PROTEIN GLYCOSYLATION — ADVANCES IN IDENTIFICATION, CHARACTERIZATION AND BIOLOGICAL FUNCTION ELUCIDATION USING MASS SPECTROMETRY

EDITED BY: Ganglong Yang, Hui Zhang, Wen Yi, Shi Yan and Liwei Cao
PUBLISHED IN: Frontiers in Chemistry

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PROTEIN GLYCOSYLATION – ADVANCES IN IDENTIFICATION, CHARACTERIZATION AND BIOLOGICAL FUNCTION ELUCIDATION USING MASS SPECTROMETRY

Topic Editors:
Ganglong Yang, Jiangnan University, China
Hui Zhang, Johns Hopkins University, United States
Wen Yi, Zhejiang University, China
Shi Yan, University of Veterinary Medicine Vienna, Austria
Liwei Cao, Johns Hopkins University, United States

Table of Contents

Editorial: Protein Glycosylation—Advances in Identification, Characterization and Biological Function Elucidation Using Mass Spectrometry

Ganglong Yang[1]*, Hui Zhang[2,3], Wen Yi[4], Shi Yan[5] and Liwei Cao[2]

[1]*The Key Laboratory of Carbohydrate Chemistry & Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, China,* [2]*Department of Pathology, Johns Hopkins University, Baltimore, MD, United States,* [3]*Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, United States,* [4]*Department of Hepatobiliary and Pancreatic Surgery, Zhejiang Provincial Key Laboratory of Pancreatic Disease, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China,* [5]*Institut für Parasitologie, Veterinärmedizinische Universität, Wien, Austria*

Editorial on the Research Topic

Protein Glycosylation—Advances in Identification, Characterization and Biological Function Elucidation using Mass Spectrometry

Glycosylation is one of the most common post-translational modifications, with over 50% of human proteins having been reported to be glycosylated. The identification and functional validation of protein glycosylation are important to reveal the roles of glycoproteins in biological processes. Until now, fewer than 20% of predicted glycoproteins in Uniprot were experimentally verified. Therefore, a comprehensive, rapid and sensitive methodology to map the protein glycosylation events is required more than ever.

A number of high-quality articles relevant to characterizing glycoproteins are reported in our Research Topic "Protein Glycosylation-Advances in Identification, Characterization and Biological Function Elucidation using Mass Spectrometry," which describe a variety of experimental techniques especially high-resolution mass spectrometry (MS). At first, fast and highly specific enrichment methods for glycoprotein/glycopeptides preparation are imperative for broad application of glycoprotein analysis in high-throughput clinical samples. Fan et al. developed an ultrafast sample preparation method for N-glycoproteome using thermoresponsive magnetic fluid-immobilized enzymes, in which protein digestion and deglycosylation of glycopeptide were done within 3 min. Chen et al. developed a method to sequentially enrich glycopeptides and phosphopeptides using a TiO_2 material, which were then sequentially subjected to MS for characterization of glycosylation sites and phosphosites, respectively. The two methods facilitated comprehensive analyses of N-linked glycosylation events by MS.

Glycosylation is critical for physiological and pathological functions of glycoproteins. Tumor growth benefits from the ability of cancer cells to bypass cellular division checkpoints, evade death signals and immune surveillance, and migrate to metastatic sites. Glycosylation plays critical roles in all of these processes (Reily et al., 2019). In order to dissect the functional roles of glycosylation in human diseases, new technologies are needed for improvement of the depth and breadth of glycoproteome analysis, especially for the characterization of intact glycopeptides. Analysis of

intact glycopeptides could provide information regarding to glycosylation sites, glycan compositions at specific glycosylation sites of glycoproteins. It preserves biological context of the modification and enables us to understand proteome-wide glycan heterogeneity. In this Research Topic, Tabang et al. comprehensively reviewed the advances in MS-based technologies focusing on characterization of glycome and glycoproteome of diabetes, pancreatitis, and pancreatic cancer. The future of MS-based methods was forecasted to discover new protein targets for early detection of human diseases. Liu et al. developed a N-linked and O-linked glycosylation mapping method to identify and quantify the occurrence of each glycoform of human plasma fibronectin in a site-specific manner, in which CID and stepped normalized collision energy (sNCE)-HCD tandem MS were deployed. These site-specific glycosylation patterns of fibronectin in human plasma can facilitate functional analyses and development of corresponding therapeutics. Zhao et al. integrated the proteomic and glycoproteomic data to identify potential clinical (glyco)protein targets for early detection of type 2 diabetes (T2D). They identified alterations of site-specific N-linked glycosylation related to the complement activation pathways in T2D patients, which were not observed by corresponding proteome measurements. It demonstrated that MS-based glycoproteomics method is a powerful tool to understand (pre)diabetes. Shu et al. applied the lectin blot and intact O-linked glycopeptide MS analysis to discover aberrant O-glycosylation of Haptoglobin in hepatocellular carcinoma, revealing up-regulation of most O-glycopeptides in HCC patient serum. The study by Demus et al. sheds new insights on glycosylation of apo-CIII that is able to regulate the triglyceride clearance, facilitating our understanding of the role of apo-CIII in the regulation of lipid metabolism in various disease settings.

Glycomics has become more essential and significant in the cancer studies and glycans were deemed as potential biomarkers in cancer. The liquid biopsies, e.g., blood and urine, are great sources for screening and characterization of targets for early detection of human diseases. Moreover, mass spectrometry imaging (MSI) is a well-established technique to spatially map biomolecules across fresh frozen or formalin-fixed paraffin-embedded (FFPE) tissue sections (Briggs et al., 2019). Blaschke et al. analyzed the N-glycome of urine, urine EPS, prostatic fluids, urine EPS-derived extracellular vesicles using MALDI-MS, and the N-glycan profile of prostatic tissue using MSI. Over 100 N-linked glycan compositions were detected, and a subset of N-glycans present in fluids were found to be derived from the gland lumens. The developed N-glycan profiling method is able to analyze large clinical cohorts, and is adaptable to characterize other biofluids. Shu et al. deployed the lectin microarray and MS to analyze protein glycosylation of saliva of the patients with Esophageal squamous cell carcinoma (ESCC). The specific glycopatterns, e.g., sialylation and fucosylation, were significantly altered in ESCC patients in comparison to healthy controls, and DSA detection was thought to be a potential diagnostic tool.

SARS-CoV-2, the causative pathogen of COVID-19, induces fever, severe respiratory illness, and pneumonia (Watanabe et al., 2020), which caused approximate 270 million infections and more than 5 million deaths until now. The SARS-CoV-2 S gene encodes a glycoprotein with 22 potential N-linked glycosites and dozens of potential O-linked glycosites, which likely play critical roles in protein folding and immune evasion. Wang et al. compared the N-glycosylation profiles of recombinant S proteins produced by CHO and HEK cells, revealing higher levels of complex type and sialylation type of glycans on CHO-expressed S protein and a decreased level of high-mannose glycans on this protein relative to the S protein produced in HEK cells. Zhang et al. compared the O-glycosylation profiles of recombinant S proteins produced by insect and HEK cells using HCD and EThcd MS-based methods, and found that most of the O-glycosites in S protein from human cells were sialylated. To comprehensively analyze O-glycosylation of S protein, Cui et al. developed an O-glycopeptide enrichment method with a dual-functional histidine-bonded silica material, resulting in identification of 46 O-glycosites.

Due to the microheterogeneity and nonlinear structural complexity of glycans, characterization of glycan structures is one of challenge tasks in the field. MS is a powerful tool to analyze the glycan structures and there is a great progress partially facilitated by the advance of instruments in the past decades. However, processing and characterization of numerous fragmental glycan spectrum from MS raw files are still challenging. Zhang et al. developed an automatically GUI-based tool (GlycanGUI) for annotation and quantification of glycan compositions. GlycanGUI could interpret the data generated by different separation columns, MS instruments and/or buffers, and even different laboratories. Huang et al. presented an automated tool for processing MALDI-MS-based glycan isotope labeling data (gQuant), which was designed with a set of dedicated algorithms, including spectral preprocessing, glycan mapping, quantitation, and ratio calculation. Wang et al. also presented a new algorithm (hepPareser) to decipher the main components of Low-molecular-weight heparins based on the LC/MS data. These tools will be greatly beneficial for structural and functional studies of protein glycosylation.

