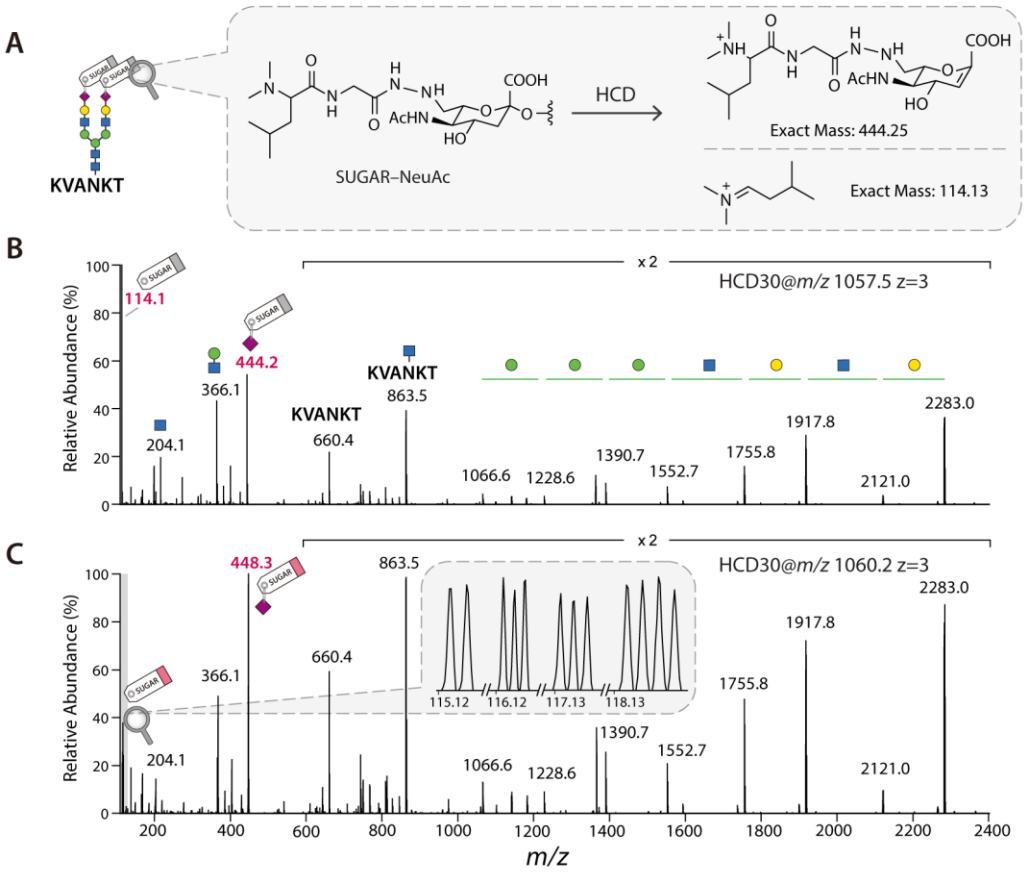


Figure 3. Evaluation of the LC-MS/MS performance of the SUGAR labeled SGP standard. (A) Structure of the labeled sialic acid and product ions generated upon HCD fragmentation. Representative HCD-MS/MS spectra for (B) NI SUGAR tag labeled SGP standard and (C) 12-plex SUGAR tag labeled SGP standard.

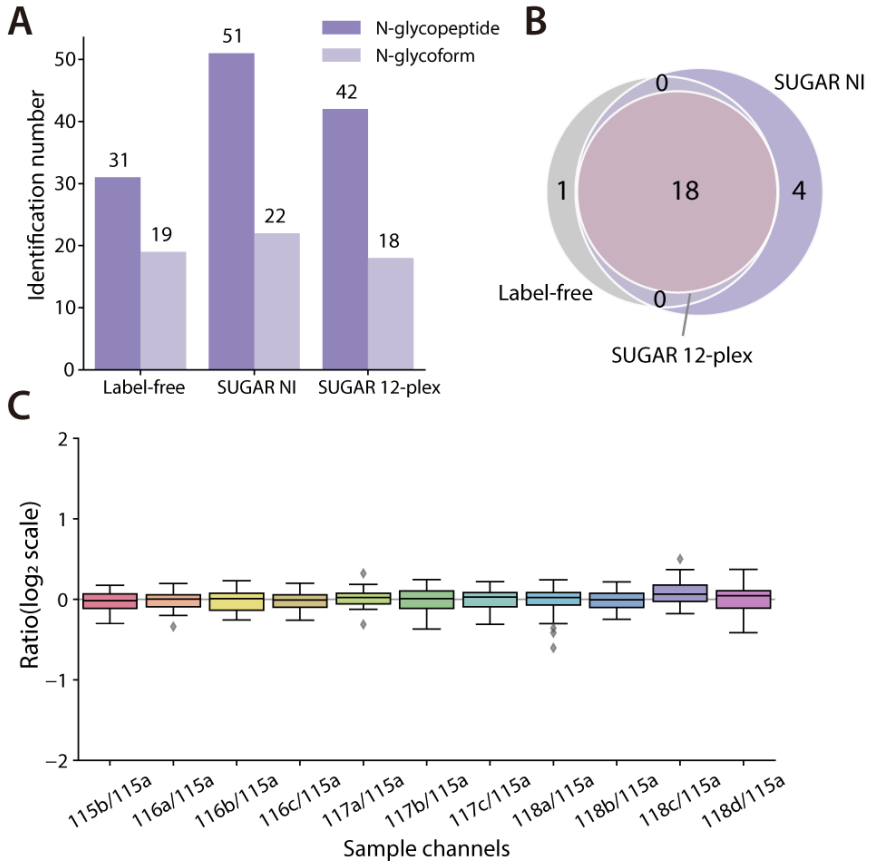


Figure 4. Evaluation of the labeling strategy with bovine fetuin tryptic digests. (A) Identification number of sialylated N-glycopeptides and N-glycoforms of the label-free, NI SUGAR-labeled, and 12-plex SUGAR-labeled samples. (B) Overlap of identified N-glycoforms among three experiments. (C) Measured quantitative ratios of N-glycopeptides labeled by 12-plex SUGAR tag and mixed at a 1:1:1:1 ratio. Boxplots indicate the median (line), the 25th and 75th percentile (interquartile range; box), and 1.5 times the interquartile range (whiskers).

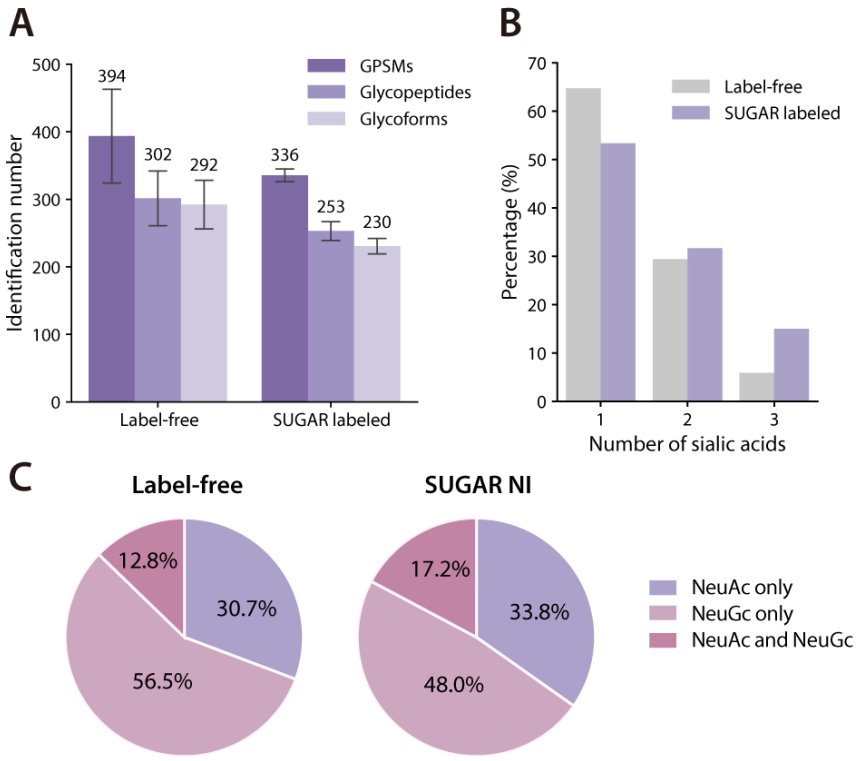


Figure 5. Comparison of SGP identification using label-free and SUGAR labeling approach. (A) The identification numbers of sialylated GPSMs, glycopeptides and glycoforms. Error bars represent technical duplicates. (B) Distribution of the number of sialic acids on identified glycans. Percentage of NeuAc- and NeuGc-containing SGPs in the (C) label-free and (D) the labeled experiment.

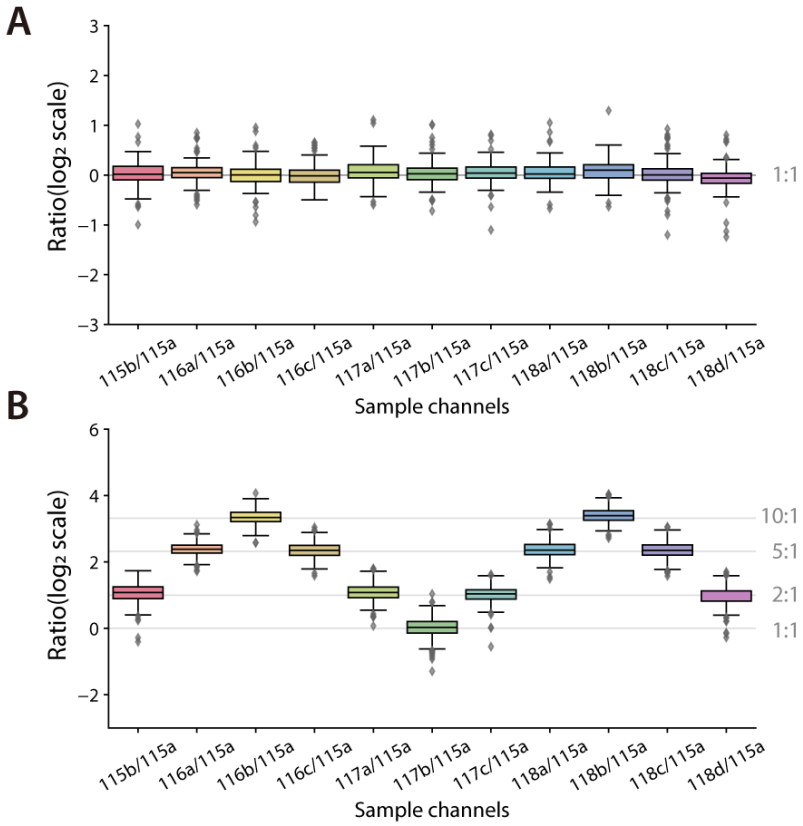


Figure 6. Quantitative performance in complex biological samples. The 12-plex SUGAR-labeled mouse heart tissue digests were combined in two known ratios of (A) 1:1:1:1 and (B) 1:2:5:10. The boxplots depict the measured ratios of each detected SGP. Boxplots indicate the median (line), the 25th and 75th percentile (interquartile range; box), and 1.5 times the interquartile range (whiskers). Gray lines indicate the theoretical ratios.

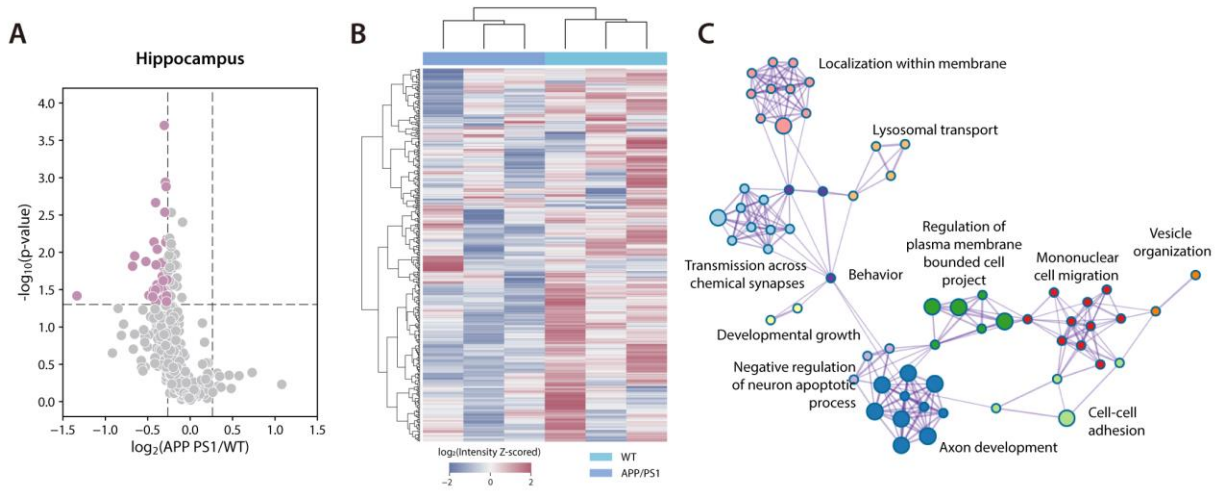


Figure 7. Quantitative intact SGP analysis between WT and APP/PS1 mice in hippocampus. (A) Volcano plot displays dysregulated SGP glycoforms. Student's *t*-test, $p < 0.05$, fold-change > 1.2. (B) Hierarchical clustering of SUGAR reporter ion intensities of all quantified sialylated glycoforms. (C) Network of biological processes enriched from proteins with significantly changed sialylated glycoforms. Each node represents an enriched term, and its size is proportional to the number of input genes fall under that term. Different terms were grouped into clusters based on their similarities, and the cluster name was annotated with the most statistically significant term.

Supplemental Information

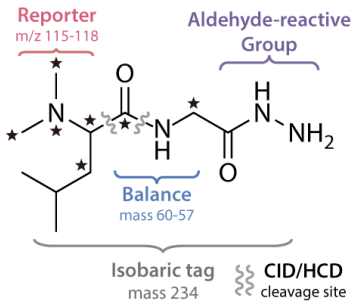
Experimental section

Offline HpH Fractionation. HpH fractionation was performed on a Waters Alliance e2695 HPLC using a 150 mm x 2.1 mm, 5mm, 100 Å, C18 column (Phenomenex) operating at 0.2 mL/min. Mobile phase A was composed of 10 mM ammonium formate in water with pH adjusted to10, and mobile phase B was 10 mM ammonium formate in 90% acetonitrile (ACN) with pH adjusted to10. The gradient used for separation was as follows: 1% B (0–5 min), 1–40% B (5–50 min), 40–60% B (50–54 min), 60–70% B (54–58 min), and 70–100% B (58–59 min). Fractions were collected every 4 minutes and 14 fractions were collected in total. Nonadjacent fractions were concatenated into 4 tubes for MS analysis.

Matrix-Assisted Laser Desorption/Ionization (MALDI)-MS analysis. SGP standards were analyzed on a Bruker Rapiflex MALDI-TOF/TOF instrument (Bruker Daltonik, Bremen, Germany) with DHB matrix (150 mg/mL in 50%ACN/H_2O(v/v), 15mM $NaHCO_3$). Ion source parameters: laser energy 80%, frequency 5000 Hz and each data collection of 2000 shots.

LC-MS/MS analysis. SGP standard and bovine fetuin samples were analyzed on an Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled with a Waters nanoAcquity UPLC (Milford, MA). Binary buffers (A, 0.1% FA; B, 100% ACN, 0.1%FA) were used in LC and the linear gradient was from 3 to 75% B for 60 min. MS survey scans of peptides were acquired from 300 to 2000 *m/z* at a resolution of 70K, using an AGC target setting of 1E6 and a maximum injection time of 50 ms. For MS2 scan, the top 15 precursor ions were selected for fragmentation by stepped higher-energy collision dissociation (HCD) with normalized collision energy (NCE) of

25 ± 5%. The MS2 data acquisition was performed at a resolution of 70K, an isolation width of 0.9 Da, a lower mass limit of *m/z* 110, an AGC target of 2E5, and an IT of 100 ms.



SUGAR Label	Stable Isotopes		Reporter Mass
	Reporter	Balance	
115a	●	● ● ●	115.12531
115b	●	● ● ●	115.13163
116a	● ●	● ●	116.12866
116b	● ●	● ●	116.13498
116c	● ●	● ●	116.14083
117a	● ● ●	●	117.13202
117b	● ● ●	●	117.13786
117c	● ● ●	●	117.14418
118a	● ● ● ●		118.13537
118b	● ● ● ●		118.14122
118c	● ● ● ●		118.14754
118d	● ● ● ●		118.15338
	● ^{13}C ● ^{2}H ● ^{15}N		

Figure S1. The structure and isotope configuration of 12-plex SUGAR tags.

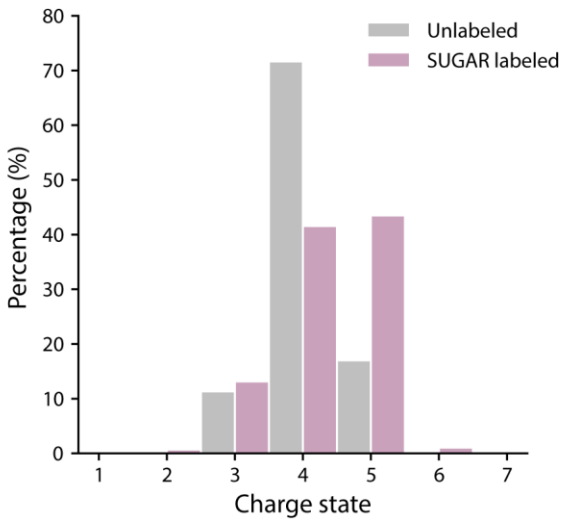


Figure S2. The charge state distribution between unlabeled and SUGAR labeled SGP standard.

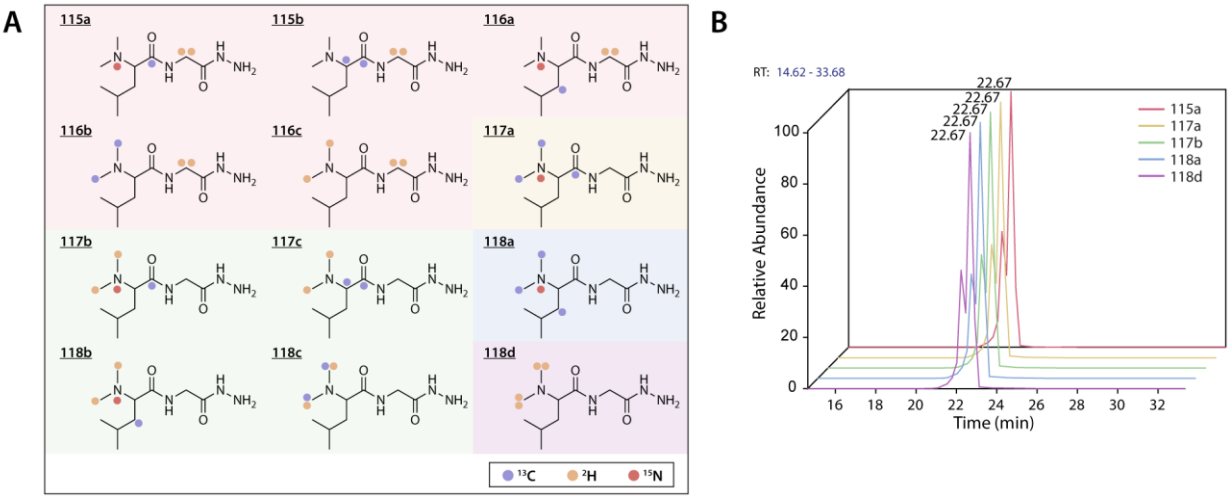


Figure S3. Evaluation of the chromatographic property of the 12 plex SUGAR tag labeled SGP. (A) 12 channels were divided into 5 groups based on the number of deuterium ions; (B) Extracted ion chromatograms (EIC) of five representative channels of the SUGAR tag labeled SGP.

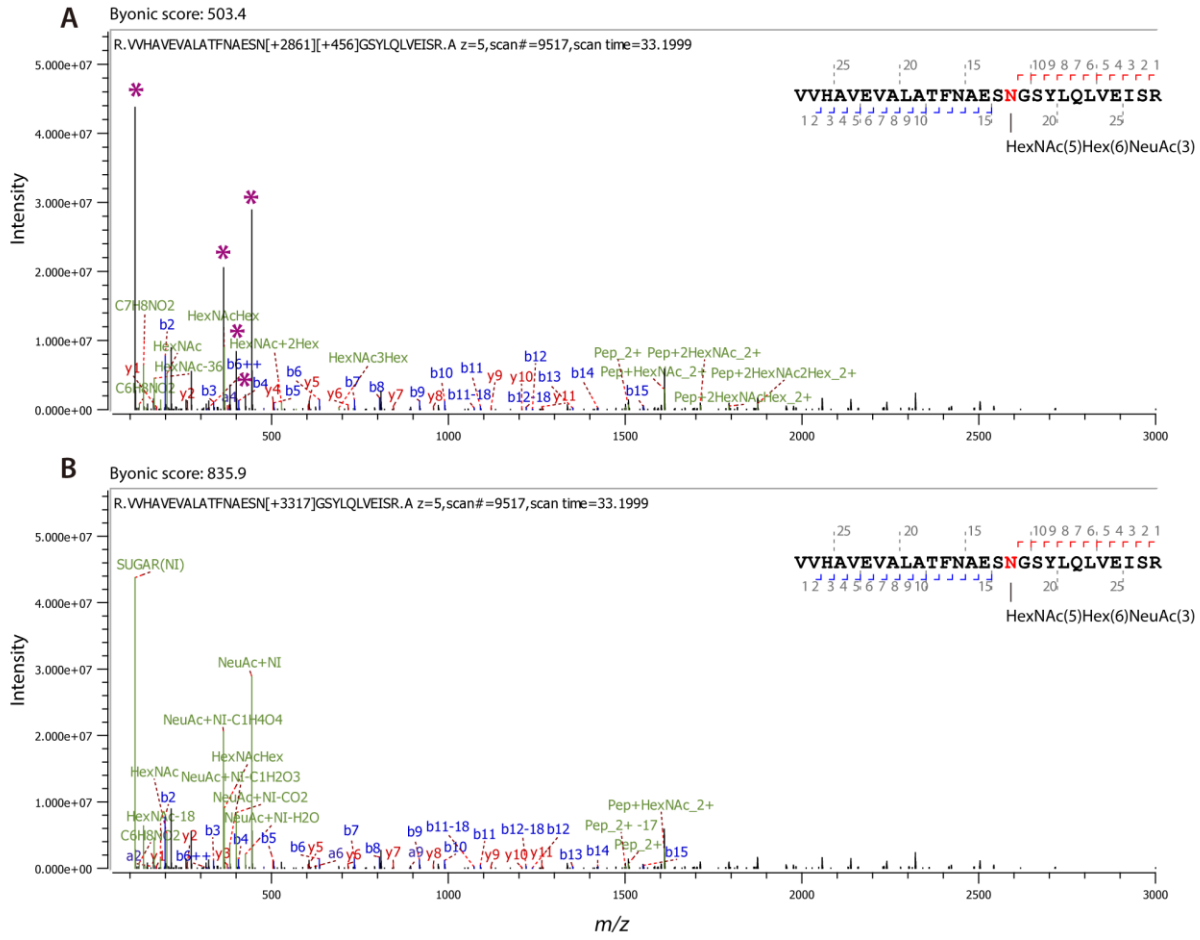


Figure S4. Representative Byonic identification result of one MS/MS spectrum from the SUGAR NI-labeled bovine fetuin digests by using (A) conventional searching approach and (B) Custom Peaks searching approach. The glycopeptide was identified with a glycan composition of HexNAc(5)Hex(6)NeuAc(3). The unrecognized reporter ion and oxonium ions in the conventional approach are annotated with "*".

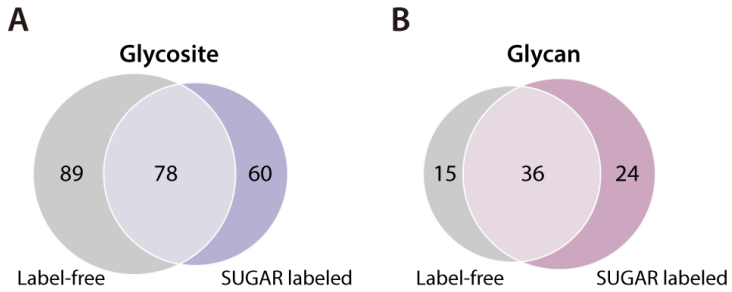


Figure S5. Overlap of (A) glycosites and (B) glycan identified in the mouse heart tissue samples.

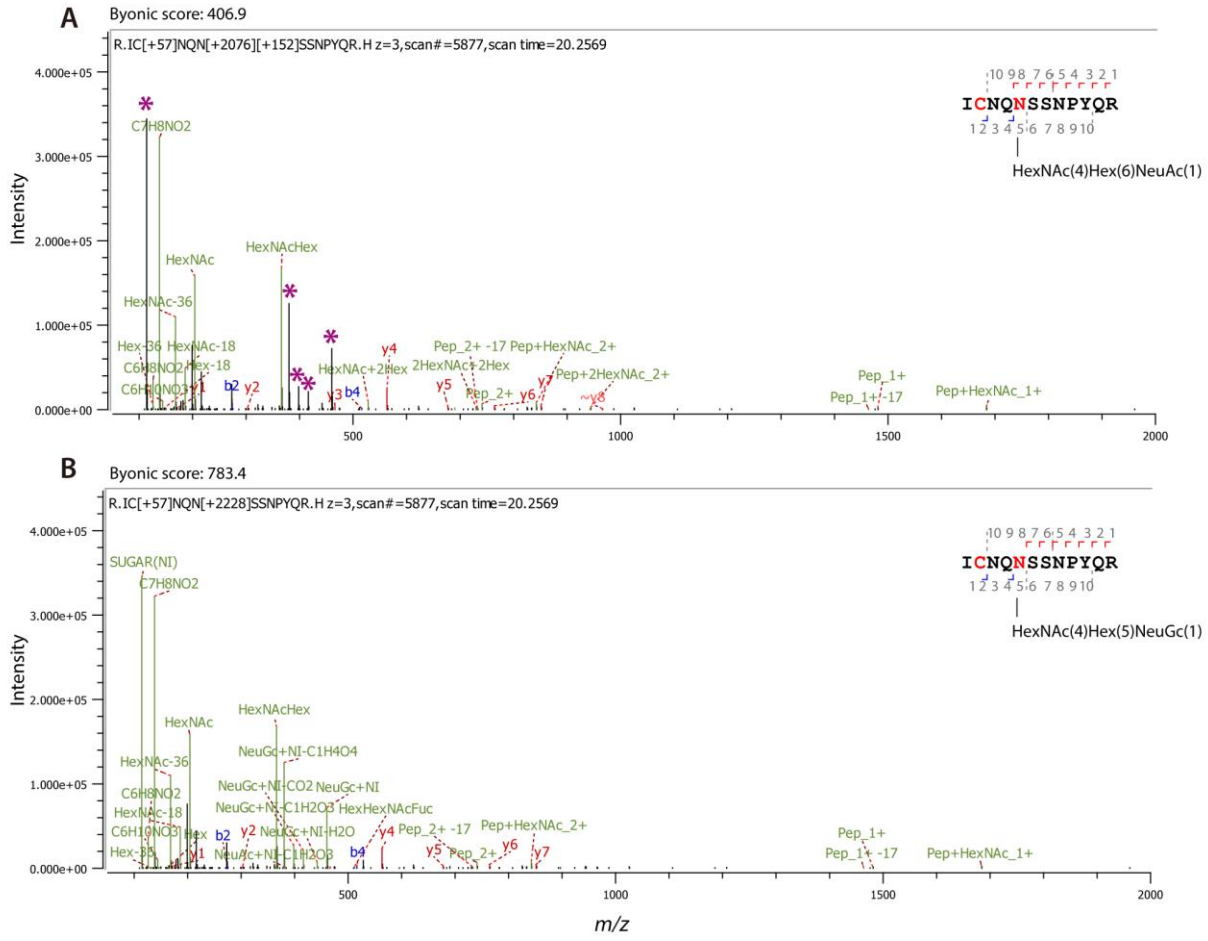


Figure S6. Representative Byonic identification result of one MS/MS spectrum from the SUGAR NI-labeled mouse heart tissue digests by using (A) conventional searching approach and (B) Custom Peaks searching approach. The conventional searching approach assigned an incorrect glycan composition of HexNAc(4)Hex(6)NeuAc(1) and ignored the sialic acid oxonium ions. By using the new searching approach, the glycan composition is correctly assigned as HexNAc(4)Hex(5)Fuc(1)NeuGc(1), as evidenced by the corresponding oxonium ions. The unrecognized reporter ion and oxonium ions in the conventional approach are annotated with "*".

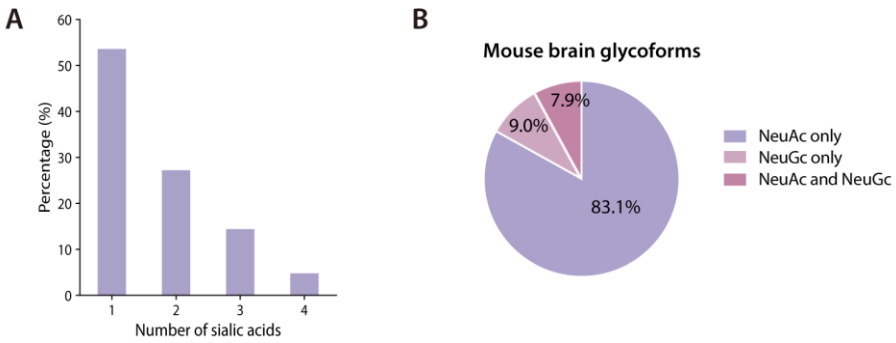


Figure S7. Identification of SGPs in mouse brain. (A) Distribution of the number of sialic acids on identified glycans. (B) Percentage of NeuAc- and NeuGc-containing SGPs.

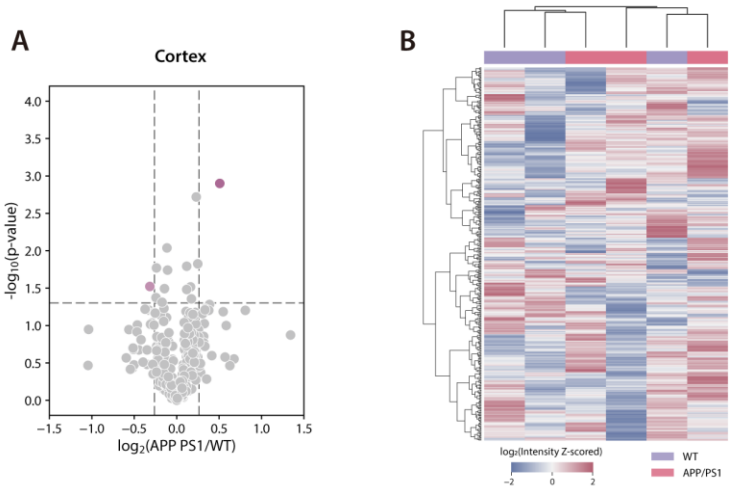


Figure S8. Quantitative intact SGP analysis between WT and APP/PS1 mice in cortex. (A) Volcano plot displays dysregulated SGP glycoforms. Student's *t*-test, $p < 0.05$, fold-change > 1.2. (B) Hierarchical clustering of SUGAR reporter ion intensities of all quantified sialylated glycoforms.

Table S1. Bovine fetuin N-glycan database.

N-glycan composition
HexNAc(4)Hex(5)NeuAc(1)
HexNAc(4)Hex(5)NeuAc(2)
HexNAc(5)Hex(6)NeuAc(1)
HexNAc(5)Hex(6)NeuAc(2)
HexNAc(5)Hex(6)NeuAc(3)
HexNAc(5)Hex(6)NeuAc(4)

Table S2. Signature ion list for SUGAR-labeled SGP

Signature ions	**NI SUGAR**	**12-plex SUGAR**
Reporter ion	114.1283	118.1354 (118a)
NeuAc+SUGAR	444.2458	448.2619
NeuAc+SUGAR-H_2O	426.2352	430.2513
NeuAc+SUGAR-CO_2	400.2560	404.2720
NeuAc+SUGAR-CO_2-H_2O	382.2454	386.2615
NeuAc+SUGAR-CO_2-$2H_2O$	364.2349	368.2509
NeuGc+SUGAR	460.2408	464.2569
NeuGc+SUGAR-H_2O	442.2302	446.2463
NeuGc+SUGAR-CO_2	416.2510	420.2670
NeuGc+SUGAR-CO_2-H_2O	398.2404	402.2565
NeuGc+SUGAR-CO_2-$2H_2O$	380.2299	384.2459

Table S3. Identified SGP glycoforms in bovine fetuin.

Sample	Glycoform
Label-free	P12763_N156_HexNAc(4)Hex(5)NeuAc(1)
	P12763_N156_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(1)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(4)
	P12763_N176_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(4)
	P12763_N99_HexNAc(4)Hex(5)NeuAc(1)
	P12763_N99_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(4)
	Q58D62_N271_HexNAc(4)Hex(5)NeuAc(2)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(2)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(3)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(4)
SUGAR NI	P12763_N156_HexNAc(4)Hex(5)NeuAc(1)
	P12763_N156_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(4)
	P12763_N176_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(4)
	P12763_N99_HexNAc(4)Hex(5)NeuAc(1)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(1)
	P12763_N99_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(4)
	Q58D62_N137_HexNAc(5)Hex(6)NeuAc(3)
	Q58D62_N137_HexNAc(5)Hex(6)NeuAc(4)

	Q58D62_N271_HexNAc(4)Hex(5)NeuAc(1)
	Q58D62_N271_HexNAc(4)Hex(5)NeuAc(2)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(2)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(3)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(4)
SUGAR 12-plex	P12763_N156_HexNAc(4)Hex(5)NeuAc(1)
	P12763_N156_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N156_HexNAc(5)Hex(6)NeuAc(4)
	P12763_N176_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N176_HexNAc(5)Hex(6)NeuAc(4)
	P12763_N99_HexNAc(4)Hex(5)NeuAc(1)
	P12763_N99_HexNAc(4)Hex(5)NeuAc(2)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(2)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(3)
	P12763_N99_HexNAc(5)Hex(6)NeuAc(4)
	Q58D62_N271_HexNAc(4)Hex(5)NeuAc(2)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(2)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(3)
	Q58D62_N271_HexNAc(5)Hex(6)NeuAc(4)

Table S4. Dysregulated SGP glycoforms in mouse brain.

Brain region	Glycoform	T: Gene name	p-value	log2(Fold change)
Cortex	Q9QXL2_N506_N6H6F1S1	Kif21a	0.0300469	-0.316729
	P43006_N205_N5H6F3S1	Slc1a2	0.00125917	0.507676
Hippocampus	Q8CBW3_N117_N5H6S2	Abi1	0.0436141	-0.389958
	P14094_N158_N5H4F1S1	Atp1b1	0.031222	-0.351966
	P14094_N158_N6H4F1S1	Atp1b1	0.0200026	-0.308453
	Q8R5M8_N116_N6H4F1S1	Cadm1	0.00289956	-0.298786
	Q8VBY2_N147_N6H5F1S1	Camkk1	0.0222248	-0.328993
	Q9JLQ0_N196_N7H8F1S4	Cd2ap	0.0489768	-0.325768
	P70232_N513_N6H7S2G1	Chl1	0.0304095	-0.375573
	Q9WUM4_N185_N9H10F1S4	Coro1c	0.0209947	-0.327956
	Q9WVJ3_N394_N6H4S1	Cpq	0.0171645	-0.303553
	P22935_N3_N5H4F1S1	Crabp2	0.0283811	-0.347553
	Q9Z218_N258_N6H4S1	Dpp6	0.00909946	-0.387095
	Q9D6F4_N157_N5H6F1S3	Gabra4	0.00019947	-0.304732
	Q6I6G8_N1114_N4H4F1S1	Hecw2	0.0382684	-1.33403
	Q6I6G8_N1114_N8H9F1S3	Hecw2	0.00215985	-0.40681
	Q60625_N397_N5H5S2	Icam5	0.0333496	-0.298514
	Q60625_N74_N5H4F2S1	Icam5	0.038432	-0.277937
	P11438_N70_N4H3S1	Lamp1	0.0301497	-0.370103
	P11438_N101_N3H5S1	Lamp1	0.00727451	-0.426086
	Q9JIA1_N277_N4H5S1	Lgi1	0.0239847	-0.318192
	P20917_N541_N7H8F1S3	Mag	0.0132791	-0.521683
	Q61885_N59_N4H5S1	Mog	0.0199793	-0.317364
	Q9QY06_N1563_N3H4S1	Myo9b	0.0395452	-0.440945
	P13595_N479_N5H4S1	Ncam1	0.0448001	-0.304599
	P13595_N450_N4H4F1S1	Ncam1	0.0111927	-0.653815
	Q9QZC2_N1084_N5H6F1S1	Plxnc1	0.0137636	-0.333542
	Q61207_N459_N3H4S1	Psap	0.0271821	-0.390608
	Q61207_N459_N5H7F1S1G1	Psap	0.00671537	-0.277282
	P97797_N302_N6H10F1S1	Sirpa	0.0297235	-0.339367
	P97797_N302_N6H9F1S2	Sirpa	0.0320308	-0.43126
	P97797_N92_N5H6S1	Sirpa	0.00131831	-0.283356
	P43006_N205_N5H5F1S2	Slc1a2	0.00731051	-0.286855
	P43006_N205_N5H6F1S1G1	Slc1a2	0.0388423	-0.264812
	P43006_N205_N5H6F1S3	Slc1a2	0.0147201	-0.401543

P60879_N77_N5H4F1S2	Snap25	0.015341	-0.678559
Q6PHU5_N160_N5H4F1S1	Sort1	0.0376812	-0.485675
Q8BG39_N516_N4H5F3S1	Sv2b	0.0167238	-0.27389
Q64332_N171_N4H5F1S2	Syn2	0.0417319	-0.318952
Q64332_N171_N6H7F1G1	Syn2	0.0319691	-0.408518
Q64332_N171_N7H6F1S4	Syn2	0.0469099	-0.405485
Q921I1_N513_N4H5G2	Tf	0.00115156	-0.292686
P01831_N118_N4H6F1S1	Thy1	0.0434186	-0.330075
P01831_N94_N3H4F1S1	Thy1	0.0312902	-0.321775
P01831_N94_N4H4F2S1	Thy1	0.0417257	-0.315683
Q8BKI2_N1571_N3H6F1S1	Tnrc6b	0.045738	-0.27827
Q62384_N57_N7H7F1S4	Zpr1	0.0233037	-0.272433

Chapter 8

Conclusions and Future Directions

Conclusions

In this dissertation, I have developed several novel mass spectrometry (MS)-based strategies for studying phosphoproteomics and glycoproteomics, both qualitatively and quantitatively. New enrichment methods were introduced to improve the separation of PTM peptides and enhance their MS detection. Additionally, new isobaric-labeling based quantification methods were developed to facilitate site-specific quantification of intact glycopeptides and explore altered glycosylation in disease models. The dissertation encompasses extensive method development and valuable datasets collected from a variety of biological and clinical samples, contributing to a deeper understanding of the diverse and essential roles of proteins and their PTMs in various biological systems.

Chapter 2–4 highlighted the development of new materials and methodologies for highly effective PTM enrichment. **Chapter 2** presented an innovative Cotton Ti-IMAC platform for phosphopeptide enrichment based on the phosphorylation modification of cotton fibers. The synthesis of Cotton Ti-IMAC was simple, rapid, and cost-effective, with broad applicability to any analytical or MS lab. The Cotton Ti-IMAC was shown to have excellent phosphopeptide enrichment performance, including extremely high selectivity and sensitivity. An enrichment specificity of 97.51% could be achieved when analyzing cell lysate samples. Additionally, the enrichment process can be conducted in various forms, such as in a packed spin-tip, a packed pipette tip, or simply in-solution mode, offering considerable flexibility for bioanalysis. Given the numerous merits of the material, we expect this tool to serve as a widely applicable and robust platform for large-scale phosphopeptide enrichment in any complex biological or clinical samples.

In **Chapter 3**, we introduced a novel epoxy-ATP-Ti^{4+} IMAC material that extended the enrichment approach from a single PTM to multiple PTMs. The material was easy and facile to prepare and possessed enhanced hydrophilicity and superior metal ion binding affinity, making it an effective stationary phase for both glycopeptide and phosphopeptide enrichment. In conventional IMAC mode, the material showed excellent selectivity and sensitivity in phosphopeptide enrichment. In a dual-functional IMAC mode, the epoxy-ATP-Ti^{4+} IMAC material enabled effective separation of N-glycopeptides and phosphopeptides, as well as sialylglycopeptides and M6P glycopeptides in complex biological samples. The material outperformed the well-established CAE-Ti-IMAC and is anticipated to serve as a useful tool for investigating the N-glycoproteome and phosphoproteome, and their potential crosstalk.

Apart from the IMAC-based enrichment strategy, **Chapter 4** explored ERLIC-based enrichment strategies and introduced new materials for VWAX chromatography. These materials, with similar binding capacities but neutralizing at lower pH compared to conventional anion exchange materials, allowed for pH gradient elution, enabling MS-compatible, desalt-free sample preparation. The optimized mobile phases and gradient enabled glycopeptide enrichment from various samples, including glycoprotein standards and cell lysates. The ERLIC-mode chromatography mechanism not only enabled separation of the glycopeptides from the bulk of unmodified peptides but also enabled simultaneous separation of neutral glycopeptides, sialyl glycopeptides, M6P glycopeptides, and sulfated glycopeptides. Detection of these negatively charged glycopeptides was enhanced, thereby improving the overall glycoproteome coverage.

Collectively, the approach simplified the experimental workflow, reduced sample loss, and provided deeper insights into less-studied glycosylation types in complex biological systems.

Chapter 5 extended the study to investigate O-glycosylation in addition to N-glycosylation. We developed an effective workflow to simultaneously analyze site-specific O-glycoproteome and endogenous peptides with PTMs in cerebrospinal fluid (CSF). The approach identified 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins from healthy individuals, representing the largest data set of site-specific O-glycoproteome study in CSF to date, including 154 novel O-glycosites reported for the first time. We also profiled the O-glycoproteome in MCI and AD patients. In addition, we identified a considerable number of CSF endogenous peptides to be O-glycosylated in all three states, while only a few notable peptides were N-glycosylated. Our findings shed light on the CSF O-glycoproteome landscape, dominant glycosylation differences, similarities during AD progression, and PTM-focused peptidome mapping.