O-GlcNAcylation is a prevalent form of posttranslational modifications on the hydroxyl group of serine and/or threonine residues (Torres and Hart, 1984). The systematical characterization of O-GlcNAcylation is needed for revealing its functional roles in human diseases. Zhu and Yi reviewed chemistry-assisted methods for characterization of O-GlcNacylation, which provided comprehensive insights for the labeling and identification of O-GlcNAcylation. Yin et al. systematically reviewed the MS-based methodologies for qualitative or quantitative characterization of O-GlcNAcylation.

Glycans from glycoconjugates are synthesized through a series of reactions mediated by glycosyltransferases (GTs) or glycoside hydrolases (GHs). The CRISPR-Cas9 technology enables us to edit the glyco-related gene expression, and

thus facilitates our understanding of the functions of glycosylation in cells. Yang et al. applied the intact glycopeptides analysis method to reveal differential expression of core-fucosylation between normal CHO cells and CHO cells with Fut8 knock-out, revealing knock-out of FUT8 influenced core-fucosylation of glycoproteins as well as other processes of glycosylation synthesis, resulting in alteration of protein glycosylation. With this, the relationship of glycan compositions, structures and modified proteins could be mapped (Narimatsu et al., 2019; Huang et al., 2021).

REFERENCES

Briggs, M. T., Condina, M. R., Ho, Y. Y., Everest-Dass, A. V., Mittal, P., Kaur, G., et al. (2019). MALDI Mass Spectrometry Imaging of Early- and Late-Stage Serous Ovarian Cancer Tissue Reveals Stage-specific N-Glycans. *Proteomics* 19, e1800482. doi:10.1002/pmic.201800482

Huang, Y.-F., Aoki, K., Akase, S., Ishihara, M., Liu, Y.-S., Yang, G., et al. (2021). Global Mapping of Glycosylation Pathways in Human-Derived Cells. *Develop. Cel* 56, 1195–1209. doi:10.1016/j.devcel.2021.02.023

Narimatsu, Y., Joshi, H. J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., et al. (2019). An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Mol. Cel* 75, 394–e5. doi:10.1016/j.molcel.2019.05.017

Reily, C., Stewart, T. J., Renfrow, M. B., and Novak, J. (2019). Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* 15, 346–366. doi:10.1038/s41581-019-0129-4

Torres, C. R., and Hart, G. W. (1984). Topography and Polypeptide Distribution of Terminal N-Acetylglucosamine Residues on the Surfaces of Intact Lymphocytes. Evidence for O-Linked GlcNAc. *J. Biol. Chem.* 259, 3308–3317. doi:10.1016/s0021-9258(17)43295-9

In conclusion, the articles collected in this Research Topic will facilitate the identification, characterization and functional elucidation of protein glycosylation, and pave the way to future ambitious challenges in glycobiology.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 369, 330–333. doi:10.1126/science.abb9983

The Abnormal Glycopatterns of Salivary Glycoproteins in Esophageal Squamous Cell Carcinoma Patients

Jian Shu[1], Jun Ma[2,3], Xiameng Ren[1], Jian Wang[4], Yan Wang[1], Kun Zhang[1], Hanjie Yu[1], Xiangqian Guo[5]* and Zheng Li[1]*

[1]Laboratory for Functional Glycomics, College of Life Sciences, Northwest University, Xi'an, China, [2]Institute of Digestive Disease of Zhengzhou University, Zhengzhou, China, [3]Department of Clinical Laboratory, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [4]Department of Oncology, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [5]Institute of Biomedical Informatics, Cell Signal Transduction Laboratory, Bioinformatics Center, Henan Provincial Engineering Center for Tumor Molecular Medicine, School of Basic Medical Sciences, Henan University, Kaifeng, China

Glycosylation is one of the most crucial posttranslational modifications of proteins, containing a remarkable amount of biological information. The alteration of glycosylation is closely associated with certain diseases. Exploring glyco-code in the development of diseases is a hot topic in recent years. Esophageal squamous cell carcinoma (ESCC) is the primary pathological histology in developing countries and a severe threat to human health. Although the glycan profiles in the blood samples of ESCC patients were analyzed using glycomic and glycoproteomic methods, the difference of salivary glycopatterns between healthy subjects and ESCC patients is not explicit yet. In the present study, ESCC patients (n = 16) and healthy volunteers (HVs, n = 25) were enrolled. The glycomic strategy combining lectin microarray and lectin blotting was employed to investigate and confirm the altered salivary glycopatterns. Datura stramonium (DSA) was selected to isolate the GlcNAc or Galβ1-4GlcNA-containing glycoproteins due to the distinct difference between ESCC patients and HVs. The N-glycans from DSA-enriched glycoproteins were released by PNGase F and further identified by MALDI-TOF/TOF-MS to obtain the precise structural information of the altered glycans. As a result, the glycopatterns recognized by 13 lectins (e.g., ECA, RCA120, and DSA) showed significant alterations in ESCC patients' saliva. The ESCC patients showed higher levels of GalNAc and Gal, sialic acid, and GlcNAc expression profiles and lower levels of mannose and fucose expression profiles. The MALDI-TOF/TOF-MS results indicated that the proportion of the GlcNAc or Galβ1-4GlcNAc-containing N-glycans was increased in ESCC patients (79.04%) compared with HV (63.20%), which was consistent with the results of lectin microarrays. Our findings provide comprehensive information to understand the complex physiological changes in ESCC patients. And the altered salivary glycopatterns such as GlcNAc or Galβ1-4GlcNAc-containing N-glycans recognized by DSA might serve as potential biomarkers for the diagnosis of ESCC patients.

Keywords: esophageal squamous cell carcinoma, protein glycosylation, saliva, lectin microarrays, MALDI-TOF/ TOF-MS

INTRODUCTION

Esophageal cancer (EC) ranks as the seventh most prevalent cancer with more than 570 thousand new cases and the sixth most lethal cancer with over 500 thousand deaths worldwide (Bray et al., 2018). EC is also a common type of malignant cancer in China, ranking fourth in terms of diagnosis and death (Chen et al., 2016). China is even responsible for more than 50% of the world's EC new cases because of the large population size and high incidence. As one of the major histologic types of EC, esophageal squamous cell carcinoma (ESCC) is still accounted for the vast majority of EC in developing countries, while the other crucial histologic kind of EC, esophageal adenocarcinoma, has become the predominant type in developed countries, and the incidence rate continues to rise (Malhotra et al., 2017). ESCC is a serious threat to human health, is distributed anywhere in the esophagus, and is considered positively associated with heavy alcohol drinking, tobacco smoking, and nutrients lacking. It is reported that more than 50% of ESCC patients were diagnosed only at the advanced stages due to the unspecific symptoms and lack of early biomarkers, resulting in poor prognosis and a low 5-year survival rate (Ohashi et al., 2015; Kojima and Doi, 2017).

The assembly of glycans is complex enzymatic progress catalyzed by a series of glycosyltransferases and glycosidases, which starts in the endoplasmic reticulum and matures in the Golgi apparatus (Chao et al., 2020). As the most prevalent posttranslational modifications, glycosylation occurs on approximately 70% of human proteins and involves a wide range of biological processes, which could partially reflect human's physiological states (Yu et al., 2020a; de Haas et al., 2020). The alterations of glycosylation have been widely detected in various cancers and have shown a profound correlation with carcinogenesis (Oliveira-Ferrer et al., 2017; Verhelst et al., 2020). Overexpressions of sialylation, fucosylation, branched glycans, and truncated O-glycan are considered the most widely occurring cancer-associated glycans changes in tissue and cell samples (Pinho and Reis, 2015). Although altered glycans were detected in human cancer tissue, cells, serum, and urine samples decades ago, investigations in the past few years have demonstrated that altered glycosylation is also present in saliva samples. Compared with the most commonly used diagnostic fluids of blood and urine, saliva offers the advantages of being noninvasive, easy, secure, and cost-effective, which prompts an interest in evaluating its diagnostic utility value (Nunes et al., 2015). Previous studies have demonstrated that different types of diseases (e.g., gastric cancer (GC), type 2 diabetes mellitus (T2DM), hepatopathy, breast disease oral lichen planus, Keshan disease, and oral ulcer) could induce the different alternations of the salivary protein glycopatterns (e.g., fucosylation in GC and sialylation in T2DM) (Zhang et al., 2016; Fang et al., 2017; Wang et al., 2017; Liu et al., 2018; Shu et al., 2018; Yu et al., 2020b). The accumulating data also indicated that the saliva glycopatterns might serve as potential biomarkers to classify cancer cases from controls. The high-throughput glycomic method of lectin microarray could be used to directly analyze the glycosylation of crude samples without the liberation of glycans, providing a global snapshot

TABLE 1 | Demographic and clinical information of ESCC patients and HV subjects in this study.