In **Chapters 6–7**, our focus was shifted to quantitative strategies for investigating changes in PTM expression in biological samples. **Chapter 6** introduced the Boost-DiLeu strategy for enhanced isobaric tag labeled quantitative intact glycoproteomic analysis. We incorporated the one-tube sample processing workflow into DiLeu labeling experiments, which not only simplified experimental steps but also greatly reduced sample loss and increased proteome coverage. During labeling, we utilized a boosting channel consisting of relatively large amounts of biological samples analogous to samples of interest to enhance the MS signal, which further increased glycoproteome coverage and ensured quantification performance. Eventually, the strategy was successfully applied to human CSF samples from AD patients and non-AD donors, demonstrating

its feasibility for glycoproteomic analysis of quantity-limited clinical samples, which would be essential for future biomarker discovery and personalized medicine development.

Chapter 7 concentrated on sialylation quantification, presenting a high-throughput quantification strategy for intact SGP analysis using mild periodate oxidation and 12-plex SUGAR isobaric tag labeling. The optimized reaction conditions simplified the workflow and minimized sample loss. Validated with peptide standards and complex biological samples, the strategy showed promising glycoproteome coverage and quantification accuracy. Furthermore, we applied this method to explore sialylation changes between AD and WT mouse model, which provided new insights into brain region-specific PTM expression during disease progression.

Future Directions

In **Chapter 2**, we demonstrated the exceptional performance of Cotton-IMAC as a platform for phosphopeptide enrichment. Preliminary data also demonstrated its potential for enriching intact phosphoproteins from protein mixtures. In the future, this platform can be expanded to enrich phosphate-containing samples at both protein and peptide levels. These two levels of enrichment can also be combined together and developed into an *in situ* sample processing strategy, enabling sequential phosphoprotein enrichment, enzymatic digestion, phosphopeptide enrichment and isotopic labeling of phosphopeptides on the same platform, as described in a previous literature.[1] This *in situ* workflow can greatly simplify the experimental procedure and reduce sample loss and contamination. In addition, given the flexibility of Cotton-IMAC, it can be used together with other enrichment materials, such as silica materials or nanoparticles. For example,

they can be packed in a form of cotton-spin tip, facilitating multifunctional enrichment within the same experiment.

Chapter 3 introduced the simultaneous enrichment workflow for N-glycopeptides and phosphopeptides using the epoxy-ATP-Ti^{4+} IMAC material. The dual-functional Ti-IMAC material is also suitable for O-glycosylation enrichment, and a method based on dual-functional Ti-IMAC materials has been established to enrich the intact O-GalNAc glycopeptides exclusively.[2] However, several challenges must be overcome when taking both N-glycopeptides and O-glycopeptides into consideration: First, O-glycopeptides are much less abundant than N-linked glycopeptides and they have different hydrophilicity due to the different glycan sizes. The current eluting gradient might not be optimal for O-glycopeptides. Second, when searching the dataset with Byonic, more false-positive matches may occur if N-glycopeptides and O-glycopeptides are searched simultaneously. To eliminate the interference from N-glycopeptides, PNGase F deglycosylation of N-linked glycosylation is required before O-glycopeptide enrichment. Future efforts will focus on incorporating O-glycopeptide enrichment into the current workflow, including optimizing the gradient for PNGase F deglycosylated samples and exploring an improved database search strategy.

Chapter 4 introduced VWAX chromatography and the desalt-free offline fractionation method for glycopeptide enrichment. With the use of MS-compatible volatile mobile phases, this method can be further developed into an online method coupled with MS analysis directly. Moreover, the online ERLIC separation can be extended to other PTMs, such as non-canonical phosphorylation that occurs on histidine (His), arginine (Arg), lysine (Lys), aspartate (Asp),

glutamate (Glu) and cysteine (Cys) residues. Unlike the more common and stable phosphate esters (pSer, pThr, pTyr), the phosphoramidate bond in pHis, pLys and pArg, and the phosphorothioate bond of pCys are highly susceptible to hydrolysis at low pH and/or at elevated temperature, preventing their enrichment with regular IMAC methods using highly acidic buffers.[3,4] The unique features of the VWAX column allow for the direct and rapid elution of peptides into a mass spectrometer with minimal hydrolysis. As the VWAX column loses its positive charge at pH 6, multiply phosphorylated peptides can also be easily eluted, enabling more comprehensive phosphoproteome profiling compared to current strategies.

In addition, the qualitative work completed in **Chapter 2–5** can be further integrated with quantitative aspects. Specifically, in **Chapter 5**, we pooled the CSF samples in each group to reduce biological variation and enable deeper profiling of low-abundance glycoproteins. However, this approach reduced the statistical power and was insufficient for conducting quantitative analysis for in-depth glycoproteomic biomarker discovery. Future investigations will employ the isobaric reagents to label individual CSF samples to explore potential O-glycosylated or endogenous proteolytic biomarker candidates in a more quantitative manner. This approach will more efficiently evaluate patient-to-patient variations and pinpoint specific changes in glycoproteins/glycosites/glycoforms of interest.

In **Chapter 6**, we developed the Boost-DiLeu strategy with a 12-plex DiLeu tag set. A similar strategy could be adapted to the 21-plex DiLeu tags for higher multiplex capacity without apparent isotopic interference from the boosting channel.[5] Future implementation of this strategy could also

be extended to other DiLeu-labeled PTM analyses of size-limited biological samples, such as phosphorylation and citrullination in clinical specimens.[6,7]

For **Chapter 7**, there are two possible future directions: First, since our method only derivatized the *cis*-diol groups on SGPs, the carboxyl groups could be further derivatized to differentiate the sialic acid α2,3- and α2,6-linkage isomers. Combining the two derivatization strategies would enable high-throughput quantification of SGP linkage isomers. Second, while the current study focused on N-linked sialylation, preliminary data collected from glycoprotein standards suggested its feasibility on O-linked sialyl glycopeptide quantification. To date, quantification strategies for intact O-glycopeptides are not as well-developed as those for N-glycopeptides. Challenges primarily stem from the lack of effective enrichment strategies, macro-heterogeneity of O-glycosylation, and the absence of suitable software for data interpretation. Future work will explore and establish an O-glycopeptide targeted quantification strategy.

Collectively, the methodologies developed in this dissertation demonstrate significant potential for large-scale MS profiling of glycosylation and phosphorylation, as well as high-throughput quantification. It is anticipated that the continuous and expansive applications of these novel analytical approaches will ultimately contribute to the pharmaceutical and biomedical fields in the future.

Reference

(1) Huang, J.; Qin, H.; Dong, J.; Song, C.; Bian, Y.; Dong, M.; Cheng, K.; Wang, F.; Sun, D.; Wang, L.; Ye, M.; Zou, H. In Situ Sample Processing Approach (ISPA) for Comprehensive Quantitative Phosphoproteome Analysis. *J Proteome Res* **2014**, *13* (9), 3896–3904.

(2) Yue, X.; Qin, H.; Chen, Y.; Fang, Z.; Liu, L.; Zhu, H.; Liu, X.; Zhou, J.; Tian, K.; Qiao, X.; Ye, M. Highly Efficient Enrichment of O-GalNAc Glycopeptides by Using Immobilized Metal Ion Affinity Chromatography. *Anal. Chem.* **2021**, *93* (21), 7579–7587.

(3) Potel, C. M.; Lin, M.-H.; Heck, A. J. R.; Lemeer, S. Widespread Bacterial Protein Histidine Phosphorylation Revealed by Mass Spectrometry-Based Proteomics. *Nat. Methods* **2018**, *15* (3), 187–190.

(4) Hardman, G.; Perkins, S.; Brownridge, P. J.; Clarke, C. J.; Byrne, D. P.; Campbell, A. E.; Kalyuzhnyy, A.; Myall, A.; Eyers, P. A.; Jones, A. R.; Eyers, C. E. Strong Anion Exchange-mediated Phosphoproteomics Reveals Extensive Human Non-canonical Phosphorylation. *Embo J.* **2019**, *38* (21), e100847.

(5) Frost, D. C.; Feng, Y.; Li, L. 21-Plex DiLeu Isobaric Tags for High-Throughput Quantitative Proteomics. *Anal. Chem.* **2020**, *92* (12), 8228–8234.

(6) Zhong, X.; Lietz, C. B.; Shi, X.; Buchberger, A. R.; Frost, D. C.; Li, L. Highly Multiplexed Quantitative Proteomic and Phosphoproteomic Analyses in Vascular Smooth Muscle Cell Dedifferentiation. *Anal. Chim. Acta* **2020**, *1127*, 163–173.

(7) Li, Z.; Wang, B.; Yu, Q.; Shi, Y.; Li, L. 12-Plex DiLeu Isobaric Labeling Enabled High-Throughput Investigation of Citrullination Alterations in the DNA Damage Response. *Anal. Chem.* **2022**.

Appendix I

Publications and Presentations

Publications

1. **Wang, D.**; Liu, P.-K.; Gu, T.-J.; Wang, B.; Alpert, A.; Li, L. Very Weak Anion Exchange (VWAX) Chromatography for Glycopeptide Enrichment and Separation. (In preparation)
2. **Wang, D.**[#]; Li, M.[#]; Wang, Z.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, L. High-throughput Quantification of Intact Sialylated Glycopeptides Enabled by 12-plex SUGAR Isobaric Tags. (To be submitted; [#]co-first authors)
3. **Wang, D.**[#]; Haque, M.E.[#]; Rigby, M.; Gu, T.-J.; Ma, M.; Li, L.; Puglielli, L. ER Acetylation Transferases Atase 1 and Atase 2 Differentially Impact N-glycosylation. (In preparation)
4. **Wang, D.**[#]; Cui, Y.[#]; Rigby, M.; Ma, M.; Puglielli, L.; Li, L. Integrated Quantitative Proteomic, Phosphoproteomic and Glycoproteomic Characterization of APP/PS1 Mouse Brain through Electrostatic Repulsion-Hydrophilic Interaction Chromatography. (In preparation)
5. **Wang, D.**[#]; Huang, J.[#]; Zhang, H.; Gu, T.-J.; Li, L. Cotton Ti-IMAC: Developing Phosphorylated Cotton as a Novel Platform for Phosphopeptide Enrichment. (To be submitted; [#]co-first authors)
6. **Wang, D.**[#]; Huang, J.[#]; Zhang, H.; Ma, M.; Xu, M.; Cui, Y.; Shi, X.; Li, L. ATP-Coated Dual-functionalized Titanium (IV) IMAC Material for Simultaneous Enrichment and Separation of Glycopeptides and Phosphopeptides. (*J. Proteome Res.* **2023**, **accepted**; [#]co-first authors)
7. **Wang, D.**[#]; Ma, M.[#]; Huang, J.; Gu, T.-J.; Cui, Y.; Li, M.; Wang, Z.; Zetterberg, H.; Li, L. Boost-DiLeu: Enhanced Isobaric *N*,*N*-Dimethyl Leucine Tagging Strategy for a Comprehensive Quantitative Glycoproteomic Analysis. *Anal. Chem.* **2022**, 94 (34), 11773–11782. ([#]co-first authors), featured as a Supplemental Cover of the journal.
8. Chen, Z.[#]; **Wang, D.**[#]; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Yu, Q.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients. *ACS Chem. Biol.* **2022**, *17* (11), 3059–3068. ([#]co-first authors)
9. Gu, T.-J.; Liu, P.-K.; **Wang, D.**; Li, L. A Complementary Ion-based Strategy Enables Multiplexed Quantification in ETD Analysis Using DiLeu Isobaric Tags. (In preparation)
10. Lu, H.; Wang, B.; Liu, Y.; **Wang, D.**; Fields, L.; Zhang, H.; Li, M.; Shi, X.; Zetterberg, H.; Li, L. DiLeu isobaric labeling coupled with limited proteolysis mass spectrometry for high-throughput profiling of protein structural changes in Alzheimer's disease. *Anal. Chem.* **2023**. (Under review)
11. Gu, T.-J.; Feng, Y.; **Wang, D.**; Li, L. Simultaneous Multiplexed Quantification and C=C Localization of Fatty Acids with LC-MS/MS Using Isobaric Multiplex Reagents for Carbonyl-Containing Compound (SUGAR) Tags and C=C Epoxidation. *Anal. Chim. Acta* **2022**, 340215.

12. Tabang, D. N.; **Wang, D.**; Li, L. A Spin-Tip Enrichment Strategy for Simultaneous Analysis of N-Glycopeptides and Phosphopeptides from Human Pancreatic Tissues. *J. Vis. Exp.* **2022**, No. 183, e63735.

13. Huang, J.; **Wang, D.**; Shipman, R. D.; Zhu, Z.; Liu, Y.; Li, L. Simultaneous Enrichment and Separation of Neutral and Sialyl Glycopeptides of SARS-CoV-2 Spike Protein Enabled by Dual-Functionalized Ti-IMAC Material. *Anal. Bioanal. Chem.* **2021**, 1–9.

14. Huang, J.; Liu, X.; **Wang, D.**; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576.

Book Chapter

1. **Wang, D.**; Chen, Z.; Li, L. Profiling CSF Endogenous Peptidome in Alzheimer's Disease. In *Peptidomic. Methods in Molecular Biology*, 2nd ed.; Humana Press, **2023**. (Submitted)

Conference Presentations

1. **Wang, D.**; Liu, P.-K.; Gu, T.-J.; Wang, B.; Alpert, A.; Li, L. Very Weak Anion Exchange (VWAX) Chromatography for Glycopeptide Enrichment and Separation. *71th ASMS Conference*, Houston, TX, June 2023, Poster.

2. **Wang, D.**; Li, M.; Wang, Z.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, L. High-throughput Quantification of Intact Sialylated Glycopeptides Enabled by 12-plex SUGAR Isobaric Tags. *19th Annual US HUPO Conference*, Chicago, IL, March 2023, **Oral**.

3. **Wang, D.**; Huang, J.; Zhang, H.; Ma, M.; Xu, M.; Cui, Y.; Shi, X.; Li, L. ATP-Coated Dual-functionalized Titanium (IV) IMAC Material for Simultaneous Enrichment and Separation of Glycopeptides and Phosphopeptides. 2022 Wisconsin Human Proteomics Symposium, Madison, WI, Aug 2022, Poster.

4. **Wang, D.**; Huang, J.; Zhang, H.; Ma, M.; Xu, M.; Cui, Y.; Shi, X.; Li, L. ATP-Coated Dual-functionalized Titanium (IV) IMAC Material for Simultaneous Enrichment and Separation of Glycopeptides and Phosphopeptides. *70th ASMS Conference*, Minneapolis, MN, June 2022, Poster.

5. **Wang, D.**; Ma, M.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, M.; Wang, Z.; Zetterberg, H.; Li, L. Boost-DiLeu: Enhanced Isobaric *N*,*N*-Dimethyl Leucine Tagging Strategy for a Comprehensive Quantitative Glycoproteomic Analysis. *69th ASMS Conference*, Philadelphia, PA, Nov 2021, **Oral**.

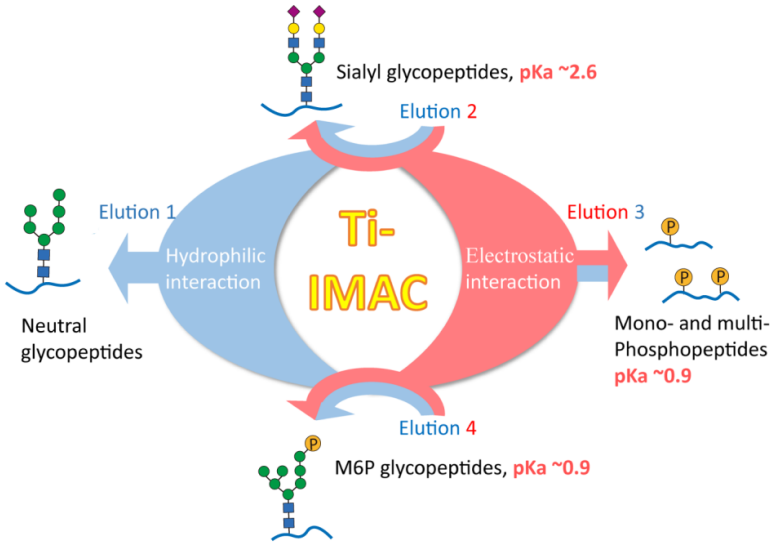
6. **Wang, D.**; Huang, J.; Zhang, H.; Ma, M.; Xu, M.; Cui, Y.; Shi, X.; Li, L. ATP-Coated Dual-functionalized Titanium (IV) IMAC Material for Simultaneous Enrichment and Separation of Glycopeptides and Phosphopeptides. *1st CASMS Conference on Advancing Biological and Pharmaceutical Mass Spectrometry*, Virtual, August 2021, Poster.

7. **Wang, D.**; Huang, J.; Ma, M.; Li, L. Hydrophilic enhanced dual-functionalized Titanium (IV) IMAC material for enrichment and separation of glycopeptides and phosphopeptides. *68th ASMS Conference*, Virtual, June 2020, Poster.

8. **Wang, D.**; Huang, J.; Liu, Y.; Cui, Y.; Johnson, J.; Li, L. Large-scale profiling of mannose-6-phosphate glycoproteome from human cells by Ti(IV)-IMAC, *67th ASMS Conference*, Atlanta, GA, June 2019, Poster.

Appendix II

Dual-functional Ti(IV)-IMAC material enables simultaneous enrichment and separation of diverse glycopeptides and phosphopeptides



Adapted from: Huang, J.; Liu, X.; **Wang, D.**; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L. Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* **2021**, *93* (24), 8568–8576. D.W. contributed to sample preparation, data collection, and figure preparation.

Abstract

Simultaneous enrichment and fractionation of diverse proteins/peptides possessing different post-translational modifications (PTMs) from the same biological samples is highly desirable to reduce sample consumption, avoid complicated sample processing, and enable studies of potential crosstalks between different PTMs. In this work, we report a new approach to enable the simultaneous enrichment and separation of glycopeptides, phosphopeptides and mannose-6-phosphate (M6P) glycopeptides by using dual-functional Ti(IV)-IMAC material. Moreover, we also made the separation of neutral and sialyl glycopeptides, mono- and multi-phosphopeptides possible by performing different elution according to the differences in their electrostatic or hydrophilic properties. These separations are effective and efficient to eliminate the signal suppression from neutral glycopeptides for sialyl glycopeptide detection, allowing separation of mono-phosphopeptides from multi-phosphopeptides, as well as detection of M6P glycopeptides that are free from the above-mentioned modifications. This new strategy significantly improves the coverage and identification numbers of glycopeptides, phosphopeptides and M6P glycopeptides by 1.9, 2.3, and 4.3-fold compared with the conventional method, respectively. This is the first report on simultaneous enrichment and separation of neutral and sialyl glycopeptides, mono- and multi-phosphopeptides and M6P glycopeptides via dual-functional Ti(IV)- IMAC, revealing novel insights into potential crosstalk among these important PTMs.

Introduction

Protein post-translational modifications (PTMs) are covalent and generally enzymatic modifications on specific protein sites, which increase the functional diversity of the proteins and

affect almost all aspects of biological processes of a cell.[1] It is reported that numerous proteins can bear multiple PTMs, which suggests the possibility of regulatory crosstalk between unique modification events.[2] Liquid chromatography tandem mass spectrometry (LC-MS/MS) based proteomics strategy has become the crucial tool for the analysis of protein PTMs. To comprehensively profile protein PTMs, an efficient enrichment method is a prerequisite due to the typical low abundance and sub-stoichiometry of most protein PTMs in complex biological samples.[3] Numerous enrichment methods have been developed, however, each method may only focus on one type of PTM.[4] For example, immobilized metal affinity chromatography (IMAC) has been developed for phosphopeptide enrichment and hydrophilic interaction chromatography (HILIC) has been utilized primarily for enrichment of glycopeptides.[5,6] Crosstalk analysis of different PTMs must perform different enrichment methods on different samples, which is not only tedious and time-consuming but also requires more precious samples that may not be available.[3] Carr and coworkers developed a PTM crosstalk analysis strategy called serial enrichments of different post-translational modifications (SEPTM) that enabled enrichment of multiple PTMs from the same biological sample.[7] Similar strategy have also been applied to enrich glycopeptides and phosphopeptides from plasma-derived extracellular vesicles.[8] However, these methods still suffer from the requirement of different enrichment materials. Several research groups have developed new IMAC materials to simultaneous enrich and separate glycopeptides and phoshopetides.[9-12] However, without proper fractionation, negatively charged sialylated glycopeptides will be supressed by more abundant neutral glycopetides in the positive ion mode

LC-MS/MS analysis, and multi-phosphopeptides and M6P glycopeptides will be supressed by mono-phosphopeptides as well.

Previously, we demonstrated that M6P glycopeptide identifications increased significantly when using the dual-functional Ti-IMAC materials for enrichment while simultaneously eliminating the interference of phosphopeptides. [13] This strategy is based on the special properties of the Ti(IV)-IMAC materials we developed, which not only possesses substantial phosphate chelated Ti(IV) ions, enabling enrichment of phosphopeptides through electrostatic interaction, but also contain numerous hydroxyl, amine and phosphate groups on the material surface that are highly hydrophilic, enabling enrichment of glycopeptides through hydrophilic interactions.[14,15] As M6P glycopeptides possess features of both phosphopeptides and glycopeptides, they could be captured by the dual-functional Ti(IV)-IMAC material through the synergistic electrostatic and hydrophilic interactions. Therefore, by performing HILIC-mode stepwise elution, M6P glycopeptides can be separated from phosphopeptides. However, in this strategy, glycopeptides were discarded together with the non-phosphopeptides.