Variables	HVs	ESCC	*p*-value[a]
Numbers (n)	25	16	
Age (years)			0.137
≤ 60	11 (44%)	7 (43.75%)	
> 60	14 (56%)	9 (56.25%)	
Gender			0.852
Male	16 (64%)	10 (62.5%)	
Female	9 (37%)	6 (37.5%)	
Pathological (AJCC)[b]			
I-II	—	7 (43.75%)	
III-IV	—	9 (56.25%)	
Differentiation			
Well	—	3 (18.75%)	
Moderate	—	8 (50%)	
Poor	—	5 (31.25%)	

[a]p value was derived from the nonparametric Mann–Whitney test.
[b]AJCC, American Joint Committee on Cancer staging system (7th edition).

of glycosylation state in its native context (Hirabayashi et al., 2013; Yu et al., 2020a). As an efficient screening tool, the lectin microarrays usually work together with lectin blotting and mass spectrometry for screening, analysis, and validation of the important glycopatterns of glycoprotein.

In this study, the altered salivary glycopatterns related to ESCC were investigated and confirmed by the integrated glycomics strategy. To acquire precise structural information of the altered glycans between ESCC and healthy volunteers (HV) cases, the glycoproteins were isolated by lectin-mediated affinity capture, and the N-glycans were released and purified from the isolated glycoproteins and then identified using MALDI-TOF/TOF-MS. The purpose of this study is to clarify the altered salivary protein glycopatterns related to ESCC and to identify the precise structures of the altered salivary glycopatterns, which may help us to understand the complex physiological changes in ESCC patients.

MATERIALS AND METHODS

Study Approval and Population

The collection and use of the whole saliva for research in this study were approved by the Ethical Committee of Northwest University (Xi'an, China) and the Second Affiliated Hospital of Zhengzhou University (Zhengzhou, China). Written informed consent was received from participants. And this study was conducted by the ethical guidelines of the Declaration of Helsinki. After a standardized endoscopic procedure and histopathological evaluation, the individuals diagnosed with primary ESCC (n =16) were enrolled in this study. The age- and sex-matched HVs (n = 25) were recruited from the health checkup center in the same hospital where they underwent a medical examination. All the participants had no significant differences between the two groups regarding demographic, socioeconomic, and lifestyle characteristics, and those who received preoperative radiotherapy, chemotherapy, chemoradiotherapy, or antibiotic therapy were excluded from this study. The clinical characteristics of HVs and ESCC patients were summarized in **Table 1**.

FIGURE 1 | The salivary glycopatterns of HVs and ESCC patients. **(A)** The layout of the lectin microarray. A total of 37 lectins were dissolved in the recommended buffer and spotted on lectin microarray; each lectin was spotted in triplicate per block. **(B)** Glycopattern of HVs **(left)** and ESCC patients **(right)** detected by the lectin microarray. The lectins with increased NFIs in ESCC patients are marked with red boxes, and those with decreased NFIs are marked with white boxes. **(C)** Unsupervised hierarchical clustering analysis of lectins and all samples. The samples were listed in columns, and the lectins were listed in rows. HVs and ESCC patients were clustered based on the correlation similarity of salivary glycopatterns. The color and intensity of each square indicated expression levels relative to other data in the row. Red, high; green, low; black, medium. **(D)** Principal component analysis of 37 lectins for all samples. HV and ESCC samples were indicated by azure circles and blue squares, respectively. PC1: principal components 1. PC2: principal components 2.

Whole Saliva Collection and Preparation

The collection protocol has been described in previous literature (Liu et al., 2018; Shu et al., 2018). The collected whole saliva was centrifuged at 10,000 g at 4°C for 15 min to remove insoluble components. And the cocktail of protease inhibitor (Sigma-Aldrich, United States) was added to the collected supernatant according to the manufacturer's recommendations and lyophilized at −80°C until use.

Lectin Microarrays and Data Analysis

The lectin microarrays were produced using 37 lectins with different binding preferences covering N- and O-linked glycans (Shu et al., 2017; Yu et al., 2020b). The Cy3-labeled glycoproteins were incubated on the lectin microarray with gentle rotation in the dark (37°C, 3 h). The microarrays were washed with PBST and PBS, respectively, centrifuged dry, and scanned immediately using a Genepix 4000B confocal scanner (Axon Instruments, United States). The generated images were analyzed by Gene pix software (version 6.0, Axon Instruments Inc., Sunnyvale, CA). The unsupervised average hierarchical cluster analysis (HCA) and principal component analysis (PCA) were performed by Expander 6.0 (http://acgt.cs.tau.ac.il/expander/) and Multivariate Statistical

Package (United Kingdom), respectively. And the p value was derived from the nonparametric Mann–Whitney test using the GraphPad Prism software (Version 7.0, GraphPad Software, Inc., San Diego, CA).

Lectin Blotting Analysis

The expression levels of glycan structures were analyzed by lectin blotting as described previously (Zhang et al., 2019; Yu et al., 2020b). The pooled salivary protein was separated by 10% SDS-PAGE and transferred onto PVDF membrane (0.22 mm Millipore, Bedford, MA, United States). The PVDF membrane was blocked by the Carbo-Free Blocking Solution (Vector Labs, Burlingame, CA) and incubated with Cy5-labeled DSA, ECA, LCA, PSA, and UEA-I, respectively. The membrane was scanned by the STORM FluorImager (Molecular Dynamics, Sunnyvale, CA, United States) and measured using the ImageJ software (NIH).

Preparation of DSA-Magnetic Particle Conjugates

The epoxy-coated magnetic particles (2 mg) were rinsed with ethanol and coupling buffer (5 mM NaB_4O_7, 180 mM H_3BO_4,

FIGURE 2 | Alterations of salivary glycopatterns in ESCC patients compared to HVs. **(A)** The NFIs of eight lectins were significantly increased in ESCC patients compared to HVs. **(B)** The NFIs of five lectins were significantly decreased in ESCC patients compared to HVs. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
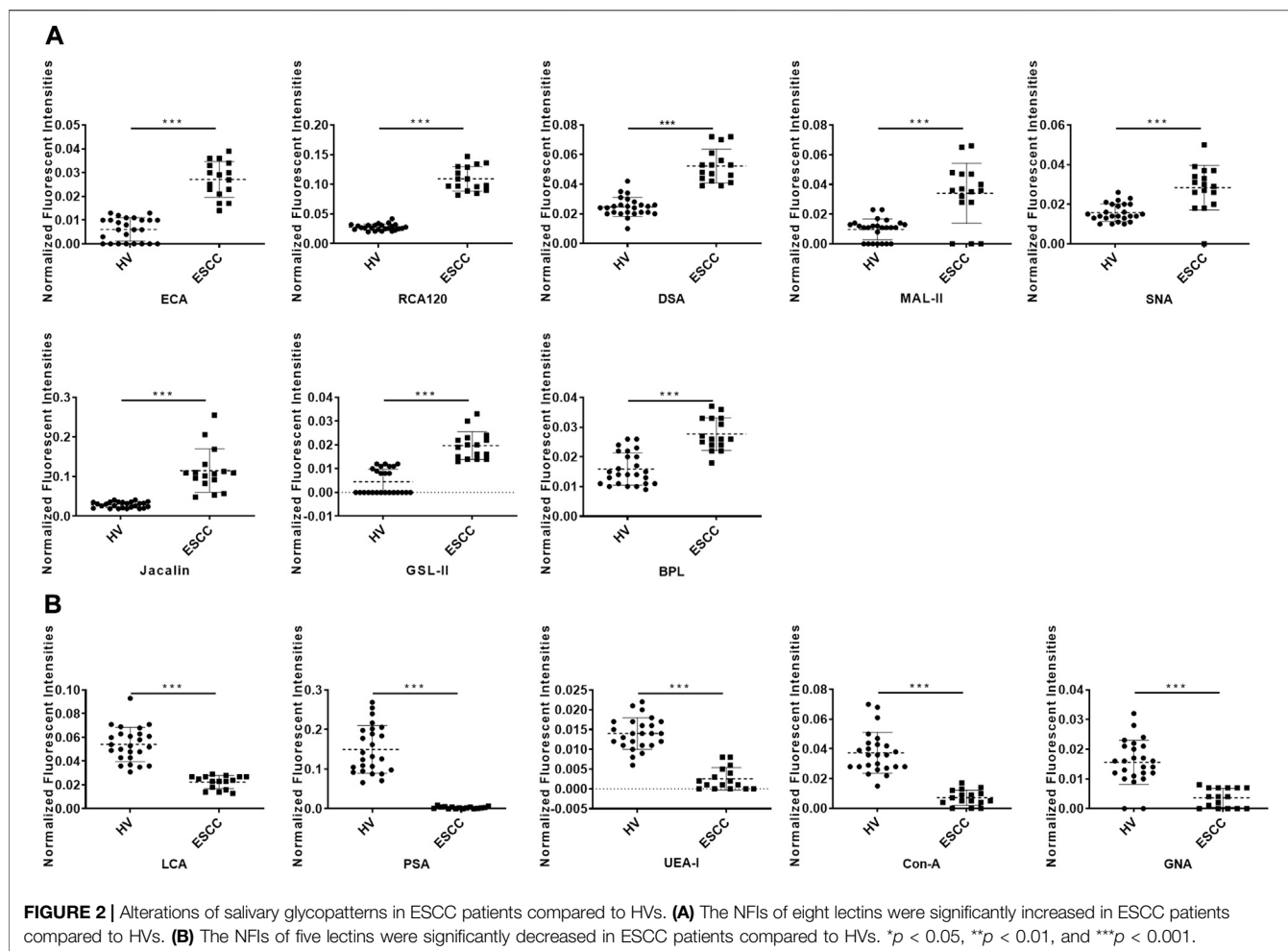
150 mM Na$^+$, pH 7.4), respectively, and reacted with DSA solution (DSA dissolve in coupling buffer) according to the protocol (Zhang et al., 2019; Yang et al., 2020b). Then, conjugates were washed with a coupling buffer to remove the unbound lectins.

Selective Isolation of Glycoprotein Fractions from Saliva by DSA-Magnetic Particle Conjugates

The GlcNAc and Galβ1-4GlcNAc-containing glycoproteins were isolated from ESCC patients and HVs using DSA-magnetic particle conjugates as described previously (Zhang et al., 2019; Yang et al., 2020b). Briefly, the pooled salivary protein was diluted with the binding buffer (100 Mm Tris-HCl; 150 mM NaCl; 1 mM CaCl$_2$, MgCl$_2$, and MnCl$_2$; pH 7.4) and incubated with the DSA conjugates (room temperature, 3 h). After 3 h gentle shaking, the conjugates were washed using washing buffer (0.1% Tween-20 in binding buffer, pH 7.2) to remove unbound proteins, and the glycoproteins bound to the conjugates were eluted with an eluting buffer (8 M urea, 40 mM NH$_4$HCO$_3$).

Isolation and Purification of N-Linked Glycans

The isolation and purification of N-glycans were performed based on previously described methods (Qin et al., 2017; Yang et al., 2020b). The glycoproteins were concentrated and desalted by Amicon Ultra-0.5 3 KDa ultrafiltration units (Millipore, United States) and then denatured by the addition of 8 M urea, 10 mM DTT, and 10 mM IAM. The denatured glycoproteins were exchanged to 40 mM NH$_4$HCO$_3$ buffer and incubated with trypsin (37°C, overnight). The trypsin in the mixture was inactivated by heating (80°C, 5 min) and then the PNGase F (New England Biolabs, Beverly, MA) was added to release the N-glycans (37°C, overnight). Subsequently, the digest was subjected to HyperSep Hypercarb SPE cartridges (25 mg, 1 mL; Thermo Scientific) to remove peptides. The purified N-glycans were collected and lyophilized.

Characterization of N-Glycans by MALDI-TOF/TOF-MS

The N-glycans were characterized by matrix-assisted laser desorption ionization time-of-flight/time-of-flight mass

TABLE 2 | Altered glycopatterns of salivary glycoproteins between HVs and ESCC patients.

Monosaccharide specificity	Lectin	Preferred glycan structure (terminal epitope)	ESCC
Gal	ECA	Galβ1-4GlcNAc (type II), Galβ1-3GlcNAc (type I)	⬆
	RCA120	Galβ1-4GlcNAc (type II), Galβ1-3GlcNAc (type I)	⬆
	BPL	Galα1-3GalNAc, GalNAc	⬆
	Jacalin	Galβ1-3GalNAcα-Ser/Thr(T), sialyl-T(ST)	⬆
Sialic acid	MAL-II	Siaα2-3Gal	⬆
	SNA	Siaα2-6Gal	⬆
GlcNAc	DSA	GlcNAc, Galβ1-4GlcNAc	⬆
	GSL-II	GlcNAc, agalactosylated tri/tetra antennary glycans	⬆
Mannose	ConA	High-mannose, Manα1-6(Manα1-3)Man	⬇
	GNA	High-mannose, Manα1-3Man	⬇
	LCA	Fucα1-6GlcNAc, high-mannose	⬇
	PSA	Fucα1-6GlcNAc, high-mannose	⬇
Fucose	LCA	Fucα-1,6GlcNAc, high-mannose	⬇
	PSA	Fucα-1,6GlcNAc, high-mannose	⬇
	UEA-I	Fucα1-2Galβ1-4Glc(NAc)	⬇

spectroscopy (MALDI-TOF/TOF-MS, UltrafleXtreme, Bruker Daltonics; Bremen, Germany) as described previously (Qin et al., 2017; Zhang et al., 2019; Yang et al., 2020b). Glycans were resuspended and spotted onto an MTP AnchorChip sample target. Then, 2 μL of 10 mg/mL 2,5-dihydroxybenzoic acid (DHB) with 1 mM NaCl in 50% (v/v) methanol solution was spotted to recrystallize the N-glycans and vacuum dried for

analysis. Peptide calibration standards (250 calibration points; Bruker) were used as mass calibration. Positive ion reflection mode was performed, and a mass range of 1,000–4,000 Da was analyzed. Representative MS spectra of N-glycan mass peaks with signal-to-noise ratio above three were generated and annotated using FlexAnalysis and GlycoWorkbench software.

RESULTS

Overall Salivary Protein Glycopatterns in Esophageal Squamous Cell Carcinoma Patients

A total of 16 ESCC and 25 HV samples were detected by lectin microarrays independently to investigate the altered salivary protein glycopatterns associated with ESCC patients. The layout of the lectin microarrays and typical representative images were listed in **Figure 1A,B**. HCA and PCA were executed to evaluate the glycan expression profiles of ESCC and HV cases and to provide a graphical representation of relationships among the subjects and diseases. The normalized fluorescent intensities (NFIs) for each lectin were distributed in the heat map by unsupervised clustering method to achieve the hierarchical relationship of the samples based on similarities in their glycan expression pattern. As shown in **Figure 1C**, 16 ESCC cases were classified into one category, and 25 HV cases were classified into another class, indicating that the salivary glycopatterns identified by these lectins were different between HVs and ESCC patients. A similar result was obtained by PCA as well; the subjects clustered separately by

principal components 1 and 2 to form HV and ESCC clusters with different colors and symbols in **Figure 1D**, representative of the salivary glycopatterns differences for the two groups in a manner.