In order to overcome the shortcomings of the above methods, in the current study, we improved and optimized the previous strategy and continually explored the hydrophilic characteristics of the Ti(IV)-IMAC material to enable simultaneous enrichment and separation of neutral and sialyl glycopeptides, mono- and multi-phosphopeptides and M6P glycopeptides (**Figure 1**). This strategy improves the quality of each individual PTM analysis significantly and can be effectively implemented and integrated to perform crosstalk analysis of these three important PTMs in a single experiment from the same biological sample.

Experimental section

Sample preparation and protein digestion. Details are provided in **Supporting Information**.

HILIC-mode Ti(IV)-IMAC enrichment and elution. Typical enrichment and elution method for Ti(IV)-IMAC and HILIC are provided in **Supporting Information.** For HILIC-mode Ti(IV)-IMAC enrichment the standard (10 μg each sample) and mouse lung protein digests (500 μg each sample) were loaded onto Ti(IV)-IMAC materials with 80% ACN/3%TFA buffer and then Ti(IV)-IMAC materials captured samples were washed three times with the loading buffer, which was similar with the conventional HILIC enrichment. The flowthrough and washing buffer of the standard proteins were collected for MALDI MS analysis. After enrichment, the standard protein samples were eluted with following two elution procedures: 1)10% NH_4OH (v/v) 10 min; 2) 0.1% FA (v/v) 10 min and 10% NH_4OH (v/v) 10min, respectively. Samples were immediately dried down in SpeedVac. Samples were re-dissolved in 0.1% formic acid (FA) and load 5% of each sample for MALDI analysis (details are provided in **Supporting Information**).

For mouse lung protein sample, the elution was also performed with two methods: 1) direct eluted with 10% NH_4OH (v/v); 2) sequentially eluted with the following four elution procedures: Elution 1 uses weak acidic buffer (0.1 TFA%) with decreased ACN gradient (1^{st} fraction 60%, 2^{nd} fraction 40%, and 3^{rd} fraction 20% ACN (v/v)); Elution 2 uses strong acidic condition (6 TFA%) with decreased ACN gradient (40%, 20%, and 0% ACN (v/v)), the three elution were combined for analysis (4^{th} fraction); elution 3 uses basic buffer (10% NH_4OH) with high and medium ACN content (5^{th} fraction 60% and 6^{th} fraction 40% ACN (v/v)); elution 4 uses basic buffer (10% NH_4OH) with low ACN content (20%, 10%, and 0% ACN (v/v)), the three elution were combined

for analysis (7^{th} fraction). The elution was immediately dried down in SpeedVac. Samples were stored at -20 ^{o}C and re-dissolved in 1% formic acid (FA) and load 15% of each sample for LC-MS/MS analysis (details are provided in **Supporting Information**). It is worth mentioning that the number of the elution steps was determined based on the sample complexity. For standard samples containing only a few neutral glycopeptides and several phosphopeptides, 2-step elution was sufficient. For complex biological samples like mouse lung protein digests, 4 or more steps of elutions were needed to enable better separation and fractionation of multiple PTMs peptide.

Data processing. Byonic software (Protein Metrics, San Carlos, CA) was used to analyze the acquired MS and MS/MS spectra of intact N-linked glycopeptides and M6P glycopeptides. Raw files were searched against *Mus musculus* protein database of reviewed (Swiss-Prot) sequences downloaded from Uniprot. Precursor ion mass tolerance of 10 ppm and fragment ion mass tolerance of 0.01 Da were selected. The phosphorylation of serine (S), threonine (T) and tyrosine (Y) and the oxidation of methionine (M) were set as variable modifications. Meanwhile, the carbamidomethylation of cysteine (C) was set as fixed modification. Common N-linked glycopeptides searching used a mammalian N-glycome database that contains 309 glycans in addition to M6P glycans. Peptide identifications were filtered at two-dimensional false discovery rate (2D FDR) <1%, PEP 2D <0.05, |Log Prob|>1, and Byonic Score >150. Manual inspection of MS/MS spectra of M6P glycopeptides was performed to examine if Byonic identification results contained phosphorylated hexose diagnostic ions. With the same raw files and database, phosphopeptides were searched against MaxQuant software (Version, 1.5.8.3). Precursor ion mass tolerance of 4.5 ppm and fragment ion mass tolerance of 0.05 Da were selected. Phosphopeptides

with the false discovery rate <0.01 and minimum score of 40 were accepted as confident identifications.

Results and Discussion

Conventional IMAC and HILIC approaches for glycopeptide, phosphopeptide and M6P glycopeptide enrichment. In general, IMAC is suitable for phosphopeptide and M6P glycopeptide enrichment as they both contain phosphate group and HILIC is applicable to enrich common glycopeptides and M6P glycopeptides as both types of peptides possess hydrophilic glycans. To investigate glycopeptide, phosphopeptide and M6P glycopeptide profiling status by these two conventional methods, we first performed typical Ti(IV)-IMAC and HILIC (homemade spin-tip) enrichment procedures using mouse lung tissue samples.[16,17] As shown in **Figure 2A**, the conventional IMAC method enabled profiling of 2404 phosphopeptides, however, only resulted in the identification of 7 common N-glycopeptides and 11 M6P glycopeptides from 25 μg mouse lung protein digests. This result suggested that majority of glycopeptides cannot be effectively enriched by Ti(IV)-IMAC materials using its typical enrichment procedures. Furthermore, the M6P glycopeptide signals can be suppressed by the coeluted phosphopeptides in LC-MS/MS analysis. As comparison, conventional HILIC method enabled profiling of 2014 common N-glycopeptides and 15 M6P glycopeptides, however, only identified 233 phosphopeptides from the same amount of sample (**Figure 2B**). This result demonstrated that only a small fraction of phosphopeptides were enriched by HILIC materials due to their low hydrophilicity and the M6P glycopeptide signals were also suppressed by the coeluted common N-glycopeptides in LC-MS/MS analysis. Therefore, development of a new enrichment strategy to

simultaneously enrich and separate glycopeptides, phosphopeptides, and M6P glycopeptides from the same biological samples would overcome ion suppression issue during MS analysis, which is highly beneficial for the global mapping of these PTMs.

Exploring hydrophilicity of the Ti(IV)-IMAC material. The Ti(IV)-IMAC material we developed not only possesses substantial phosphate chelated Ti(IV) ions, which can be used to capture phosphopeptides through electrostatic interactions, but also contains numerous hydroxyl, amine and phosphate groups on the material surface. We previously demonstrated the high hydrophilicity of Ti(IV)-IMAC material and used this characteristic to separate the Ti(IV)-IMAC enriched phosphopeptides and M6P glycopeptides.[13] However, this strategy was not able to enrich common glycopeptides due to non-HILIC mode loading procedures. To overcome this problem, we continually optimize the loading procedure and change it to HILIC mode. We aim to simultaneously capture glycopeptides during the phosphopeptide and M6P glycopeptide enrichment.

To test this possibility, we first used standard phosphoprotein β-casein and standard glycoprotein RNase B digests to perform the enrichment. Standard M6P glycoprotein was not found commercially available, so it was not included in the initial test experiment. The sample loading buffer was changed from 40% ACN/3%TFA to 80% ACN/3%TFA, which was similar with the conventional HILIC enrichment protocols. The increased ACN content can help to generate an aqueous layer across the substantial hydroxyl, amine groups and phosphate chelated Ti (IV) ions on the material surface, therefore, the glycopeptides can be captured on materials through hydrophilic interaction. Initially, the aqueous buffer of 10% ammonia was used as a single

elution buffer because it can interrupt both electrostatic and hydrophilic interactions and should be able to elute captured phoshopeptides and glycopeptides simultaneously.

The flow-through and eluted samples were analyzed by MALDI-TOF MS and the results were shown in **Figures 3A** and **3B**. **Figure 3A** showed that flow-through did not reveal the trace of phosphopeptides and glycopeptides, which demonstrated that majority of these modified peptides were captured on the materials. And in **Figure 3B**, the two intense peaks labeled with red pound signs can be annotated as the two phosphopeptides of β-casein. There are also some peaks that can be annotated to the glycopeptides of RNase B, however, their intensities are very low. We infer that the glycopeptides of RNase B are either not effectively enriched by Ti (IV)-IMAC materials or their signals are greatly supressed by the coeluted phosphopeptides. To evaluate these possibilities, we performed another elution experiment, that is the glycopeptides and phosphopeptides were eluted separately. As 0.1% FA can interrupt the hydrophilic interactions, we used it to elute the glycopeptides of RNase B, and then 10% ammonia was used to elute phosphopeptides. **Figure 3C** was similar with **Figure 3A** and no trace of phosphopeptides and glycopeptides was found. As shown in **Figure 3D**, significantly enhanced and increased glycopeptide peaks were observed with the 0.1% FA elution. And in **Figure 3E**, we detected the two phosphopeptides of β-casein without signals from glycopeptides (Annotated MALDI-TOF MS peaks were also shown in **Table S1**). These results demonstrated that Ti (IV)-IMAC materials can be used to simultaneously enrich glycopeptides and phosphopeptides from the same sample. And separating the glycopeptides from the phosphopeptides after enrichment can increase the MS signals for glycopeptides and improving their detection and coverage.

Applying dual-functional Ti(IV)-IMAC material to enrichment of PTM peptides from complex biological samples. After demonstrating the potential of dual-functional Ti(IV)-IMAC material for the enrichment and separation of both glycopeptides and phosphopeptides. We then applied this novel strategy to the enrichment of N-glycopeptides, phosphopeptides and M6P glycopeptides from complex biological samples consisting of mouse lung protein digests. Initially, the samples were loaded onto the Ti(IV)-IMAC materials with conventional HILIC mode. In this way, neutral glycopeptides were captured on the materials through hydrophilic interaction, phosphopeptides were captured with electrostatic interaction and sialyl and M6P glycopeptides were captured by the synergistic electrostatic and hydrophilic interaction. We then performed stepwise elution of these captured peptides. Generally, as shown in **Figure 4A**, in Elution 1 (fractions 1-3), the numbers of identified glycopeptides were 888, 1630, and 1705, respectively, and majority of them were neutral glycopeptides. In Elution 2 (4^{th} fraction), 1787 glycopeptides were identified, while the glycopeptides with sialylated glycan were continually increased to 834. Higher number of sialylated glycopeptides was profiled in this fraction. This observation is likely due to the fact that the pKa value of sialic acid is 2.1, therefore sialic acids may not be fully protonated in low TFA elution buffer and they can still be captured by Ti(IV)-IMAC materials through electrostatic interaction. However, if we increased the TFA content from 0.1% to 6%, more sialic acids can be protonated and continually be eluted. It is worth mentioning that this strong acid condition did not elute much phosphopeptides and M6P glycopeptides, as the pKa value of phosphate group is 0.7-1.0, which is much lower than sialic acid and is hard to be protonated.

In fractions 1-4, in which majority of glycopeptides were identified, we found that the ratios of sialylated glycopeptides in all identified glycopeptides were 3.4%, 16.4%, 40.3%, and 46.7%, respectively, and the ratios were gradually increased in these four fractions. Moreover, the ratios of multi-sialylated glycopeptides in all identified sialylated glycopeptides in these four fractions were 10.0%, 7.1%, 23.7%, and 40.7%, respectively, and the ratios also have a general trend of increasing (**Figure 5A, Figure S1**). This is because the sialic acid is more hydrophilic than other monosaccharides and it is also negatively charged, eluting sialylated glycopeptides from dual-functional Ti(IV)-IMAC material is more difficult than other neutral glycopeptides due to stronger retention. The increasing trend of sialylated and multi-sialylated glycopeptide ratio corresponded well to the decreasing ACN content, increasing TFA content. These results demonstrated that the dual-functional Ti(IV)-IMAC materials can not only effectively enrich glycopeptides with excellent efficiency, it can also separate sialylated glycopeptides from neutral glycopeptides. As sialylation has been widely implicated in various biological processes. Therefore, comprehensive sialylation analyses will help to elucidate their roles in the related biological process and facilitate the discovery of novel biomarkers for diseases and drug targets.[18] We assume that by performing negative ion mode MS,[19] or explore sialic acid derivatization strategy[20-22] and combine these approaches with our high efficiency enrichment and separation strategy, the coverage of sialylated glycopeptide analyses will continually be improved.

As a comparison, we also performed fractionation for conventional HILIC enriched glycopeptides. In total, four fractions were collected, and the elution buffers all contained 0.1% FA and gradient decrease of ACN content at 60%, 40%, 20%, and 0%, respectively. The identified

glycopeptides were 1684, 1404, 711, and 188, respectively, and the sialylated glycopeptides were 691, 572, 183, and 22 (**Figure S2A**). The data showed that conventional HILIC did not fractionate the glycopeptides like the dual-functional Ti-IMAC done. The greatest number of sialylated glycopeptides were identified in the first fraction and it decreased with the reduction of ACN content and the total number of glycopeptides in other three fractions, the sialylated glycopeptides ratio have the same trends (**Figure S2B**), which showed opposite trends with the dual-functional Ti-IMAC fractionation. This demonstrated that sialylated glycopeptides were not separated well from neutral glycopeptides in conventional HILIC fractionation. Moreover, the identified M6P glycopeptides in the four fractions were 15, 9, 2, and 0, respectively, which was also lower than the dual-functional Ti-IMAC approach profiled (**Figure S2C**).

After eluting most glycopeptides, materials were resuspended in 80% ACN to recover the aqueous layer and continually eluted with gradient ACN containing ammonia. In Elution 3 (5^{th} and 6^{th} fraction), only the majority of phosphopeptides were eluted, the results in **Figure 4B** demonstrated that in the 5^{th} fraction, 3883 phosphopeptides were profiled, of which only 150 were multi-phosphopeptides, while in the 6^{th} fraction, 2634 phosphopeptides were identified and the identified multi-phosphopeptides were increased to 703. In Elution 4 (7^{th} fraction), the ACN content was continually decreased to the aqueous elution buffer, in this fraction, the total number of identified phosphopeptides was decreased to 1504, with 489 of them being multi-phosphopeptides. In the 5-7^{th} fractions, the multi-phosphopeptides ratios were 3.9%, 26.7%, and 32.5%, respectively. It was contrary to the decreasing trend of phosphopeptides, the ratio of multi-phosphopeptides increased gradually (**Figure 5B**). This counter-intuitive result is likely due to the

strong hydrophilicity of the phosphate group on phosphopeptides,[23] and in general multi-phosphopeptides are more hydrophilic than mono-phosphopeptides, which will enhance their retention on materials under HILIC conditions.[24] These results demonstrated dual-functional Ti(IV)-IMAC materials can not only effectively enrich phosphopeptides with high efficiency, but can also separate multi-phosphopeptides from mono-phosphopeptides according to their differences in hydrophilicity. Comprehensive isolation and characterization of both mono- and multi-phosphopeptides is important for understanding the graded regulation of phosphorylation signal that can more precisely modulate the switch of signal transduction pathways.[25-27] Previously, separation of multi-phosphopeptides from mono-phosphopeptides was achieved by applying two enrichment materials with different mono- and multi-phosphopeptide selectivity.[28-30] However, these methods cost additional enrichment material and often lead to greater sample loss during the sample transfer.[31] Therefore, our simultaneous enrichment and separation strategy through dual-functional Ti(IV)-IMAC materials have unique advantages to enable more comprehensive profiling of mono- and multi-phosphopeptides. It is worth noting that, in the 7th fraction, we also identified 361 glycopeptides with 69.0% being sialylated glycopeptides and 61.9% being multi-sialylated glycopeptides (**Figure 4A, Figure 5A**), which were remnants of strong acid elution and were continually eluted by aqueous and basic elution conditions.

In Elution 1-3, only a handful of M6P glycopeptides were identified, if any. However, in Elution 4, we profiled 57 M6P glycopeptides (**Figure 4C**). These results demonstrated that M6P glycopeptides have the strongest retention on Ti(IV)-IMAC materials due to the unique properties of this class of glycopeptides possessing both hydrophilic glycans and negatively charged

phosphate groups. Only the highest elution strength with aqueous buffer plus electrostatic interaction interruption reagent could effectively elute these peptides. Therefore, the separation of M6P glycopeptides from common N-glycopeptides and phosphopeptides based on their differences in hydrophilicity and electrostatic interactions using the dual-functional Ti(IV)-IMAC materials can be achieved with systematic optimization.

Through this sequential elution and separation, we identified 3896 N-glycopeptides, 5539 phosphopeptides, and 65 M6P glycopeptides in mouse lung and the identified glycopeptides, phosphopeptides and M6P glycopeptides by the new strategy were 1.9, 2.3, and 4.3 times higher than those achieved using the conventional method, respectively (**Figure 4D**). And we also performed the HILIC mode Ti-IMAC without fractionation, the result was similar with the standard protein result. The identified phosphopeptides and M6P glycopeptides were similar with the conventional Ti-IMAC method. However, the enriched N-glycopeptides only increased to 295, which was much lower than the conventional HILIC method and demonstrated that they were also suppressed by the co-eluted phosphopeptides in LC-MS/MS analysis. By analyzing the overlap of identified glycoproteins, phosphoproteins, and M6P glycoproteins, we observed that 2 proteins bear all three PTMs, 3 proteins have both M6P glycosylation and phosphorylation, 19 proteins have both common N-glycosylation and M6P glycosylation, and 76 proteins possess both glycosylation and phosphorylation sites, which shed light on the crosstalk analysis of these multiply modified peptides (**Figure 4E**).

Characterization of common N-glycoproteome and M6P glycoproteome datasets. Of the 3961 N-glycopeptides (including common N-glycopeptides and M6P glycopeptides), we mainly focus

on the analysis of unique glycoforms and remove the redundant results. This analysis resulted in the identification of 3279 unique glycoforms, corresponding to 724 N-glycosites and 388 N-glycoproteins. We compared our current results with a published mouse lung glycoproteome.[32] The comparison revealed that the two datasets profiled similar number of glycoproteins (388 *vs.* 394) and glycosites (724 *vs.* 686), and among the 724 glycosites, 335 were commonly identified in both datasets while 389 were newly identified in our dataset. Our dataset also contained more identifications of glycoforms, whereas 2433 glycoforms were uniquely identified in our dataset (**Figure S3**).

We then analyzed the heterogeneity of the profiled glycoproteome. In brief, 65.2% (253) of glycoproteins possess only one glycosite and each glycoprotein had an average of 1.9 N-linked glycosites (**Figure 6A, left**). However, the number of glycans per glycosite showed different trend, 59.1% (428) of profiled glycosites were modified with more than one glycan and 21.8% of them have more than six glycans, and each glycosite had an average of 4.5 different glycans (**Figure 6A, right**). These data demonstrated the significant heterogeneity of N-linked glycosylation.

In total, we identified 210 unique glycans and categorized the identified glycans into six groups based on the state of their biosynthetic process. The six categories represented a gradual maturation of glycans from mature high mannose, to paucimannose, mannose-6-phosphate, complex/hybrid, and finally fucosylated and sialylated. We found that 20.3% and 11.2% of the identified N-glycoforms possess high mannose and paucimannose glycans, respectively. And 1.6% of them corresponded to M6P glycosylation. 14.6% of the N-glycoforms possessed complex/hybrid glycans. Given our glycan categorization rules, any glycan with a NeuAc and/or

NeuGc moiety was categorized as sialylated, therefore some glycans in the sialylated group were also fucosylated. However, the fucosylated glycan type group contained any glycans containing a fucose moiety that was not sialylated. Fucosylated and sialylated glycans corresponded to 14.0% and 38.2% of the identified N-glycoforms (**Figure 6B**). In total, 66.8% N-glycoforms were complex/hybrid glycans or decorated with fucose or sialic acid, which was in accordance with the results obtained with conventional HILIC method revealing 67.1% of mouse lung glycoproteome contained complex/hybrid glycans. These data demonstrated that, the majority of the glycans in mouse lung have been processed to their mature state.

We identified 54 unique M6P glycoforms corresponding to 23 N-glycoproteins. It is interesting that 18 M6P glycoforms have their common N-glycoform counterparts. We selected some of them and showed in the left panel of **Figure 7**, N-acetylglucosamine-6-sulfatase (Gns) possesses the M6P glycans Man6GlcNAc2Phospho1, Man7GlcNAc2Phospho1, Man7GlcNAc2Phospho2, and Man9GlcNAc3Phospho1 at Asn354, and it also possesses Man6GlcNAc2, Man7GlcNAc2, and Man9GlcNAc2 at the same site. The situation for Tripeptidyl-peptidase 1 (Tpp1) and Cathepsin L1 (Cts1) was similar. These findings are consistent with literature reports that phosphate on M6P glycans can be removed by acid phosphatase after M6P glycoproteins are distributed to their appropriate cellular localization.[33] With this in mind, to comprehensively profile the glycosylation of M6P glycoproteins, especially to quantify their changes during specific biological process, the non-M6P counterparts must be considered. Therefore, the simultaneous enrichment strategy is essential for this situation. As shown in the right panel of **Figure 7,** conventional HILIC method revealed fewer M6P glycoforms and their

counterparts than the HILIC mode Ti-IMAC enrichment method, therefore our simultaneous enrichment and separation strategy can achieve more comprehensive profiling and quantification of M6P glycosylation in complex biological sample.