Alterations of Salivary Protein Glycopatterns in Esophageal Squamous Cell Carcinoma Patients

To further investigate the alteration of salivary protein glycopatterns in ESCC patients, the Mann–Whitney test was used to compare the variance between ESCC and HV groups. In total, 13 lectins exhibited significantly altered NFIs in ESCC cases. As shown in **Figure 2** and **Table 2**, the Galβ1-3/4GlcNAc binder LEL and MAL-I, Galα1-3GalNAc binder BPL, T antigen and sialyl-T antigen binder Jacalin, Siaα2-3/6Gal binder MAL-II/ SNA, GlcNAc and Galβ1-4GlcNAc binder DSA, and GSL-II showed increased NFIs in ESCC salivary glycoproteins against HV cases. On the contrary, the high-mannose binder of ConA and GNA, Fucα-1,6GlcNAc binder LCA and PSA, and Fucα1-2Galβ1-4Glc (NAc) binder UEA-I showed decreased NFIs in ESCC salivary glycoproteins compared with HV cases. As a result, ESCC patients showed higher levels of GalNAc and Gal

FIGURE 3 | Validation of the differential expressions of the glycopatterns in pooled saliva from HVs and ESCC patients. **(A)** Silver nitrate staining of the salivary proteins from HV and ESCC groups. **(B)** The binding pattern of glycoproteins from ESCC patients and HVs using Cy five labeled lectins of DSA, ECA, LCA, PSA, and UEA-I. The molecular weights of band 1(b1), b2, b3, and b4 were approximately 90, 55, 60, and 25 kDa. **(C)** The gray value of each apparent difference bands was measured by ImageJ software.

FIGURE 4 | MALDI-TOF/TOF-MS spectra of the purified N-glycans from the DSA-enriched salivary glycoproteins of HVs **(A)** and ESCC patients **(B)**, respectively. Detailed glycan structures were analyzed using the GlycoWorkbench software. Proposed structures and their m/z values were shown for each peak. Blue square, GlcNAc; green circle, Man; yellow circle, Gal; yellow square, GalNAc; purple diamond, NeuAc; red triangle, Fuc.

expression profile, sialic acid expression profile, GlcNAc expression profile, and lower levels of mannose expression profile and fucose expression profile.

Validation of Different Glycopatterns Between ESCC Patients and HV Cases
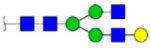
SDS-PAGE and lectin blotting analysis were performed with silver staining and Cy5-labeled lectin staining to confirm the different abundances of glycopatterns in pooled saliva from ESCC and HV cases. The results of SDS-PAGE demonstrated that the distribution and abundance of salivary protein bands were similar in ESCC and HV cases (**Figure 3A**). The result of the lectin blotting analysis showed that the obviously different bands range from 30 to 100 kDa (**Figure 3B**). The GlcNAc and Galβ1-4GlcNAc binder DSA showed a distinctly increased binding to four apparent bands with molecular weights of approximately 90 kDa (b1), 55 kDa (b2), 60 kDa (b3), and 25 kDa (b4) in the ESCC patients compared with HVs, and Galβ-1,3/4GlcNAc binder ECA staining showed stronger binding intensity to two apparent bands (b3 and b4) in the ESCC patients. On the

contrary, Fucα-1,6GlcNAc binder LCA and PSA and Fucα1-2Galβ1-4Glc (NAc) binder UEA-I showed weaker binding to b1, b2, b3, or b4 in ESCC than in HV (**Figure 3B,C**). The relative binding intensity of these randomly selected lectins to pooled saliva samples was almost coincident with the results from the lectin microarrays.

The N-Linked Glycan Profiles of the DSA-Isolated Salivary Glycoprotein
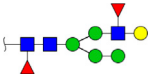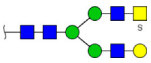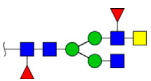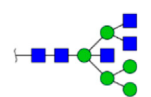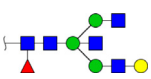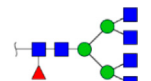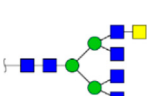
To obtain the GlcNAc, Galβ1-4GlcNAc N-glycan structures of glycoproteins in saliva from ESCC and HV cases. The glycoproteins were isolated using the DSA-magnetic particle conjugates, then N- glycans were released by PNGase F, purified by HyperSep Hypercarb SPE cartridges, and identified by MALDI-TOF/TOF-MS, respectively. A total of 48 and 56 N-glycan peaks from the pooled saliva samples of ESCC and HV were identified and annotated with proposed structures in **Figure 4A,B**. Of these, there were 44 and 52 GlcNAc or Galβ1-4GlcNAc containing N-glycans that could be recognized by DSA to be identified in HV and ESCC, respectively, and their proposed structures were listed in

TABLE 3 | The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
1	1501.529	—	1501.511	[M + Na]+		—	1.10
2	1565.491	1565.472	1565.474	[M + Na]+		5.30	5.47
3	1593.565	1593.532	1593.545	[M + Na]+		2.26	1.90
4	1606.560	1606.338	1606.341	[M + Na]+		0.91	1.15
5	1622.555	1622.528	1622.533	[M + Na]+		1.19	2.45
6	1625.605	1625.553	1625.563	[M + H]+		1.65	7.56
7	1663.581	1663.496	—	[M + Na]+		0.59	—
8	1688.613	1688.424	1688.425	[M + Na]+		0.59	0.70
9	1704.608	1704.593	1704.593	[M + Na]+		1.09	1.10
10	1730.636	1730.579	1730.606	[M + H]+		5.99	7.29
11	1746.631	1746.587	1746.592	[M + H]+		0.92	1.72

(Continued on following page)

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
12	1752.618	—	1752.608	[M + Na]+		—	0.94
13	1784.565	—	1784.524	[M + Na]+		—	1.35
14	1834.671	1834.528	1834.575	[M + Na]+		0.73	0.71
15	1844.69	1844.591	1844.593	[M + H]+		3.96	4.45
	1866.661	1866.558	1866.574	[M + Na]+		2.03	0.62
16	1850.666	—	1850.642	[M + Na]+		—	0.94
17	1891.692	1891.683	1891.660	[M + Na]+		0.66	0.76
18	1926.732	1926.675	—	[M + H]+		0.66	\
19	1946.661	—	1946.668	[M + Na]+		—	2.28
20	1955.697	—	1955.679	[M + Na]+		—	1.25

(Continued on following page)

16

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
21	1987.687	1987.535	1987.558	[M + Na]+		0.69	0.90
22	2006.732	2006.574	2006.546	[M + H]+		1.60	4.10
23	2015.768	2015.721	2015.725	[M + H]+		2.28	2.47
24	2018.639	—	2018.568	[M + Na]+		—	0.91
25	2028.714	—	2028.582	[M + Na]+		—	0.46
26	2084.740	2084.620	2084.533	[M + Na]+		2.15	0.63
27	2094.772	2094.581	—	[M + Na]+		0.56	—
28	2100.735	—	2100.549	[M + Na]+		—	0.51
29	2108.713	2108.626	—	[M + Na]+		0.79	—
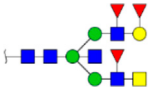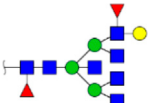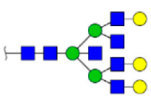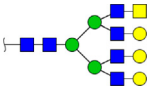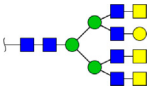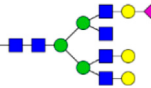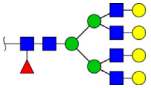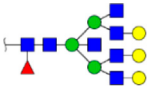
(Continued on following page)

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
30	2129.811	2129.695	2129.742	[M + H]+		1.90	1.52
31	2158.777	—	2158.768	[M + Na]+		—	0.42
32	2174.772	2174.600	—	[M + Na]+		0.53	—
33	2193.816	2193.721	2193.798	[M + H]+		0.88	0.45
34	2240.830	—	2240.745	[M + Na]+		—	0.38
35	2281.832	2281.683	2281.705	[M + H]+		0.99	0.37
36	2297.851	2297.918	2297.956	[M + Na]+		1.05	1.08
37	2344.841	—	2344.843	[M + Na]+		—	3.19
38	2377.851	2377.667	—	[M + Na]+		0.32	—
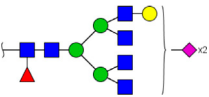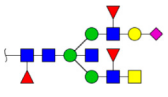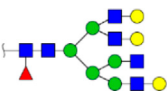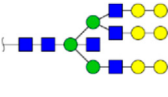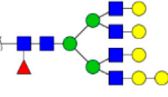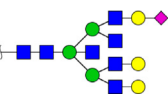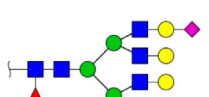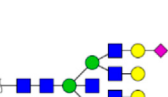
(Continued on following page)