Characterization of phosphoproteome dataset and its crosstalk with glycoproteome. Another great advantage of our method is that we can also perform comprehensive profiling of the phosphoproteome in the same sample. In total, we identified 5539 phosphopeptides corresponding to 2004 phosphoproteins. To examine the potential cross-regulation between glycosylation and phosphorylation, we analyzed the proteins containing these two PTMs. A total of 76 proteins possesses both glycosylation and phosphorylation sites (**Figure 4B**). A representative example is angiotensin-converting enzyme (ACE), a zinc-dependent peptidase responsible for converting angiotensin I into the vasoconstrictor angiotensin II, which plays a critical role in the blood pressure regulation.[34] Moreover, as a relatively nonspecific peptidase that can cleave a wide range of substrates, it also affects many other physiological processes including hematopoiesis, reproduction, renal development, renal function, and immune response.[35] It has been demonstrated that N-linked glycosylation is crucial for maintaining the stability and enzymatic activity of ACE, since the expressed ACE without glycosylation is catalytically inactive and subjects to rapid degradation.[36] Phosphorylation of ACE cytoplasmic tail at site Ser1305 regulates its retention in the endothelial cell plasma membrane and affect it secretion to extracellular space to perform its function.[37] As shown in **Figure 8**, with our HILIC mode Ti-IMAC enrichment method, we successfully profiled 129 unique glycoforms of ACE across its nine glycosites and identified a phosphopeptide with phosphorylation at site Ser1305. As a comparison, in a published mouse lung

glycoproteome, 133 unique glycoforms were mapped on the same nine glycosites in ACE, which was similar with our data (**Table S2**), however the phosphorylation site was not reported for ACE in that study.[32] Our simultaneous enrichment strategy will facilitate more in-depth investigation of the cross-regulation of glycosylation and phosphorylation on ACE function in future studies. This new strategy will also shed light on the large-scale crosstalk analysis of protein glycosylation and phosphorylation based on the simultaneously enriched glycoproteome and phosphoproteome data collected from the same biological sample.

Conclusion

We explore and demonstrate the potential of Ti-IMAC for the simultaneous enrichment of common N-glycopeptides, phosphopeptides, and M6P glycopeptides from the same complex biological sample. Moreover, as majority of glycopeptides were captured on materials with only hydrophilic interaction, they were eluted first with aqueous buffer containing formic acid (FA) or trifluoracetic acid (TFA). Phosphopeptides were mainly captured on the materials with electrostatic interaction. Therefore, they can be eluted with high acetonitrile (ACN) content buffer containing ammonia. Finally, M6P glycopeptides captured on materials with synergistic electrostatic and hydrophilic interaction can be eluted with low ACN content buffer containing ammonia. The three classes of PTM peptides were successfully separated with appropriate fractionation, therefore, the interference among these three classes of peptides can be substantially eliminated and the coverage of N-glycopeptides, phosphopeptides and M6P glycopeptides were significantly increased to 1.9, 2.3, and 4.3-fold respectively compared with conventional

enrichment method. The enrichment and fractionation of three different PTMs from the same biological sample can also facilitate the crosstalk analysis of these three PTMs.

Acknowledgements

This work was supported, in part, by the National Institutes of Health Grants U01CA231081, RF1AG052324, and R01 DK071801 (to L.L.), and funds from the National Key R&D Program of China (2016YFA0501402, to M.Y.), the National Natural Science Foundation of China (21535008 and 21525524, to M.Y.). The MS instruments were purchased through the support of NIH Shared Instrument Grants (NIH-NCRR S10RR029531 and S10OD025084 to L.L.) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. The authors wish to pay tribute to the late Prof. Hanfa Zou in memory of his great contribution that led the development of Ti(IV)-IMAC materials.

References

(1) Beltrao, P.; Albanèse, V.; Kenner, Lillian R.; Swaney, Danielle L.; Burlingame, A.; Villén, J.; Lim, Wendell A.; Fraser, James S.; Frydman, J.; Krogan, Nevan J. Systematic Functional Prioritization of Protein Posttranslational Modifications. *Cell* **2012**, *150*, 413-425.

(2) Swaney, D. L.; Beltrao, P.; Starita, L.; Guo, A.; Rush, J.; Fields, S.; Krogan, N. J.; Villen, J. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods* **2013**, *10*, 676-682.

(3) Huang, J.; Wang, F.; Ye, M.; Zou, H. Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications. *J. Chromatogr. A* **2014**, *1372*, 1-17.

(4) Doll, S.; Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **2015**, *10*, 63-71.

(5) Chen, Z.; Huang, J.; Li, L. Recent advances in mass spectrometry (MS)-based glycoproteomics in complex biological samples. *Trends Analyt. Chem.* **2019**, *118*, 880-892.

(6) Huang, J.; Qin, H.; Dong, J.; Song, C.; Bian, Y.; Dong, M.; Cheng, K.; Wang, F.; Sun, D.; Wang, L.; Ye, M.; Zou, H. In Situ Sample Processing Approach (iSPA) for Comprehensive Quantitative Phosphoproteome Analysis. *J. Proteome Res.* **2014**, *13*, 3896-3904.

(7) Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.; Gillette, M. A.; Jaffe, J. D.; Carr, S. A. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods* **2013**, *10*, 634-637.

(8) Andaluz Aguilar, H.; Iliuk, A. B.; Chen, I. H.; Tao, W. A. Sequential phosphoproteomics and N-glycoproteomics of plasma-derived extracellular vesicles. *Nat. Protoc.* **2020**, *15*, 161-180.

(9) Hong, Y.; Zhao, H.; Pu, C.; Zhan, Q.; Sheng, Q.; Lan, M. Hydrophilic Phytic Acid-Coated Magnetic Graphene for Titanium(IV) Immobilization as a Novel Hydrophilic Interaction Liquid Chromatography–Immobilized Metal Affinity Chromatography Platform for Glyco- and Phosphopeptide Enrichment with Controllable Selectivity. *Anal. Chem.* **2018**, *90*, 11008-11015.

(10) Tang, R.; Yu, Y.; Dong, J.; Yao, Y.; Ma, S.; Ou, J.; Ye, M. Facile preparation of bifunctional adsorbents for efficiently enriching N-glycopeptides and phosphopeptides. *Anal. Chim. Acta* **2021**, *1144*, 111-120.

(11) Zou, X.; Jie, J.; Yang, B. Single-Step Enrichment of N-Glycopeptides and Phosphopeptides with Novel Multifunctional Ti4+-Immobilized Dendritic Polyglycerol Coated Chitosan Nanomaterials. *Anal. Chem.* **2017**, *89*, 7520-7526.

(12) Wang, Z.; Wang, J.; Sun, N.; Deng, C. A promising nanoprobe based on hydrophilic interaction liquid chromatography and immobilized metal affinity chromatography for capture of glycopeptides and phosphopeptides. *Anal. Chim. Acta* **2019**, *1067*, 1-10.

(13) Huang, J.; Dong, J.; Shi, X.; Chen, Z.; Cui, Y.; Liu, X.; Ye, M.; Li, L. Dual-Functional Titanium(IV) Immobilized Metal Affinity Chromatography Approach for Enabling Large-Scale Profiling of Protein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. *Anal. Chem.* **2019**, *91*, 11589-11597.

(14) Zhou, H.; Ye, M.; Dong, J.; Han, G.; Jiang, X.; Wu, R.; Zou, H. Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J. Proteome Res.* **2008**, *7*, 3957-3967.

(15) Yu, Z.; Han, G.; Sun, S.; Jiang, X.; Chen, R.; Wang, F.; Wu, R.; Ye, M.; Zou, H. Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. *Anal. Chim. Acta* **2009**, *636*, 34-41.

(16) Glover, M. S.; Yu, Q.; Chen, Z. W.; Shi, X. D.; Kent, K. C.; Li, L. J. Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass spectrom.* **2018**, *427*, 35-42.

(17) Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J. R.; Zou, H.; Mohammed, S. Robust phosphoproteome enrichment using monodisperse microsphere–based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **2013**, *8*, 461-480.

(18) Li, F.; Ding, J. Sialylation is involved in cell fate decision during development, reprogramming and cancer progression. *Protein & Cell* **2019**, *10*, 550-565.

(19) Nwosu, C. C.; Strum, J. S.; An, H. J.; Lebrilla, C. B. Enhanced Detection and Identification of Glycopeptides in Negative Ion Mode Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 9654-9662.

(20) de Haan, N.; Yang, S.; Cipollo, J.; Wuhrer, M. Glycomics studies using sialic acid derivatization and mass spectrometry. *Nature Reviews Chemistry* **2020**, *4*, 229-242.

(21) Feng, Y.; Li, M.; Lin, Y.; Chen, B.; Li, L. Multiplex Quantitative Glycomics Enabled by Periodate Oxidation and Triplex Mass Defect Isobaric Multiplex Reagents for Carbonyl-Containing Compound Tags. *Anal. Chem.* **2019**, *91*, 11932-11937.

(22) Nishikaze, T. Sialic acid derivatization for glycan analysis by mass spectrometry. *Proceedings of the Japan Academy, Series B* **2019**, *95*, 523-537.

(23) Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **1990**, *499*, 177-196.

(24) McNulty, D. E.; Annan, R. S. Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. *Molecular & Cellular Proteomics* **2008**, *7*, 971-980.

(25) Park, K.-S.; Mohapatra, D. P.; Misonou, H.; Trimmer, J. S. Graded Regulation of the Kv2.1 Potassium Channel by Variable Phosphorylation. *Science* **2006**, *313*, 976-979.

(26) Lee, C. W.; Ferreon, J. C.; Ferreon, A. C. M.; Arai, M.; Wright, P. E. Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19290-19295.

(27) Samarasimhareddy, M.; Mayer, G.; Hurevich, M.; Friedler, A. Multiphosphorylated peptides: importance, synthetic strategies, and applications for studying biological mechanisms. *Org. Biomol. Chem.* **2020**, *18*, 3405-3422.

(28) Zhong, H.; Xiao, X.; Zheng, S.; Zhang, W.; Ding, M.; Jiang, H.; Huang, L.; Kang, J. Mass spectrometric analysis of mono- and multi-phosphopeptides by selective binding with NiZnFe2O4 magnetic nanoparticles. *Nature Communications* **2013**, *4*, 1656.

(29) Tsai, C.-F.; Hsu, C.-C.; Hung, J.-N.; Wang, Y.-T.; Choong, W.-K.; Zeng, M.-Y.; Lin, P.-Y.; Hong, R.-W.; Sung, T.-Y.; Chen, Y.-J. Sequential Phosphoproteomic Enrichment through Complementary Metal-Directed Immobilized Metal Ion Affinity Chromatography. *Anal. Chem.* **2014**, *86*, 685-693.

(30) Choi, S.; Kim, J.; Cho, K.; Park, G.; Yoon, J. H.; Park, S.; Yoo, J. S.; Ryu, S. H.; Kim, Y. H.; Kim, J. Sequential Fe3O4/TiO2 enrichment for phosphopeptide analysis by liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 1467-1474.

(31) Long, X.-y.; Zhang, Z.-j.; Li, J.-y.; Sheng, D.; Lian, H.-z. A combination strategy using two novel cerium-based nanocomposite affinity probes for the selective enrichment of mono- and multi-phosphopeptides in mass spectrometric analysis. *Chem. Commun.* **2017**, *53*, 4620-4623.

(32) 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*, 438.

(33) Makrypidi, G.; Damme, M.; Müller-Loennies, S.; Trusch, M.; Schmidt, B.; Schlüter, H.; Heeren, J.; Lübke, T.; Saftig, P.; Braulke, T. Mannose 6 Dephosphorylation of Lysosomal Proteins Mediated by Acid Phosphatases Acp2 and Acp5. *Mol. Cell. Biol.* **2012**, *32*, 774-782.

(34) Bernstein, K. E.; Khan, Z.; Giani, J. F.; Cao, D.-Y.; Bernstein, E. A.; Shen, X. Z. Angiotensin-converting enzyme in innate and adaptive immunity. *Nature Reviews Nephrology* **2018**, *14*, 325-336.

(35) Bernstein, K. E.; Ong, F. S.; Blackwell, W.-L. B.; Shah, K. H.; Giani, J. F.; Gonzalez-Villalobos, R. A.; Shen, X. Z.; Fuchs, S. A Modern Understanding of the Traditional and Nontraditional Biological Functions of Angiotensin-Converting Enzyme. *Pharmacol. Rev.* **2013**, *65*, 1-46.

(36) Sadhukhan, R.; Sen, I. Different Glycosylation Requirements for the Synthesis of Enzymatically Active Angiotensin-converting Enzyme in Mammalian Cells and Yeast. *J. Biol. Chem.* **1996**, *271*, 6429-6434.

(37) Kohlstedt, K.; Shoghi, F.; Müller-Esterl, W.; Busse, R.; Fleming, I. CK2 Phosphorylates the Angiotensin-Converting Enzyme and Regulates Its Retention in the Endothelial Cell Plasma Membrane. *Circul. Res.* **2002**, *91*, 749-756.

Figures

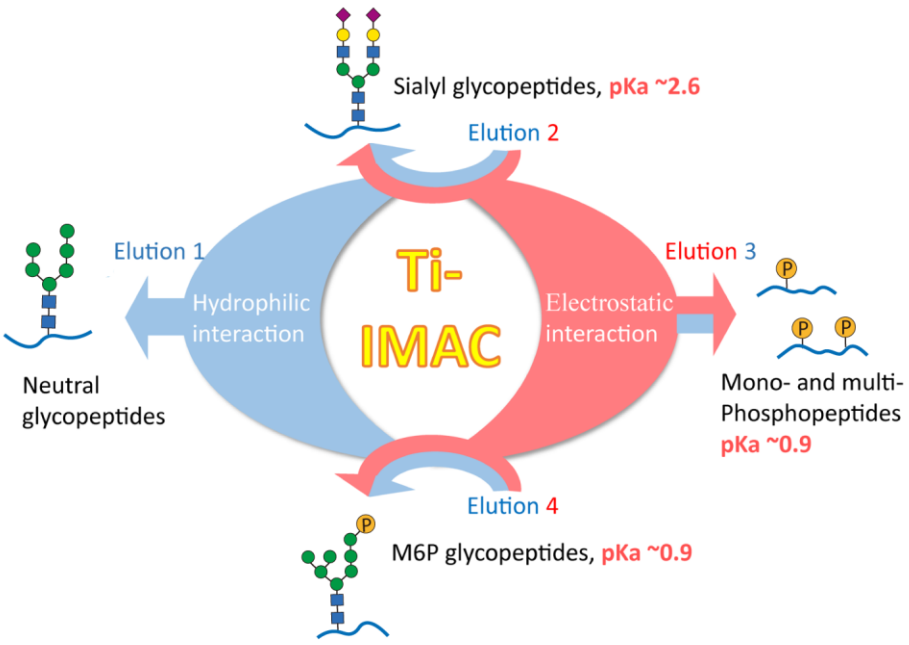


Figure 1. Scheme of IMAC and HILIC dual-mode affinity enrichment approach for the simultaneous enrichment of glycopeptides, phosphopeptides and M6P glycopeptides. Elution is performed in four steps to separate different peptides according to their chemical properties: Elution 1 uses weak acidic buffer with decreased ACN gradient and mainly elutes neutral glycopeptides; Elution 2 uses strong acidic condition that allows protonating sialylated glycopeptides to be collected and makes them isolated from the IMAC beads that bind to the unprotonated phosphopeptides and M6P glycopeptides; Elution 3 uses basic buffer with high and medium ACN content, which can elute mono- and multi- phosphopeptides, respectively; Elution 4 uses basic buffer with low ACN content and M6P glycopeptides are mainly eluted in this fraction.

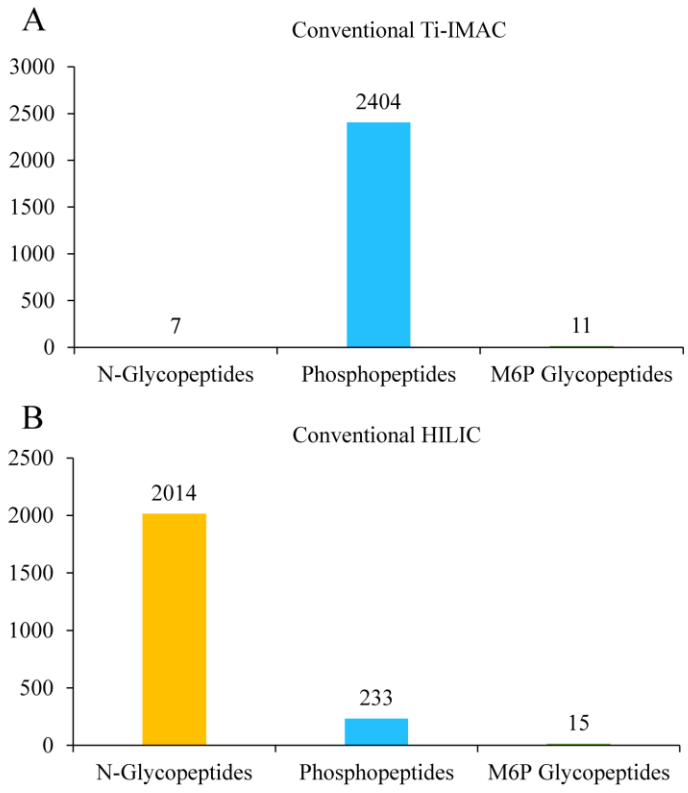


Figure 2. Common N-glycopeptide, phosphopeptide and M6P glycopeptide enrichment with conventional method. (A) conventional Ti-IMAC method. (B) conventional HILIC approach.

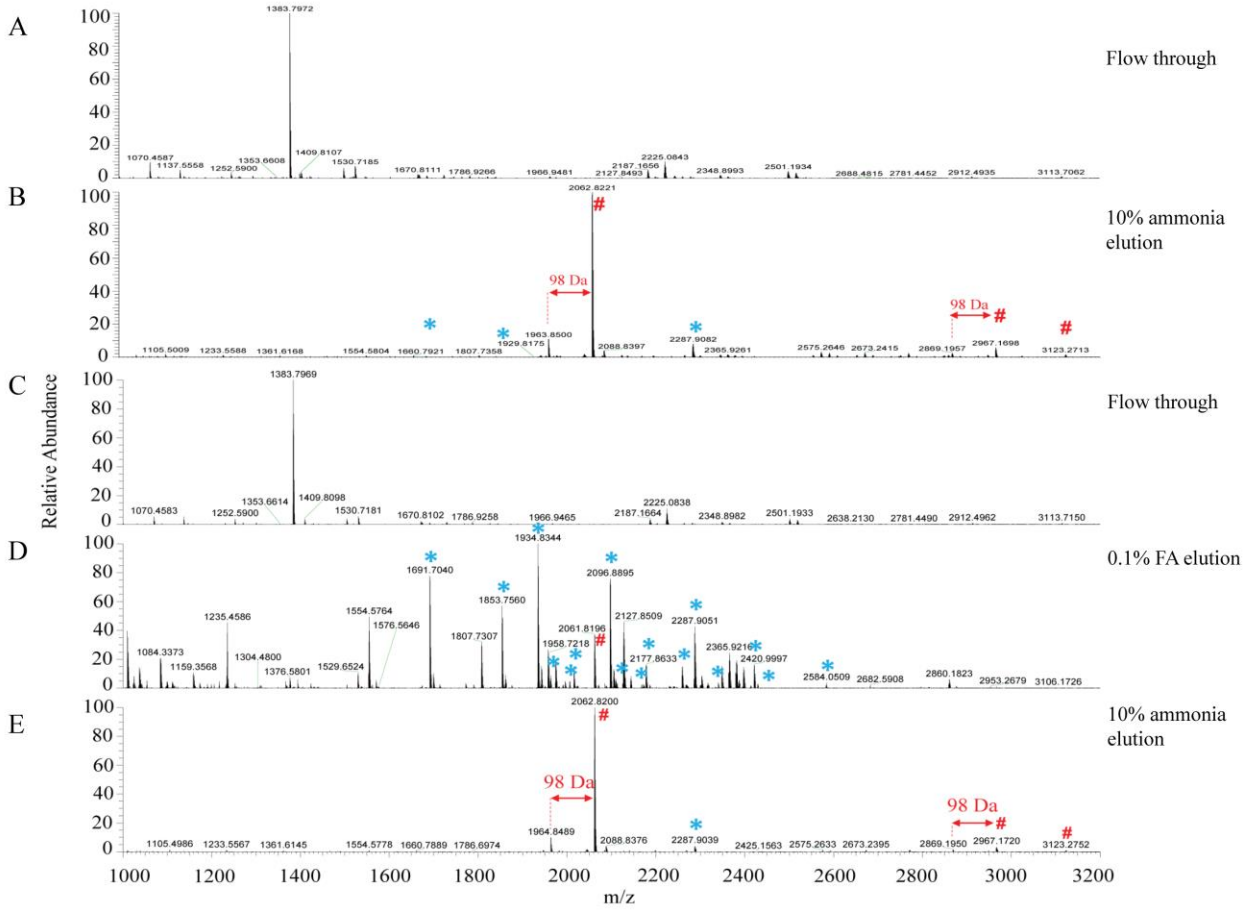


Figure 3. MALDI-TOF MS analysis of Ti-IMAC enriched β-casein and RNase B digest mixture. (A, B) Ti-IMAC enrichment in HILIC-mode and elution without fractionation. (C-E) Ti-IMAC enrichment in HILIC-mode and elution with 0.1% FA (D) and 10% ammonia (E) fractionation. (# represents phosphopeptides and * represents glycopeptides)

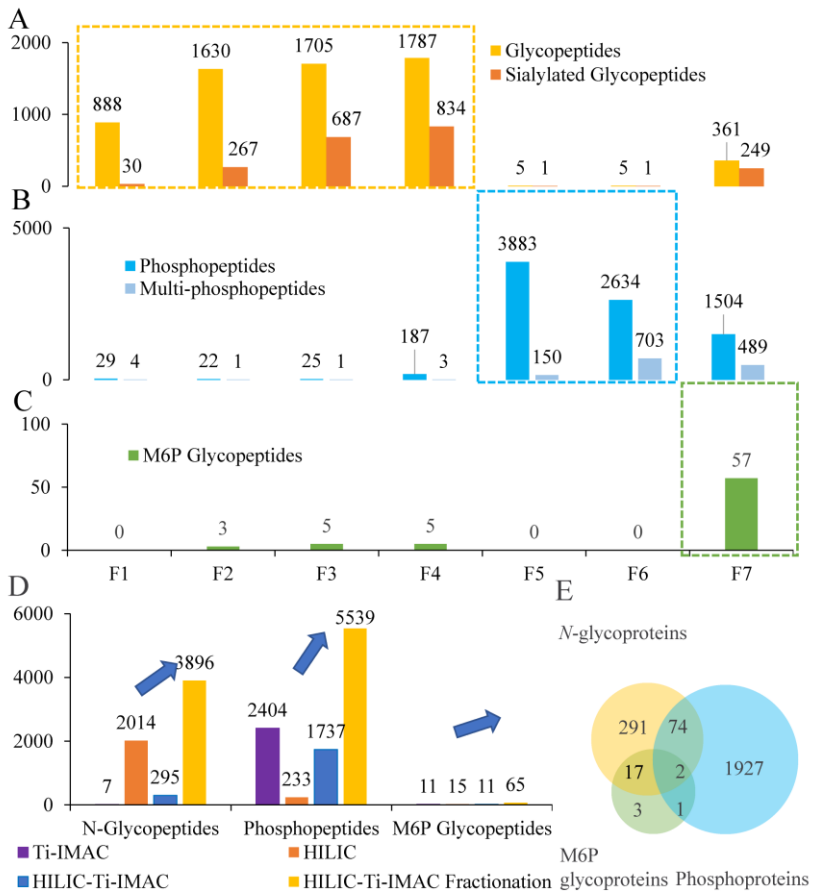


Figure 4. Dual-mode affinity enrichment approach enables the enrichment and separation of common N-glycopeptides, phosphopeptides, and M6P glycopeptides. (A) Common N-glycopeptide identification results. (B) Phosphopeptide identification results. (C) M6P glycopeptide identification results. (D) Comparison of identified N-glycopeptides, phosphopeptides and M6P glycopeptides across different enrichment methods. (E) Overlap analysis of N-glycoproteins, phosphoproteins and M6P glycoproteins profiled in the HILIC mode Ti-IMAC with fractionation method. (F1-7 represent fractions 1-7)

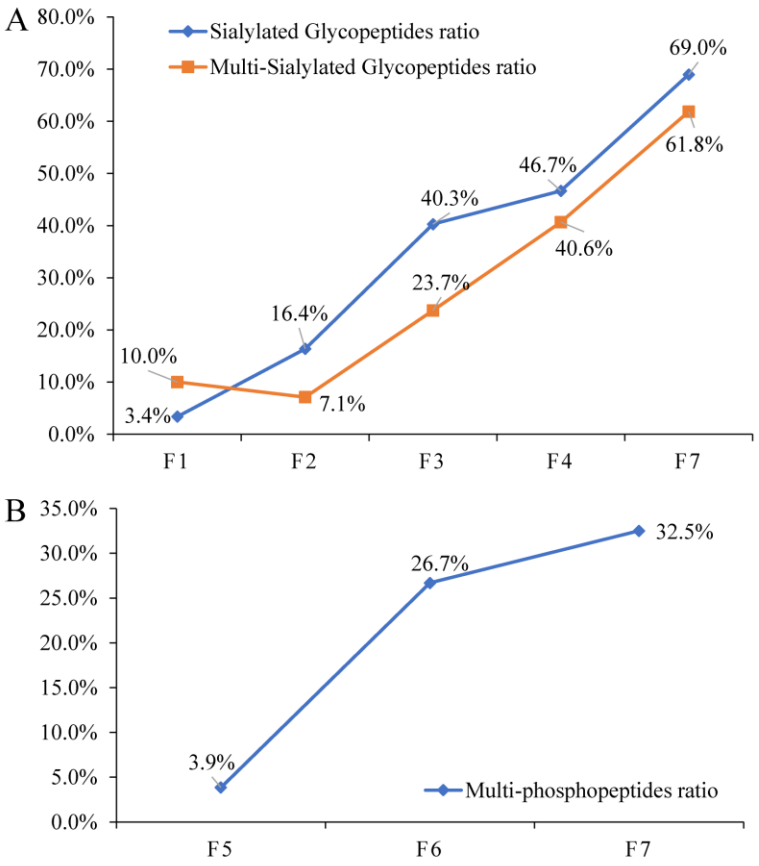


Figure 5. Fractionation efficiency of different elution conditions. (A) Sialylated and multi-sialylated glycopeptides ratio. (B) Multi-phosphopeptides ratio. (F1-7 represent fractions 1-7)

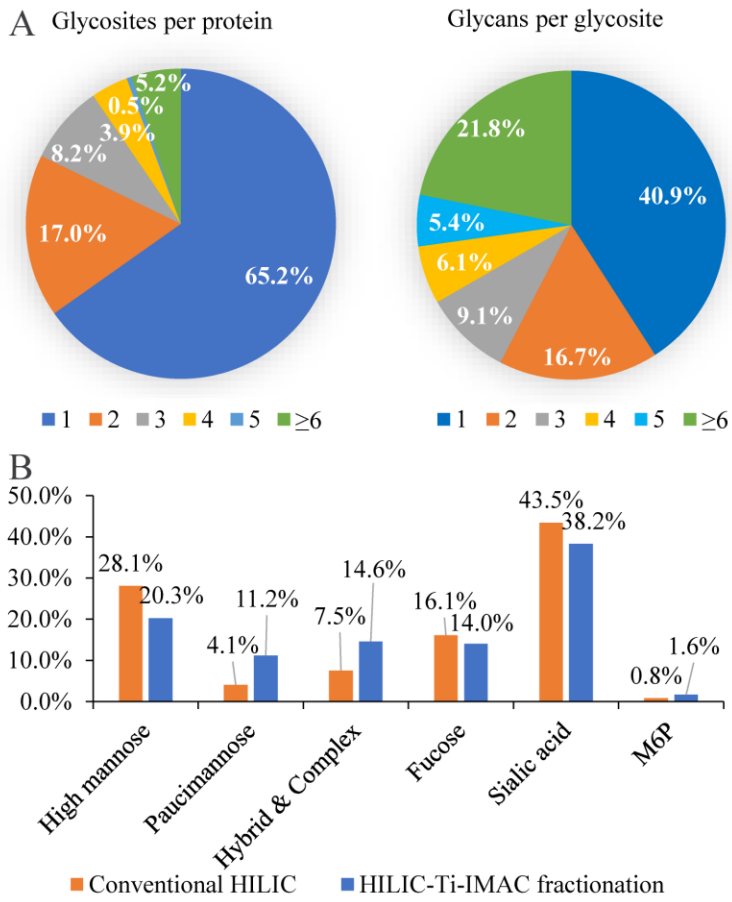


Figure 6. Glycoproteome heterogeneity analysis. (A**)** Distribution of the number of glycosites per glycoprotein (left) and the number of unique glycans per glycosites (right). (B**)** Glycan categorization of identified unique glycoforms.

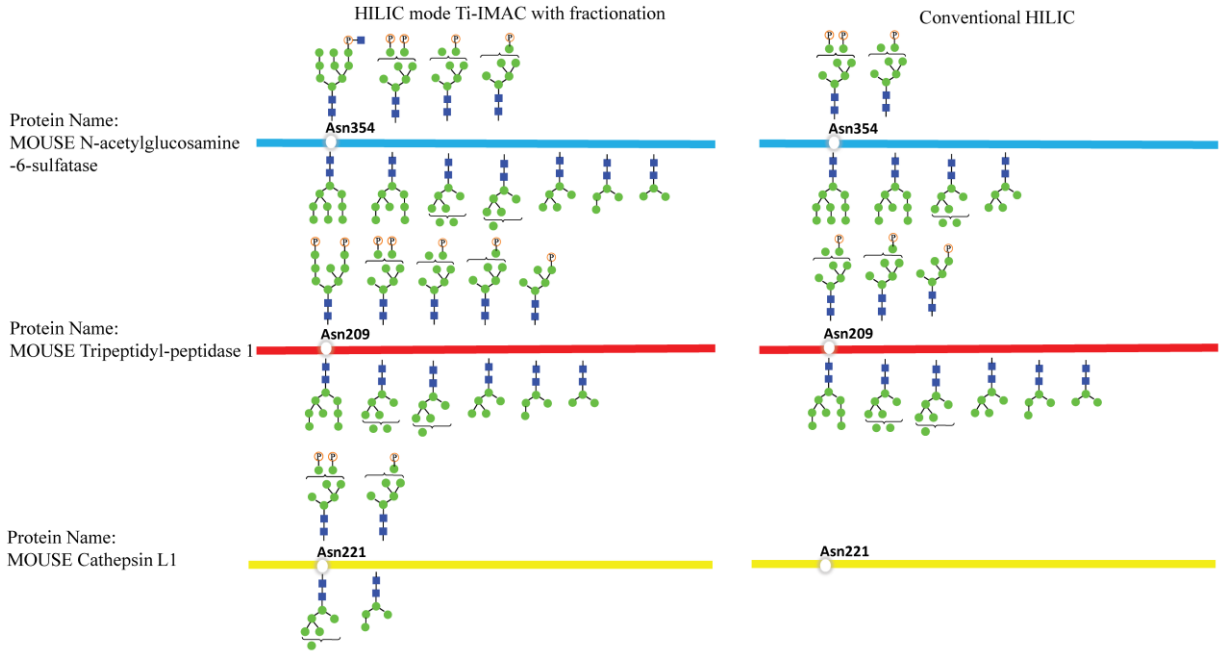


Figure 7. Mouse lung M6P glycosylation and its phosphate free counterpart revealed by the HILIC mode with fractionation and conventional HILIC mode enrichment.

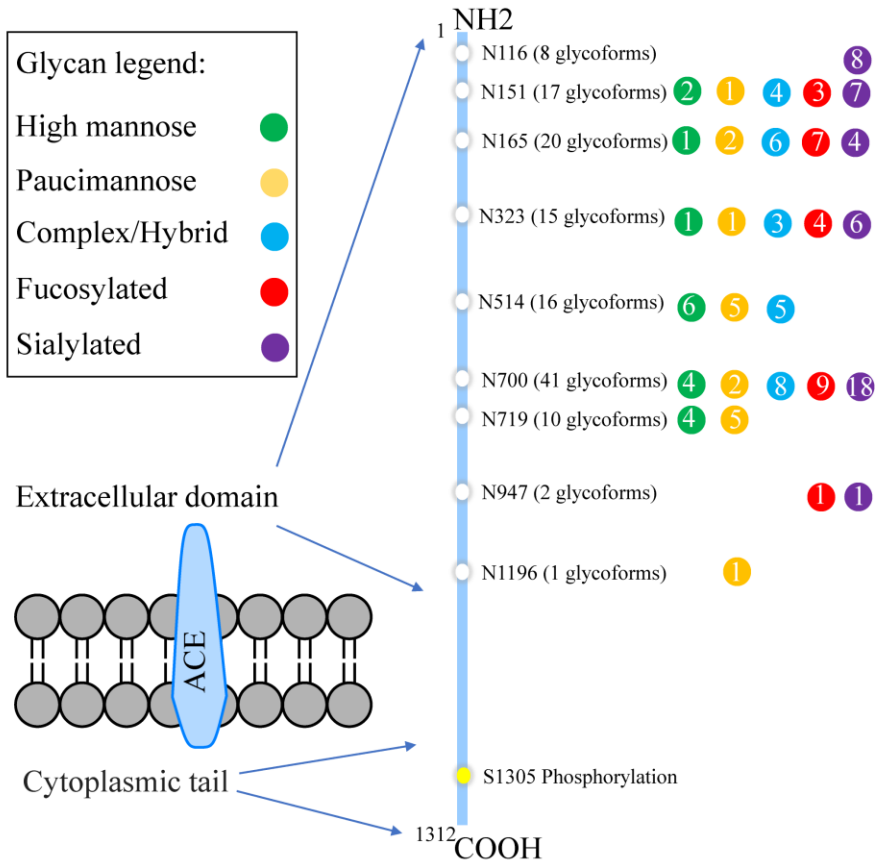


Figure 8. Glycan diversity and phosphorylation on angiotensin-converting enzyme (ACE).

Supplemental Information

Experimental Section.

Sample preparation. Weigh 1 mg of standard phosphoprotein β-casein and standard glycoprotein Ribonuclease B (RNase B) and dissolve them in 125 ml 8M urea and 50 mM TEAM buffer (pH 8.0), respectively. The mouse lung tissue was collected from male adult C57BL/6 mice. The tissue sample was sliced into 1-2 cubic millimeter pieces in clean dish with ice bath and washed with 150 mM ice cold PBS buffer for 3 times. Then they were homogenized and lysed in a homogenization buffer consisting of 4% v/v SDS, 65 mM DTT, 150 mM NaCl, and 25 mM Tris (pH 7.4) (1 protease inhibitor tablet (Roche, Mannheim, Germany) and 1 phosphatase inhibitor tablet (Roche, Mannheim, Germany) were added in every 10 mL of the lysis buffer). Lysed buffer was continually sonicated in a probe sonicator with 60 W energy and pulse 5s on 5s off for 60 cycles. The homogenates were centrifuged with 3000 x g for 15 min and the supernatants were collected and poured into 5-fold ice cold precipitation buffer (acetone: ethanol: acetic acid=50: 50: 0.1). The precipitation was put in -20 ℃ for 12 h and then was centrifuged with 3000 x g at 4 ℃ for 15 min. The protein pellet was washed with 10 ml ice cold precipitation buffer twice and centrifuged with 3000 x g for 15 min to remove the supernatants. The pellet was put in hood for 15 min to dry the remaining precipitation buffer and then was re-dissolved in 8M urea and 50 mM TEAB buffer (pH 8.0). Protein concentration was measured by a BCA assay kit (Thermo Fisher Scientific, San Jose, CA) and stored in -80 ^{o}C.

Protein digestion. 1 mg of standard phosphoprotein, glycoprotein, and mouse lung protein samples were weighed respectively. The samples were treated with 10 mM dithiothreitol (DTT) for 1 h to reduce the disulfide bonds. After reduction, the samples were alkylated with 20 mM

iodoacetamide for 30 min in the dark to block the reduced cysteine. The samples were quenched by adding 10 mM DTT for 5 min. In the end, the sample buffer was diluted to 1 M urea with 50 mM TEAB and trypsin (Promega, Madison, WI) was added to the samples with protein: trypsin ratio at 50:1 and incubated at 37 ^{o}C water bath for 16 h. Digested samples were stored at -80 ^{o}C for later use.

Typical Ti(IV)-IMAC enrichment and elution. Ti(IV)-IMAC adsorbents were home-made and the enrichment protocol was adapted from our previous protocol.[1-3] Briefly, the 200 μg tryptic mouse lung peptide mixture was added in the loading buffer (80% acetonitrile (ACN) (v/v) and 6% trifluoroacetic acid (TFA) (v/v)) with 1:1 ratio (v/v) and then they were incubated with Ti(IV)-IMAC adsorbents with a ratio of 1:10 (w/w) for 30 min. Then the non-specific adsorbed peptides on the surface of adsorbents were washed sequentially by washing buffer 1 (50% ACN (v/v) 6% TFA (v/v) and 200 mM NaCl) and washing buffer 2 (30% ACN (v/v) and 0.1% TFA (v/v)). Finally, the enriched phosphopeptides and M6P glycopeptides were eluted from the adsorbents with 10% NH_4OH (v/v). After centrifugation at 20,000 x g for 5 min, the supernatant was collected and immediately dried down in SpeedVac. Samples were stored at -20 ^{o}C and re-dissolved in 1% formic acid (FA) before LC–MS/MS analysis and load 15% of each sample for analysis.

Typical HILIC enrichment and elution. HILIC enrichment was performed with homemade HILIC stage-tip which uses a neutral, polar material (PolyHYDROXYETHYL A: 12μm and 300 Å from PolyLC Inc.). Briefly, 2 mg of cotton wool was weighed and inserted into a 200 μL pipette tip. Place the pipette tip on a 2 mL sample vial with the help of an adapter unit. HILIC beads were dissolved into 1% trifluoroacetic acid (TFA) solvent and the HILIC beads slurry was vortexed for

15 min to activate the HILIC beads. Transfer the HILIC slurry onto the top of the cotton wool. Centrifuge the sample vial to remove the solvent at 500 xg for 2 min. Add 200 µL of 1% TFA to flush the beads. Centrifuge the sample vial at 500 xg for 2 min to remove the liquid. Condition the HILIC beads with 200 µL 80% ACN/1% TFA for three times. 200 µg of tryptic mouse lung sample was dissolved in 80% ACN/1% TFA and loaded onto the HILIC stage-tip. Centrifuge the sample vial at 500 xg for 2 min and flow through of the sample was re-loaded to the HILIC stage-tip twice. Then wash away the non-glycopeptides with 200 µL of 80% ACN/ 1% TFA for three times, and then elute the glycopeptides with 300 µL of 0.1% FA water. The sample was immediately dried down in SpeedVac. Samples were stored at -20 ^{o}C and re-dissolved in 0.1% formic acid (FA) before LC–MS/MS analysis and load 15% of each sample for analysis.

Fractionation method: Two aliquots of 250 µg of tryptic digested mouse lung sample were dissolved in 80% ACN/1% TFA and loaded onto the HILIC stage-tip. Centrifuge the sample vial at 500 xg for 2 min and the flow-through sample was re-loaded to the HILIC stage-tip twice. The non-glycopeptides were washed with 200 µL of 80% ACN/ 1% TFA for three times, followed by elution of the glycopeptides with 60% ACN (v/v) and 0.1% FA (v/v), 40% ACN (v/v) and 0.1% FA (v/v), 20% ACN (v/v) and 0.1% FA (v/v), 0.1% FA (v/v), respectively. The elutions from the two aliquots of samples were combined and were immediately dried down in SpeedVac. Samples were stored at -20 ^{o}C and re-dissolved in 0.1% formic acid (FA) before LC–MS/MS analysis and 20% of each sample was loaded for analysis.

MALDI-TOF and NanoLC-MS/MS analysis. Standard samples were tested on a Bruker Rapiflex MALDI-TOF/TOF instrument (Bruker Daltonik, Bremen, Germany) with 2,5-

Dihydroxybenzoic acid matrix (Sigma Aldrich, St. Louis, USA). Ion source parameters: laser energy 70%, frequency 200 Hz, laser 355 nm and each data collection of 10,000 shots. Complex samples were analyzed with an Ultimate 3000 nanoLC coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Re-dissolved sample was injected onto a 75 μm i.d. × 15 cm length homemade column with an integrated HF etched emitter tip and packed with 1.7 μm, 130 Å, BEH C_{18} material from a Waters UPLC column (Waters, Milford, MA). Peptides were separated with a gradient that ramped from 97% solvent A (0.1% FA in H_2O) and 3% solvent B (0.1% FA in ACN) to 30% solvent B over 80 min and then kept each of 75% solvent B, 95% solvent B, and 100% solvent A for 10 min. The flow rate was set at 300 nL/min. MS method was set according to the following parameters: MS scan range (*m/z*) = 400-2000; resolution = 120,000; AGC target = $1.0e^6$; maximum injection time = 250 ms; included charge state = 2-6; dynamic exclusion duration = 30 s. MS/MS method is a top 20 data-dependent acquisition (DDA) mode in which all MS/MS dissociations were performed with higher energy collisional dissociation (HCD): resolution = 60,000; AGC target = $2.0e^5$; maximum injection time = 150 ms; collision energy = 30%.

References

(1) Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J. R.; Zou, H.; Mohammed, S. Robust phosphoproteome enrichment using monodisperse microsphere–based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **2013**, *8*, 461-480.

(2) Huang, J.; Qin, H.; Dong, J.; Song, C.; Bian, Y.; Dong, M.; Cheng, K.; Wang, F.; Sun, D.; Wang, L.; Ye, M.; Zou, H. In Situ Sample Processing Approach (iSPA) for Comprehensive Quantitative Phosphoproteome Analysis. *J. Proteome Res.* **2014**, *13*, 3896-3904.

(3) Liu, Z.; Wang, F.; Chen, J.; Zhou, Y.; Zou, H. Modulating the selectivity of affinity absorbents to multi-phosphopeptides by a competitive substitution strategy. *J. Chromatogr. A* **2016**, *1461*, 35-41.