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
39	2345.861	—	2345.698	[M + Na]+		—	3.19
40	2380.901	—	2380.872	[M + H]+		—	1.06
41	2412.890	2412.867	2412.940	[M + H]+		0.88	0.44
42	2434.872	—	2434.812	[M + Na]+		—	0.43
43	2516.925	—	2516.893	[M + Na]+		—	0.36
44	2522.888	2522.924	2522.781	[M + Na]+		7.23	2.45
45	2539.904	2539.764	—	[M + Na]+		0.31	—
46	2580.930	2580.804	2580.830	[M + Na]+		0.53	0.26

(Continued on following page)

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
47	2613.954	—	2613.930	[M + H]+		—	0.47
48	2636.957	—	2637.031	[M + Na]+		—	0.43
49	2679.975	2679.794	2679.795	[M + H]+		5.52	1.90
50	2695.670	2695.824	—	[M + H]+		0.38	—
51	2701.957	—	2702.092	[M + Na]+		—	0.30
52	2725.968	2726.128	2726.128	[M + Na]+		0.20	0.19
53	2830.999	—	2830.999	[M + Na]+		—	1.91
54	2866.039	2866.208	—	[M + H]+		0.26	—

(Continued on following page)

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
55	2919.015	2918.965	2918.981	[M + Na]⁺		0.66	0.25
56	2938.060	2938.150	2938.126	[M + H]⁺		0.37	0.24
57	2997.122	2997.246	—	[M + H]⁺		0.32	—
58	3053.123	3053.293	3053.247	[M + H]⁺		2.18	2.73
	3075.105	—	3075.261	[M + Na]⁺		—	0.42
59	3181.157	3181.005	—	[M + Na]⁺		0.28	—
60	3213.147	3213.112	3213.305	[M + Na]⁺		0.21	0.24
61	3396.175	3396.318	—	[M + Na]⁺		0.24	—

(Continued on following page)

TABLE 3 | (*Continued*) The GlcNAc, Galβ1-4GlcNAc-containing N-glycan peaks of DSA-isolated salivary glycoproteins in the present study.

NO	Calculated m/z	Experimental mass m/z		Charge	Proposed structure[a]	% N-glycans[b]	
		HV	ESCC			HV	ESCC
62	3409.244	3409.435	—	[M + H]+		0.21	—
63	3472.252	—	3472.428	[M + Na]+		—	0.25
64	3491.297	3491.445	3491.345	[M + H]+		0.93	0.32
65	3650.300	3650.486	3491.452	[M + Na]+		0.22	0.07
	The N-glycan with GlcNAc, Galβ1-4GlcNAc moieties			Number		44	52
				Sum		63.20	79.04

[a]Monosaccharides are represented according to MS tools from the GlycoWorkbench software. GlcNAc, blue square; Man, green circle; Gal, yellow circle; Fuc, red triangle; NeuAc, purple diamond.

[b]\ not detected in the sample.
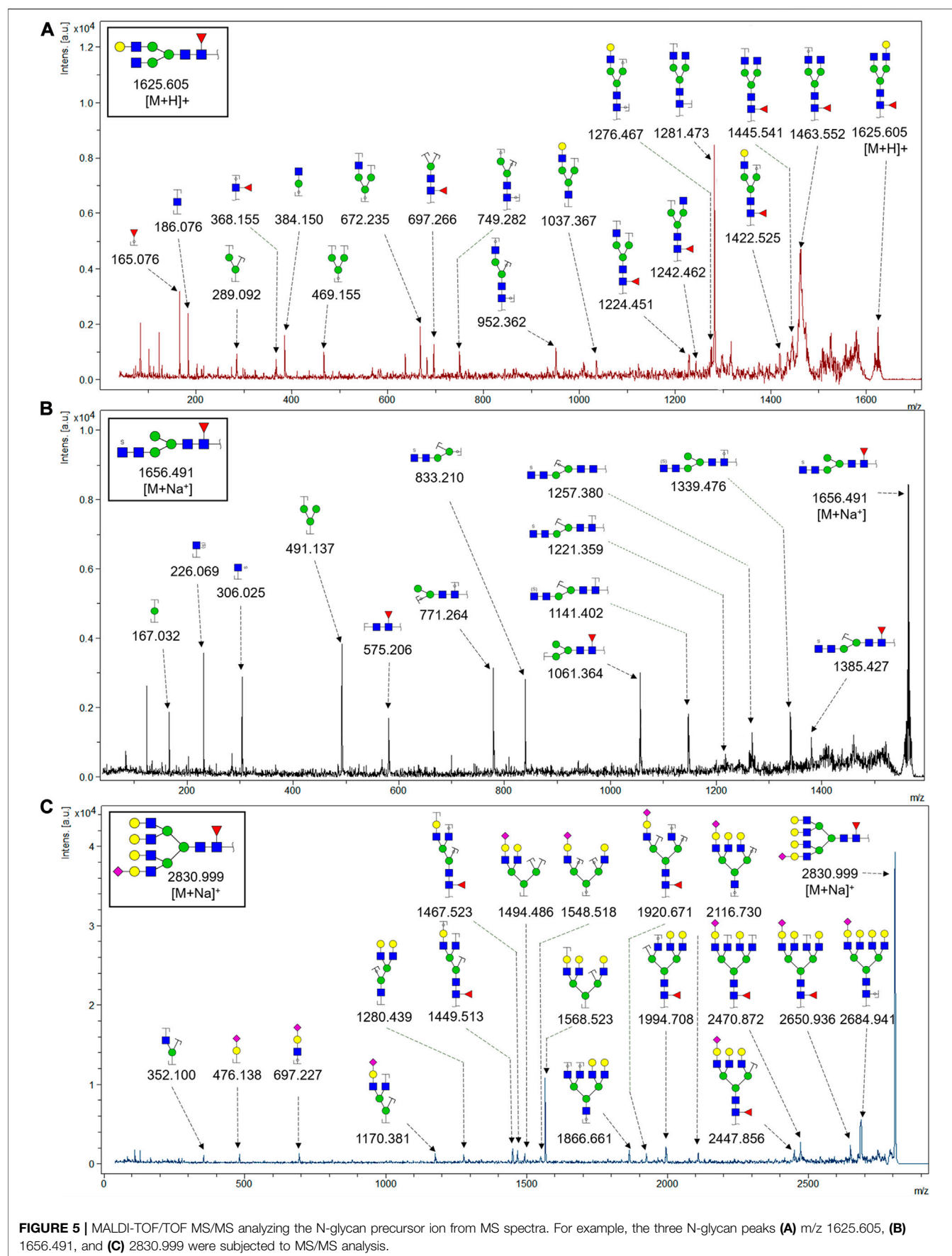
FIGURE 5 | MALDI-TOF/TOF MS/MS analyzing the N-glycan precursor ion from MS spectra. For example, the three N-glycan peaks **(A)** m/z 1625.605, **(B)** 1656.491, and **(C)** 2830.999 were subjected to MS/MS analysis.
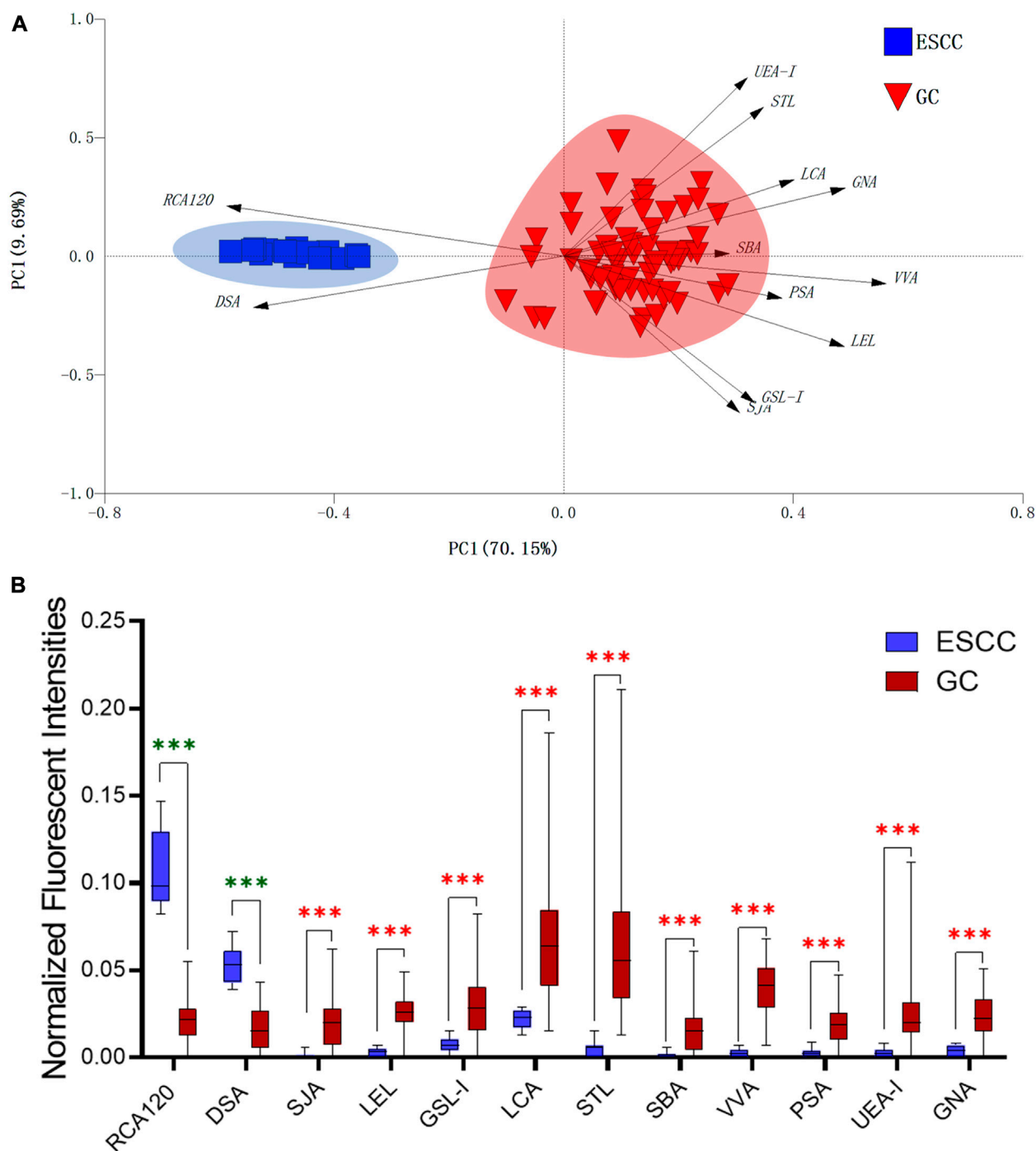
FIGURE 6 | The different salivary glycopatterns in ESCC and GC patients. **(A)** The principal component analysis of the salivary glycopatterns from ESCC and GC patients is based on the NFIs of 37 lectins. ESCC and GC cases were indicated by blue squares and red triangles. **(B)** The alterations of salivary glycopatterns in ESCC patients compared to GC patients. The NFIs of two lectins (RCA120 and DSA) increased and 10 lectins decreased in ESCC patients compared to GC patients.