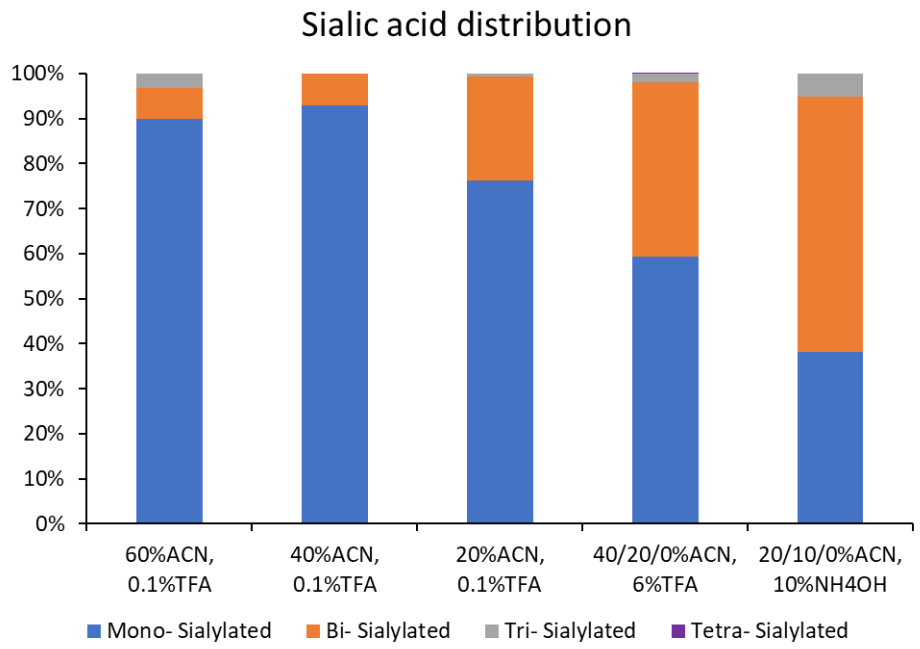


Figure S1. Distributions of sialic acid numbers of sialylated glycopeptides in each fraction.

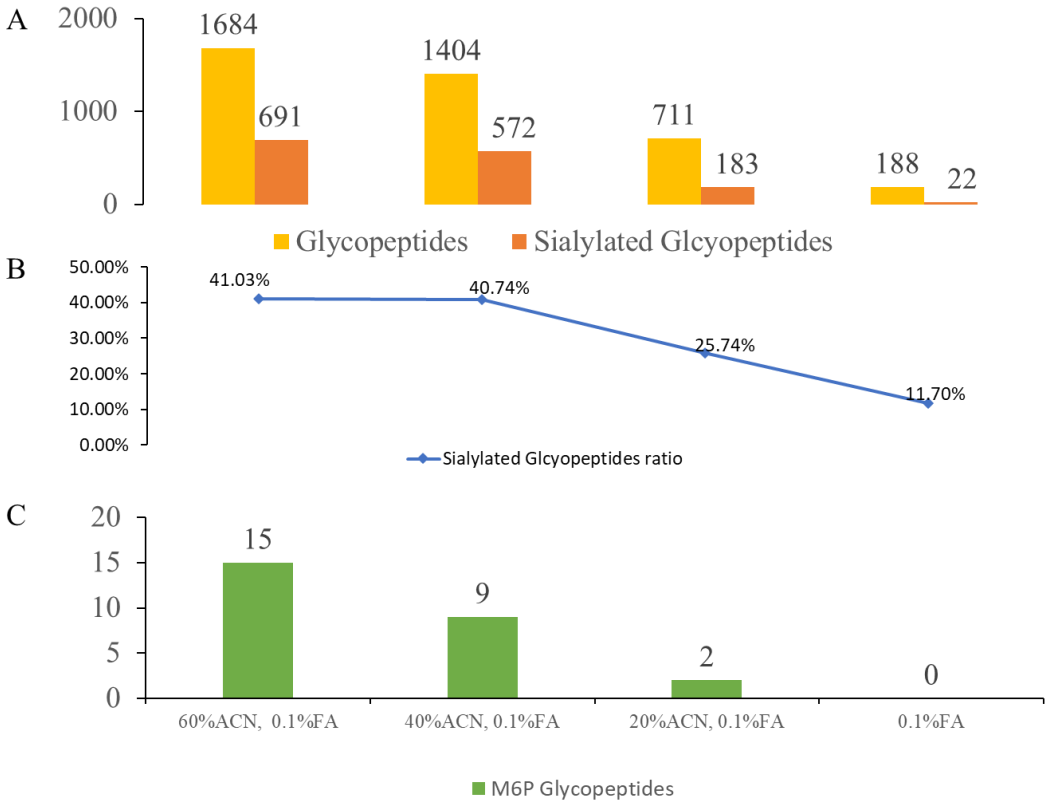


Figure S2. Glycopeptide fractionation with conventional HILIC approach. (A) Total N-glycopeptide and sialylated glycopeptide identification results. (B) Sialylated glycopeptide ratio. (C) M6P glycopeptide identification results.

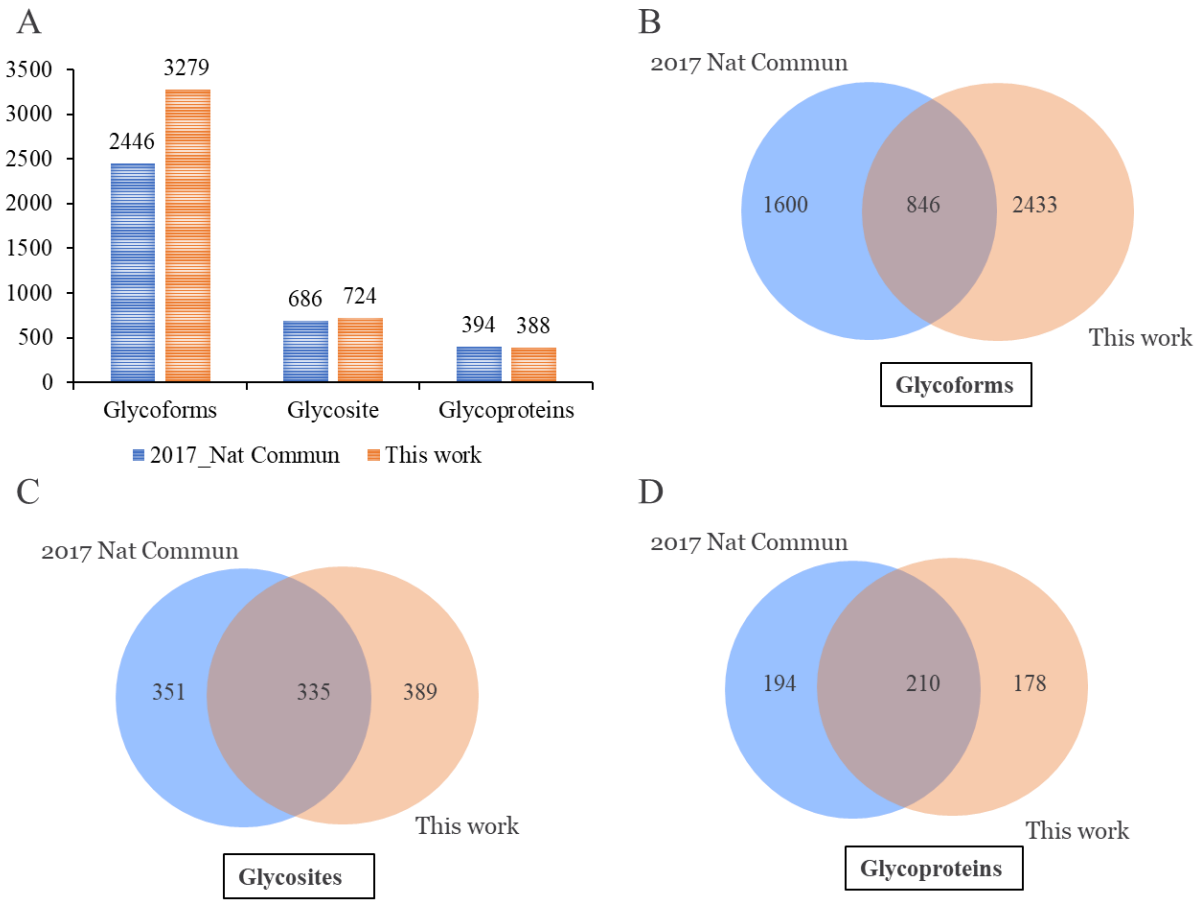


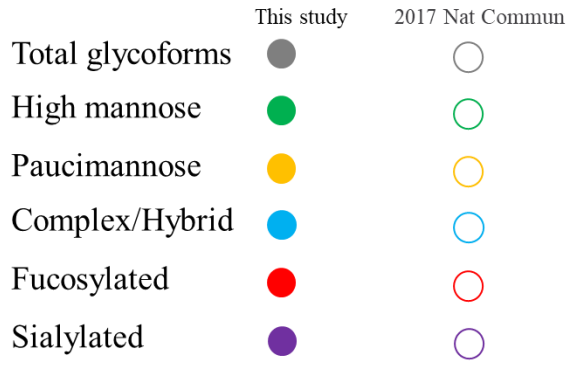
Figure S3. Comparison of mouse lung glycoproteomes profiled in this study and previously published data.

Table S1. MALDI-TOF MS analysis of HILIC-mode Ti-IMAC enriched peptides from β-casein and RNase B digest mixture with or without fractionation

Protein	Peptide	Glycosite	Glycan	$[M+H]^+$	No fractionation	Fractionation
Rnase B	NLTK	Asn-60	GlcNAc2 Man5	1691.71	1	1
Rnase B	NLTK	Asn-60	GlcNAc2 Man6	1853.76	1	1
Rnase B	SRNLTK	Asn-60	GlcNAc2 Man5	1934.84		1
Rnase B	NLTKDR	Asn-60	GlcNAc2 Man5	1962.84		1
Rnase B	NLTK(cam)DR	Asn-60	GlcNAc2 Man5	2005.84		1
Rnase B	NLTK	Asn-60	GlcNAc2 Man7	2015.82		1
Rnase B	SRNLTK	Asn-60	GlcNAc2 Man6	2096.9		1
Rnase B	NLTKDR	Asn-60	GlcNAc2 Man6	2124.89		1
Rnase B	NLTK(cam)DR	Asn-60	GlcNAc2 Man6	2167.9		1
Rnase B	NLTK	Asn-60	GlcNAc2 Man8	2177.87		1
Rnase B	SRNLTK	Asn-60	GlcNAc2 Man7	2258.95		1
Rnase B	NLTKDR	Asn-60	GlcNAc2 Man7	2286.94	1	1
Rnase B	NLTK	Asn-60	GlcNAc2 Man9	2339.92		1
Rnase B	SRNLTK	Asn-60	GlcNAc2 Man8	2421		1
Rnase B	NLTKDR	Asn-60	GlcNAc2 Man8	2449		1
Rnase B	SRNLTK	Asn-60	GlcNAc2 Man9	2583.05		1
β-casein	FQ[pS]EEQQQTEDELQDK			2061.83	1	1
β-casein	ELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR			2966.16	1	1
β-casein	RELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR			3122.27	1	1

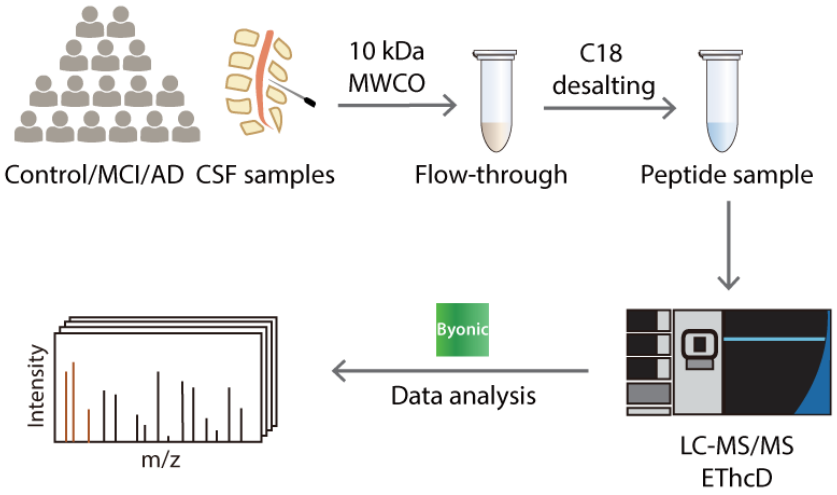
Table S2. MALDI-TOF MS analysis of glycopeptides and phosphopeptides enriched from mixture of RNase B and β-casein tryptic digests.

	N1156	N1515	N165	N323	N514	N700	N719	N947	N1196
Total glycoforms	8 16	17 2	20 25	15 5	16 26	41 23	10 6	2 24	1 6
High mannose		2	1 5	1 1	6 5	4 2	4 5		3
Pauci-mannose		1	2	1	5	2	5		1
Complex/ Hybrid	2	4 1	6 8	3	5 8	8 3	1	2	2
Fucosylated	1	3	7 9	4 2	9	9 6		1 3	
Sialylated	8 13	7 1	4 3	6 2	4	18 12		1 19	1



Appendix III

Profiling CSF Endogenous Peptidome in Alzheimer's Disease



Adapted from: **Wang, D.**; Chen, Z.; Li, L. Profiling CSF Endogenous Peptidome in Alzheimer's Disease. In *Pepidomics. Methods in Molecular Biology*, 2nd ed.; Humana Press, **2023**. (Submitted)

Abstract

Human cerebrospinal fluid (CSF) is a rich source for central nervous system (CNS)-related disease biomarker discovery due to its direct interchange with the extracellular fluid of the CNS. Though extensive proteome-level profiling has been conducted for CSF, studies targeting at its endogenous peptidome is still limited. It is harder to include the post-translational modifications (PTMs) characterization of the peptidome in the mass spectrometry (MS) analysis because of their low abundance and the challenge of data interpretation. In this chapter, we present a peptidomic workflow that combines molecular weight cut-off (MWCO) separation, EThcD fragmentation, and a three-step database searching strategy for comprehensive PTM analysis of endogenous peptides including both N-glycosylation and O-glycosylation and other common peptide PTMs. The method has been successfully adopted to analyze CSF samples from healthy donors, mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients to provide a landscape of peptidome in in different disease states.

1. Introduction

Human cerebrospinal fluid (CSF) is predominantly produced in choroid plexuses, 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 includes mechanical protection of the central nervous system (CNS), homeostasis of the interstitial fluid in the brain, and regulation of neuronal functioning.[3,4] Through direct contact with CNS, the CSF reflects the ongoing physiological or pathological state of CNS most directly.[5,6] Metabolites, peptides, proteins, enzymes, and hormones in CSF are involved in many biological processes, and changes in these compositions are viewed as a sign of pathological alterations in CNS. These biological compositional changes provide an opportunity to mine the CSF for biomarker discovery in neurological diseases.[7]

To date, many studies have characterized the proteome and the proteome-level PTMs of human CSF.[8–12] Some studies have also explored the correlation between proteome-level dysregulation with the progression of neurogenerative diseases, including Alzheimer's disease (AD) and Parkinson’s disease.[13–17] Nevertheless, few reports focus on the small proteins and peptides in CSF, which mostly derive from the processing of larger protein precursors. As one crucial but often underestimated post-translational processing step, the endogenous proteolytic activity results in specific generation of important classes of peptides, such as peptide hormones or cytokines by proteases, along with the unspecific degradation of proteins in metabolism.[18] The occurrence and quantity of the endogenous peptides as protein fragments can therefore depict a

measure of protease and protease inhibitor activities and so many more biological processes and disease states that those typically considered.[18,19]

What's more, the discussion on the post-translational modifications (PTMs) on the endogenous peptidome is scarcer due to the challenge of their mass spectrometry (MS) detection and identification. Yet the PTMs on the endogenous peptides have important functions, for example, they can alter the functional properties of some neuropeptides by increasing the binding affinity to receptors or making peptides more resistant to enzymatic degradation.[20,21] The most common PTMs are C-terminal amidation and pyroglutamic acid at the N-terminus.[22,23] In addition, endogenous peptides may carry various other PTMs, such as acetylation, oxidation, deamidation, methylation, phosphorylation, and glycosylation. Therefore, it is essential to consider multiple variable PTMs during the database search to better profile the landscape of the endogenous peptidome.

Among these PTMs, glycosylation is the most challenging one to study due to the heterogeneity of glycans and glycosylation sites. Glycosylation serves as a key regulatory mechanism controlling protein folding, molecular trafficking, cell adhesion, receptor activation and signal transduction.[24–27] Based on the amino acids that glycans attach to, glycosylation can be classified into two major categories: N-glycosylation and O-glycosylation. Biosynthesis of N-glycosylation is initiated by transferring a pre-assembled 14 monosaccharide complex glycan to asparagine residue (Asn) within the consensus motif (Asn-X-Ser/Thr, X≠P) followed by sequential addition or removal of certain monosaccharide in a well-defined process.[28,29] In contrast, O-glycosylation synthesis involves the attachment of a single monosaccharide to the serine/threonine

(Ser/Thr) residue of a polypeptide without any definable peptide consensus motif and subsequent attachment of numerous diverse monosaccharide residues. As a result, a higher degree of site occupancy, structural heterogeneity, and diversity have been observed in O-glycosylation. Similar to protein modifications, glycosylation can also happen on the endogenous peptides, including neurotransmitters and hormones. As an example, it has been reported that there is an extensive N-/O-glycosylation of gonadotropin, which is a glycoprotein polypeptide hormone.[30] Our group recently reported O-glycosylation on mouse insulin-1B and -2B chains and human insulin-B chain, as well as multiple O-glycoforms of signaling peptides.[31] We also discovered 14 O-glycosylated neuropeptides in the crustacean nervous system.[32] These findings highlight the crucial role of glycosylation on neurotransmitters and peptide hormones, and the necessity to characterize the glycosylation state of endogenous peptides.

For endogenous peptidome in CSF, previous studies only reported PTMs such as acetylation, amidation, phosphorylation, Gln to pyro-Glu conversion, and limited observation of glycan oxonium ions in some MS spectra.[33,34] Even if a small number of O-glycopeptides were eventually identified with lower collision energy during fragmentation, only two O-glycan compositions were found in this case.[34] It is highly possible that there are many more O-glycosylated endogenous peptides undiscovered. Therefore, an approach with a glycosylation-centered analysis workflow will be especially beneficial to systematically evaluate N-/O-glycosylation on endogenous peptides.

Herein, we describe a peptidomic workflow for comprehensive PTM analysis of endogenous peptides including both N-glycosylation and O-glycosylation and other common

peptide PTMs. The protocol consists of CSF endogenous peptide extraction by 10kDa molecular weight cut-off (MWCO), LC-MS/MS analysis with EThcD fragmentation, and a three-step database searching strategy for comprehensive PTM analysis (as shown in **Fig. 1**).[35] With EThcD fragmentation, site-specific information can be obtained for peptide PTM characterization. This workflow can be further adopted to study the endogenous CSF peptides in mild cognitive impairment (MCI) and AD patients, which helps to explore the alteration of peptidome during the disease progression.

2. Materials

2.1 Chemicals and Equipment

1. Ultrapure water, used for all the solutions listed below unless stated otherwise.
2. HPLC grade acetonitrile (ACN).
3. HPLC grade methanol (MeOH).
4. Formic acid (FA).
5. Microfuge tubes with low protein binding.
6. 10 kDa molecular weight cut-off (MWCO) centrifugal filter devices.
7. 50% (v/v) MeOH in water.
8. 70:30 (v/v) aqueous 1 M sodium chloride/MeOH.
9. Protease inhibitor cocktail, complete mini EDTA-free 20× stock solution (we use the product from Roche.) Dissolve 1 tablet in 500 μL water to prepare the stock solution.
10. C18 reversed-phase material to remove salts from samples. (We use Sep-Pak C18 SPE cartridges from Waters).

11. 0.1% (v/v) trifluoroacetic acid (TFA) in water for peptide desalting.
12. 0.1% (v/v) TFA in 40% ACN/water solution for peptide desalting.
13. 0.1% (v/v) TFA in 80% ACN/water solution for peptide desalting.
14. LC mobile phase A: 0.1% (v/v) FA in water. Use LC-MS grade water and FA. Prepare the solution and sonicate it for 15 min to degas the solvent.
15. LC Mobile phase B: 0.1% (v/v) FA in acetonitrile (ACN). Use LC-MS grade ACN and FA. Prepare the solution and sonicate it for 15 min to degas the solvent.
16. A C18 column (75 μm inner diameter × 15 cm, either home-packed or commercially available. We use the homemade column packed with 1.7 μm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661)).

2.2 Instrumentation and Software

1. A refrigerated centrifuge that can achieve 14000 × g and accommodate 1.5 mL microfuge tubes.
2. An ultrasonic cleaner.
3. A vacuum centrifuge/concentrator.
4. Ultraperformance liquid chromatography (UPLC) system (we use the Thermo Dionex UPLC system).
5. A Tribrid Quadrupole-Orbitrap Mass Spectrometer with ETD capability (we use the Thermo Orbitrap Fusion Lumos).
6. Database search software capable of searching intact glycopeptides (we use Byonic developed by Protein Metrics).