Table 3. It was noticeable that there was an overlap of 31 N-glycans (e.g., m/z 1565.491, 1593.565, and 1625.605) presented in both HV and ESCC cases, while 13 N-glycans (e.g., m/z 1622.555, 1926.732, and 2084.740) were observed only in HV samples and 21 N-glycans (e.g., m/z 1501.529, 1792.624, and 2018.639) detected only in ESCC samples. The increased tendency of GlcNAc or Galβ1-4GlcNAc-containing N-glycans in ESCC samples compared with HVs was presented in the numbers and relative abundance levels. The proportion of the N-glycans with the GlcNAc or Galβ1-4GlcNAc moieties was increased in ESCC (79.04%) compared with HV (63.20%), which was consistent with the results of lectin microarrays. The MS/MS analysis was further performed to determine the exact glycan structures. For example, the MS/MS spectra of the precursor ions m/z 1625.605, 1656.491, and 2830.999 were illustrated in **Figure 5**.

DISCUSSION

Despite the continuous improvement in medical technology and comprehensive treatment, the prognosis of ESCC patients remains poor (Yang et al., 2020a; Wei et al., 2020). It is reported that the overall survival rate of ESCC is as high as 20–30%, and there is still much room for improvement because the 5-year survival rate of ESCC patients could rapidly increase to 80–90% if they were detected in the early stage and received the timely intervention (Fitzmaurice et al., 2017; Hoshino et al., 2020). Unfortunately, since nearly half of early-stage ESCC patients were unlikely to show clinical symptoms, coupled with the lack of reliable noninvasive screening methods, more than 50% of ESCC patients were diagnosed at the advanced stages (Ohashi et al., 2015; Kojima and Doi, 2017). Therefore, to improve the prognosis of ESCC, it is still an urgent demand for the discovery of a novel noninvasive biomarker that can be detected even earlier.

Decades of research have demonstrated that aberrant protein glycosylation often occurs in the development of tumors, and it has been shown that specific tumor-associated glycans are expressed in the precursor lesions of different types of cancer, which makes them potentially powerful early diagnosis markers (Oliveira-Ferrer et al., 2017; Liu et al., 2018; Okumura et al., 2020; Schedin-Weiss et al., 2020; Zhang et al., 2020). Glycomic and glycoproteomic analysis of blood samples have shown that glycans or glycan profiles could be used as candidate biomarkers to distinguish EC from controls and potential predictors of disease progression (Yehia et al., 2009; Hammoud et al., 2010; Mohanty et al., 2012; Song et al., 2014). But the comprehensive information of salivary glycopatterns from ESCC and the possibility of salivary glycopatterns acting as potential biomarkers were not explicit yet.

In this study, the integrated glycomics methods were used to investigate the differences of salivary glycopatterns between ESCC and HV cases. The results of lectin microarrays showed that 13 lectins (e.g., ECA, RCA120, and DSA) revealed significant alterations of the salivary glycopatterns between ESCC and HV cases. As a result, ESCC patients showed higher levels of GalNAc and Gal expression profile, sialic acid expression profile, GlcNAc expression profile, and lower levels of mannose expression profile and fucose expression profile. The lectins of DSA, ECA, LCA, PSA, and UEA-1 were randomly selected to confirm the differentially expressed sugar patterns. The lectin of DSA has an affinity to both GlcNAc and Galβ1-4GlcNAc, which can be used as an effective tool for the analysis of complex-type N-glycans. We used the DSA-magnetic particle conjugates to isolate the glycoproteins. And the glycan profiles of DSA-isolated glycoproteins were analyzed by MALDI-TOF/TOF-MS after liberation and purification. The results indicated that the proportion of the GlcNAc or Galβ1-4GlcNAc-containing *N*-glycans was increased in ESCC (79.04%) compared with HV (63.20%), which was consistent with the results of lectin microarrays. Compared with our previously published data of the other important gastrointestinal cancer, we found that the glycopatterns of ESCC and GC are obviously different (**Figure 6A**). The NFIs of RCA120 and DSA were higher in the saliva of ESCC patients, and the NFIs of SJA, LEL, GSL-I, LCA, STL, SBA, VVA, PSA, UEA-I, and GNA were higher in GC patients (**Figure 6B**). These results indicate that different tumors exhibit the different cancer-associated glycopatterns in saliva glycoproteins, but the detailed glycol-codes of cancers still need further characterization.

Our analyses have some limitations. First, it is a pilot study, the samples recruited are not enough, and an independent validation group is absent in this study. Second, the derivatization of sialic acids, such as permethylation and ethyl esterification, is needed for more accurate measurements of the released N-glycan profile. But our study provided comprehensive information of saliva glycopatterns from ESCC and the resolution of ESCC-associated glycopatterns that may contribute to understanding the complex physiological changes of ESCC patients. And the altered salivary glycopatterns such as GlcNAc or Galβ1-4GlcNAc-containing N-glycans recognized by DSA might be served as potential biomarkers for the diagnosis of ESCC patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The collection and use of human whole saliva for research in this study were approved by the Ethical Committee of Northwest University (Xi'an, China) and the Second Affiliated Hospital of Zhengzhou University (Zhengzhou, China). Written informed consent was received from participants. And this study was

conducted in accordance with the ethical guidelines of the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

ZL, XG, HY, and JS conceived and designed experiments; JM and JW contributed to collecting the saliva samples and clinical information. JS, YW, KZ, and XR performed the experiments; all authors participated in literature research and data classification; JS and XR wrote the manuscript; ZL and XG reviewed and edited the manuscript before submission. All authors have read and approved the manuscript for publication.