3 Methods

3.1 CSF sample collection

1. All study procedures involving human subjects have been approved by the University of Wisconsin Institutional Review Board and abide by the Declaration of Helsinki principles. Each enrollee was provided a signed informed consent form before participation. 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. All MCI and AD participants were diagnosed via applicable clinical criteria in standardized and multidisciplinary consensus conferences.[36,37]
2. 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.[38,39]
3. CSF was collected by lumbar puncture of individuals under written informed consent.
4. To enable in-depth profiling of CSF peptidome, 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. (see **Note 1**)

3.2 CSF sample processing

1. Add 50 µL of Protease inhibitor cocktail 20× stock solution to 1 mL of CSF immediately before use.
2. Perform endogenous peptide separations using 10 kDa MWCO centrifugal filters. Before the MWCO separation, wash the filters three times to remove the contaminants from the filter and achieve optimal peptide coverage. The three washes are 500 µL of 50% methanol in water, 500 µL of water, and 400 µL of 70:30 aqueous 1 M sodium chloride/MeOH. For each step of washing, centrifuge the MWCO filters at 14000 × g for 5 min.
3. Add the CSF sample to the MWCO filters and centrifuge the MWCO filters at 14000 × g for 30 min at 4ºC. Collect the flow-through from this step in a 1.5 mL microfuge tube. (see **Note 2**).
4. Dry down the flow-through in vacuo.

3.3 Peptide desalting

1. Condition a C18 cartridge with 3×1 mL ACN and 3×1 mL 0.1% TFA.
2. Resuspend the peptide sample in 1 mL of 0.1% TFA and load the sample onto the desalting cartridge. (see **Note 3**)
3. Wash the mixture using 3×1 mL 0.1% TFA. If the elution is slow, apply positive pressure but avoid a flow over one drop per second. (see **Note 4**)
4. Elute peptides using 700 µL 40% ACN, 0.1% TFA followed by 700 µL 80% ACN, 0.1% TFA. Pool the eluate together.

5. Dry down the eluate in vacuo.

3.4 LC-MS/MS analysis

1. Resuspend the sample in 0.1% FA.
2. Put the resuspended sample in a vial and load the sample to LC-MS.
3. Set the LC gradient as follows: 3%–30% A (18–98min), 30%–75% A (100–108 min) and 75%–95% A (108–118min).
4. Operate the mass spectrometer in data dependent acquisition (DDA) mode. Acquire MS1 scan from from m/z 400–1800 (120,000 resolution, 4e5 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, 3e5 AGC, 100 ms injection time). Perform EThcD 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. (see **Note 5**)

3.5 Data analysis

1. Search the LC-MS/MS raw data with PTM-centric search engine Byonic.
2. Perform three consecutive searches as show in **Fig. 2**. Set precursor mass error tolerance of 10 ppm, and a fragment mass error tolerance of 0.01 Da.
3. Set unspecific enzyme digestion.
4. Use the whole human proteome database in the 1st search. Set total “rare” modification number to 1 and include oxidation of methionine residues

(+15.99492 Da), amidation at peptide C-terminus (-0.984016), acetylation at peptide N-terminus and lysine, serine (+42.010565), and Gln to pyro-Glu conversion (-17.026549) as dynamic modifications.

5. In the 2nd search, build a focused protein database on the protein precursors identified in the 1st search at 1% FDR together with proteins reported in literature.[33,34]
6. Use the same dynamic modification settings as in the 1st search, except for adding N-glycosylation and O-glycosylation as "common" modification, with total common modification set as 1. Use the human glycan database embedded in Byonic containing 182 N-glycans and 70 O-glycans.
7. The 2nd search yields a list of identified N-glycans and O-glycans. Build a focused glycan database by using the identified glycans.
8. Set the protein database as the focused protein database in the 3rd search.
9. Set dimethylation at peptide N-terminus, lysine, arginine (+28.0313), deamidation at asparagine (+0.984016), methylation at peptide N-terminus, lysine, arginine, glutamine (14.01565), phosphorylation at serine, tyrosine, threonine (+79.966331) as dynamic "rare" modifications.
10. Load the focused glycan database. Set N-glycosylation and O-glycosylation as "common" modifications, with the total common modification number set as 1. Perform Byonic search.

11. After searching, filter peptides with PSMs of FDR < 1%, Byonic score > 50 and PEP 2D < 0.01. For endogenous peptides with PTMs, include Delta Mod Score > 40 in the filtering criteria. (see **Note 6**)

4 Notes

1. Here the CSF samples from 16 individuals are pooled to reduce biological variation and allow deeper profiling of low-abundance peptides.[40] It is also recommended to study on individual control, MCI and AD subjects for studies like biomarker discovery if the size of CSF samples collected, instrument/labor time, and financial budget allows.
2. Many times, the MWCO filters become clogged when there are too many proteins loaded. To speed up the centrifugation step, the CSF samples can be loaded to separate filters, and the flow-through can be pooled together later. The centrifugation may take longer time than 30 min depending on the sample type.
3. It is important to estimate the loading capacity of the C18 cartridge and choose the right size for the experiment. For example, the peptide loading capacity estimate for the Waters C18 Sep-Pak cartridge is about 0.1% to 0.3% weight of the sorbent.
4. Unlike other Waters Oasis cartridges, Sep-Pak C18 cartridge is not a water wettable product, so be careful not to let the cartridge go dry during the desalting.
5. The commonly used HCD fragmentation generates B/Y ions and abundant oxonium ions for glycan identification and b/y ions for peptide sequence information, but it does not provide much information for PTM site localization. On the other hand, c/z ions produced in ETD-MS/MS provide information on the PTM 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. To take advantage of both modes, electron-transfer and higher-energy collision dissociation (EThcD) is developed as a "hybrid" dissociation method to better identify and localize peptide PTMs. **Fig. 3A** shows a representative EThcD spectrum of an O-glycosylated peptide. Its glycan composition can be assigned through the 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), 274.05 (NeuAc-18), 292.11 (NeuAc), HexNAcHex (366.14), and HexNAcHexNeuAc (657.23). Its O-glycosite is unambiguously localized at the 6^{th} serine residue instead of the 5^{th} threonine and 12^{th} serine residues by using the c3-c7 ions. **Fig. 3B** demonstrates the benefit of using EThcD for other labile PTMs like phosphorylation. Along with other b/y, c/z ions and the 98 Da neutral loss, the detection of c4 and c5 ions helps to unambiguously localize the phosphorylation site at the 5^{th} serine residue of an endogenous phosphorylated peptide VDPKSKEEDKH.

6. Though the output of Byonic has been filtered by several parameters to control the FDR rate, we still recommend to manually inspect the PSMs, especially the glycol-PSMs to avoid potential mismatch and misassignment.

Reference

(1) McComb, J. G. Recent Research into the Nature of Cerebrospinal Fluid Formation and Absorption. *J. Neurosurg.* **1983**, *59*, 369–383.

(2) Sakka, L.; Coll, G.; Chazal, J. Anatomy and Physiology of Cerebrospinal Fluid. *European Ann Otorhinolaryngology 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*, 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; Springer, **1998**; pp 59–65.

(5) Zhang, J. Proteomics of Human Cerebrospinal Fluid – the Good, the Bad, and the Ugly. *Proteom. - Clin. Appl* **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*, 39–64.

(7) Yuan, X.; Desiderio, D. M. Proteomics Analysis of Human Cerebrospinal Fluid. *J. Chromatogr. B* **2005**, *815*, 179–189.

(8) 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* **2008**, *7* (01), 386–399.

(9) Macron, C.; Lane, L.; Galindo, A. N.; Dayon, L. Identification of Missing Proteins in Normal Human Cerebrospinal Fluid. *J Proteome Res* **2018**, *17* (12), 4315–4319.

(10) Macron, C.; Lane, L.; Galindo, A. N.; Dayon, L. Deep Dive on the Proteome of Human Cerebrospinal Fluid: A Valuable Data Resource for Biomarker Discovery and Missing Protein Identification. *J Proteome Res* **2018**, *17* (12), 4113–4126.

(11) Jankovska, E.; Svitek, M.; Holada, K.; Petrak, J. Affinity Depletion versus Relative Protein Enrichment: A Side-by-Side Comparison of Two Major Strategies for Increasing Human Cerebrospinal Fluid Proteome Coverage. *Clin Proteom* **2019**, *16* (1), 9.

(12) Chen, Z.; Yu, Q.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific Analysis of N-Glycoproteome in Human Cerebrospinal Fluid and Glycosylation Landscape Changes in Alzheimer's Disease. *Mol. Cell Proteomics* **2021**, *20*, 100081.

(13) 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. *Proteom. - Clin. Appl* **2016**, *10* (12), 1225–1241.

(14) Dayon, L.; Galindo, A. N.; Wojcik, J.; Cominetti, O.; Corthésy, J.; Oikonomidi, A.; Henry, H.; Kussmann, M.; Migliavacca, E.; Severin, I.; Bowman, G. L.; Popp, J. Alzheimer Disease Pathology and the Cerebrospinal Fluid Proteome. *Alzheimer's Res Ther* **2018**, *10* (1), 66.

(15) Pedrero-Prieto, C. M.; García-Carpintero, S.; Frontiñán-Rubio, J.; Llanos-González, E.; García, C. A.; Alcaín, F. J.; Lindberg, I.; Durán-Prado, M.; Peinado, J. R.; Rabanal-Ruiz, Y. A Comprehensive Systematic Review of CSF Proteins and Peptides That Define Alzheimer's Disease. *Clin Proteom* **2020**, *17* (1), 21.

(16) Wang, D.; Ma, M.; Huang, J.; Gu, T.-J.; Cui, Y.; Li, M.; Wang, Z.; Zetterberg, H.; Li, L. Boost-DiLeu: Enhanced Isobaric N,N-Dimethyl Leucine Tagging Strategy for a Comprehensive Quantitative Glycoproteomic Analysis. *Anal. Chem.* **2022**, *94* (34), 11773–11782.

(17) Karayel, O.; Winter, S. V.; Padmanabhan, S.; Kuras, Y. I.; Vu, D. T.; Tuncali, I.; Merchant, K.; Wills, A.-M.; Scherzer, C. R.; Mann, M. Proteome Profiling of Cerebrospinal Fluid Reveals Biomarker Candidates for Parkinson's Disease. *Cell Reports Medicine* **2022**, *3* (6), 100661.

(18) Möhring, T.; Kellmann, M.; Jürgens, M.; Schrader, M. Top-down Identification of Endogenous Peptides up to 9 KDa in Cerebrospinal Fluid and Brain Tissue by Nanoelectrospray Quadrupole Time-of-Flight Tandem Mass Spectrometry: Top-down Identification of Endogenous Peptides. *J. Mass Spectrom.* **2005**, *40* (2), 214–226.

(19) Schrader, M.; Schulz-Knappe, P. Peptidomics Technologies for Human Body Fluids. *Trends Biotechnol* **2001**, *19*, 55–60.

(20) Hökfelt, T.; Broberger, C.; Xu, Z.-Q. D.; Sergeyev, V.; Ubink, R.; Diez, M. Neuropeptides — an Overview. *Neuropharmacology* **2000**, *39* (8), 1337–1356.

(21) Zhang, X.; Petruzziello, F.; Zani, F.; Fouillen, L.; Andren, P. E.; Solinas, G.; Rainer, G. High Identification Rates of Endogenous Neuropeptides from Mouse Brain. *J Proteome Res* **2012**, *11* (5), 2819–2827.

(22) Eipper, B. A.; Milgram, S. L.; Husten, E. J.; Yun, H.-Y.; Mains, R. E. Peptidylglycine α -Amidating Monooxygenase: A Multifunctional Protein with Catalytic, Processing, and Routing Domains: Peptidylglycine α -Amidating Monooxygenase. *Protein Sci* **1993**, *2* (4), 489–497.

(23) Hayakawa, E.; Menschaert, G.; Bock, P.-J. D.; Luyten, W.; Gevaert, K.; Baggerman, G.; Schoofs, L. Improving the Identification Rate of Endogenous Peptides Using Electron Transfer Dissociation and Collision-Induced Dissociation. *J Proteome Res* **2013**, *12* (12), 5410–5421.

(24) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126*, 855–867.

(25) 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*, 6527–6534.

(26) 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.

(27) Chen, Z.; Huang, J.; Li, L. Recent Advances in Mass Spectrometry (MS)-Based Glycoproteomics in Complex Biological Samples. *Trends Analyt Chem.* **2018**, *118* (Cell 126 2006), 880–892.

(28) Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology.*; Cold Spring Harber Laboratory Press, New York; **1998**.

(29) Chen, Z.; Glover, M. S.; Li, L. Recent Advances in Ion Mobility–Mass Spectrometry for Improved Structural Characterization of Glycans and Glycoconjugates. *Current opinion in chemical biology* **2018**, *42*, 1–8.

(30) Thotakura, N. R.; Blithe, D. L. Glycoprotein Hormones: Glycobiology of Gonadotrophins, Thyrotrophin and Free α Subunit. *Glycobiology* **1995**, *5*, 3–10.

(31) 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.

(32) Cao, Q.; Yu, Q.; Liu, Y.; Chen, Z.; Li, L. Signature-Ion-Triggered Mass Spectrometry Approach Enabled Discovery of N- and O-Linked Glycosylated Neuropeptides in the Crustacean Nervous System. *J. Proteome Res.* **2019**, *19* (2), 634–643.

(33) Hölttä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J. Peptidome Analysis of Cerebrospinal Fluid by LC-MALDI MS. *PLoS One* **2012**, *7*, e42555.

(34) 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*, 386–399.

(35) Chen, Z.; Wang, D.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Yu, Q.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients. *ACS Chem. Biol.* **2022**, *17* (11), 3059–3068.

(36) 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.

(37) 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*, 2462–2475.

(38) Apweiler, R.; Hermjakob, H.; Sharon, N. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473*, 4–8.

(39) 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*, 8260–8287.

(40) Kendziorski, C.; Irizarry, R. A.; Chen, K.-S.; Haag, J. D.; Gould, M. N. On the Utility of Pooling Biological Samples in Microarray Experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (12), 4252–4257.

Figures

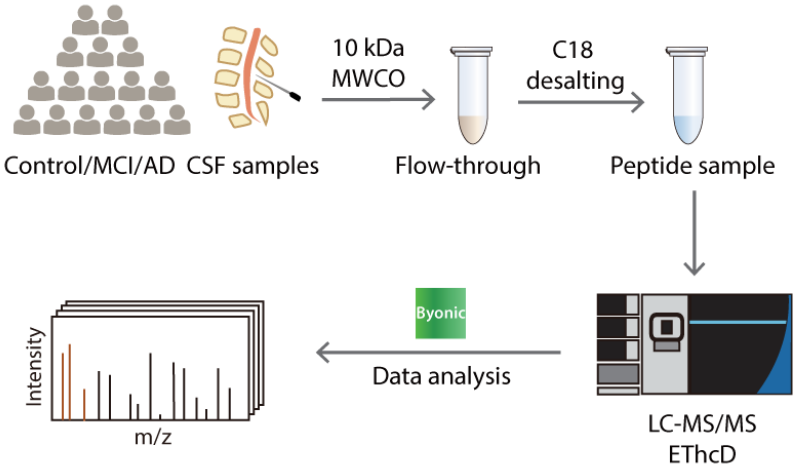


Figure 1. The general experimental workflow.

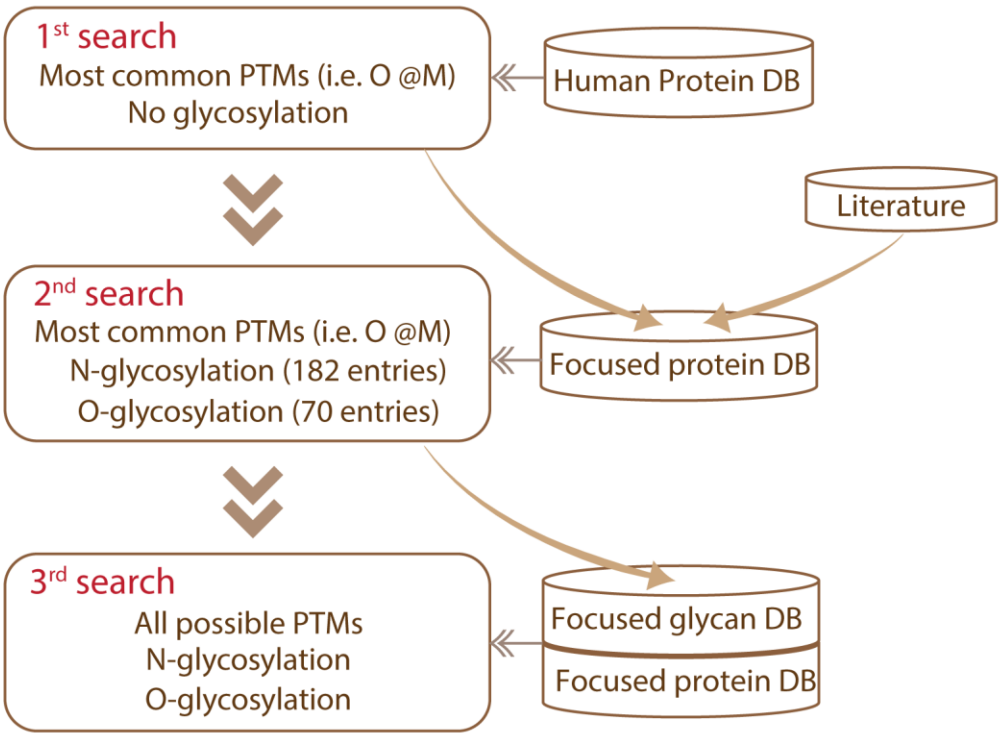


Figure 2. The three-step searching strategy for comprehensive PTMs analysis of CSF endogenous peptides.

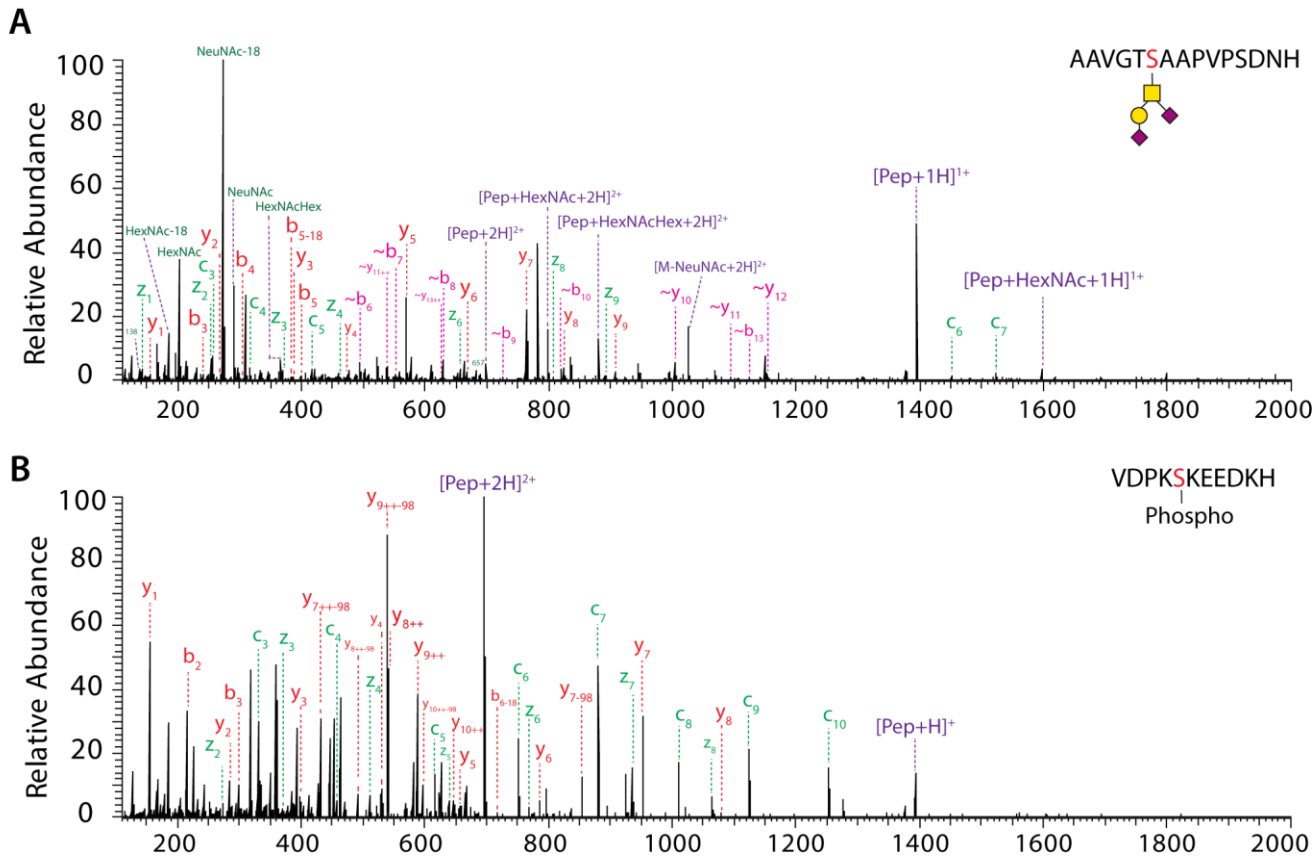


Figure 3. Representative EThcD fragmentation spectra of (A) an endogenous O-glycopeptide, and (B) an endogenous phosphorylated peptide.