REFERENCES

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Chao, Q., Ding, Y., Chen, Z. H., Xiang, M. H., Wang, N., and Gao, X. D. (2020). Recent progress in chemo-enzymatic methods for the synthesis of N-glycans. *Front. Chem.* 8, 513. doi:10.3389/fchem.2020.00513

Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., et al. (2016). Cancer statistics in China, 2015. *CA Cancer J. Clin.* 66 (2), 115–132. doi:10.3322/caac.21338

de Haas, P., Hendriks, W. J. A. J., Lefeber, D. J., and Cambi, A. (2020). Biological and technical challenges in unraveling the role of N-glycans in immune receptor regulation. *Front. Chem.* 8, 55. doi:10.3389/fchem.2020.00055

Fang, L., Liu, Q., He, P., Wang, X., Wang, Y., Wei, M., et al. (2017). Alteration of salivary glycopatterns in oral lichen planus. *Biomark.* 23 (2), 1–21. doi:10.1080/1354750X.2017.1405284

Fitzmaurice, C., Fitzmaurice, C., Allen, C., Barber, R. M., Barregard, L., Bhutta, Z. A., et al. (2017). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990–2015: a systematic analysis for the global burden of disease study. *JAMA Oncol.* 3 (4), 524–548. doi:10.1001/jamaoncol.2016.5688

Hammoud, Z. T., Mechref, Y., Hussein, A., Bekesova, S., Zhang, M., Kesler, K. A., et al. (2010). Comparative glycomic profiling in esophageal adenocarcinoma. *J. Thorac. Cardiovasc. Surg.* 139 (5), 1216–1223. doi:10.1016/j.jtcvs.2009.12.045

Hirabayashi, J., Yamada, M., Kuno, A., and Tateno, H. (2013). Lectin microarrays: concept, principle and applications. *Chem. Soc. Rev.* 42 (10), 4443–4458. doi:10.1039/c3cs35419a

Hoshino, I., Ishige, F., Iwatate, Y., Gunji, H., Shiratori, F., Kuwayama, N., et al. (2020). Usefulness of serum miR-1246/miR-106b ratio in patients with esophageal squamous cell carcinoma. *Oncol. Lett.* 20 (6), 350. doi:10.3892/ol.2020.12213

Kojima, T., and Doi, T. (2017). Immunotherapy for esophageal squamous cell carcinoma. *Curr. Oncol. Rep.* 19 (5), 33. doi:10.1007/s11912-017-0590-9

Liu, X., Yu, H., Qiao, Y., Yang, J., Shu, J., Zhang, J., et al. (2018). Salivary glycopatterns as potential biomarkers for screening of early-stage breast cancer. *Ebiomedicine* 28, 70–79. doi:10.1016/j.ebiom.2018.01.026

Malhotra, G. K., Yanala, U., Ravipati, A., Follet, M., Vijayakumar, M., and Are, C. (2017). Global trends in esophageal cancer. *J. Surg. Oncol.* 115 (5), 564–579. doi:10.1002/jso.24592

Mohanty, S., Tsiouris, A., and Hammoud, Z. (2012). Glycomic expression in esophageal disease. *Metabolites* 2 (4), 1004–1011. doi:10.3390/metabo2041004

Nunes, L. A., Mussavira, S., and Bindhu, O. S. (2015). Clinical and diagnostic utility of saliva as a non-invasive diagnostic fluid: a systematic review. *Biochem. Med.* 25 (2), 177–192. doi:10.11613/bm.2015.018

Ohashi, S., Miyamoto, S., Kikuchi, O., Goto, T., Amanuma, Y., and Muto, M. (2015). Recent advances from basic and clinical studies of esophageal squamous cell carcinoma. *Gastroenterology* 149 (7), 1700–1715. doi:10.1053/j.gastro.2015.08.054

Okumura, M., Yamanoi, K., Uehara, T., and Nakayama, J. (2020). Decreased alpha-1,4-linked N-acetylglucosamine glycosylation in biliary tract cancer progression from biliary intraepithelial neoplasia to invasive adenocarcinoma. *Cancer Sci.* 111, 4629–4635. doi:10.1111/cas.14677

Oliveira-Ferrer, L., Legler, K., and Milde-Langosch, K. (2017). Role of protein glycosylation in cancer metastasis. *Semin. Cancer Biol.* 44, 141–152. doi:10.1016/j.semcancer.2017.03.002

Pinho, S. S., and Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* 15 (9), 540–555. doi:10.1038/nrc3982

Qin, Y., Zhong, Y., Ma, T., Zhang, J., Yang, G., Guan, F., et al. (2017). A pilot study of salivary N-glycome in HBV-induced chronic hepatitis, cirrhosis, and hepatocellular carcinoma. *Glycoconj. J.* 34 (4), 523–535. doi:10.1007/s10719-017-9768-5

Schedin-Weiss, S., Gaunitz, S., Sui, P., Chen, Q., Haslam, S. M., Blennow, K., et al. (2020). Glycan biomarkers for Alzheimer disease correlate with T-tau and P-tau in cerebrospinal fluid in subjective cognitive impairment. *FEBS J.* 287 (15), 3221–3234. doi:10.1111/febs.15197

Wang, S., Fan, Z., Zhou, B., Wang, Y., Du, P., Tan, W., et al. (2017). Roles of glycoproteins in the diagnosis and differential diagnosis of chronic and latent keshan disease. *Molecules* 22 (5), 746, doi:10.3390/molecules22050746

Shu, J., Yu, H., Du, H., Zhang, J., Zhang, K., Li, X., et al. (2018). Identification of N- and O-linked glycans recognized by AAL in saliva of patients with atrophic gastritis and gastric cancer. *Cancer Biomark.* 22 (4), 669–681. doi:10.3233/CBM-171087

Shu, J., Yu, H., Li, X., Zhang, D., Liu, X., Du, H., et al. (2017). Salivary glycopatterns as potential biomarkers for diagnosis of gastric cancer. *Oncotarget* 8 (22), 35718–35727. doi:10.18632/oncotarget.16082

Song, E., Zhu, R., Hammoud, Z. T., and Mechref, Y. (2014). LC-MS/MS quantitation of esophagus disease blood serum glycoproteins by enrichment with hydrazide chemistry and lectin affinity chromatography. *J. Proteome Res.* 13 (11), 4808–4820. doi:10.1021/pr500570m

Verhelst, X., Dias, A. M., Colombel, J. F., Vermeire, S., Van Vlierberghe, H., Callewaert, N., et al. (2020). Protein glycosylation as a diagnostic and prognostic marker of chronic inflammatory gastrointestinal and liver diseases. *Gastroenterology* 158 (1), 95–110. doi:10.1053/j.gastro.2019.08.060

Wei, L., Wang, B., Hu, L., Xu, Y., Li, Z., Shen, Y., et al. (2020). MEX3A is upregulated in esophageal squamous cell carcinoma (ESCC) and promotes development and progression of ESCC through targeting CDK6. *Aging* 12, 21091–21113. doi:10.18632/aging.103196

Yang, H., Su, H., Hu, N., Wang, C., Wang, L., Giffen, C., et al. (2020). Integrated analysis of genome-wide miRNAs and targeted gene expression in esophageal squamous cell carcinoma (ESCC) and relation to prognosis. *BMC Cancer* 20 (1), 388. doi:10.1186/s12885-020-06901-6

Yang, J., Liu, X., Shu, J., Hou, Y., Chen, M., Yu, H., et al. (2020). Abnormal galactosylated-glycans recognized by bandeiraea simplicifolia lectin I in saliva of patients with breast cancer. *Glycoconj. J.* 37 (3), 373–394. doi:10.1007/s10719-020-09910-6

Yehia, M., Ahmed, H., Slavka, B., Vitara, P., Min, Z., Dobrolecki, Lacey. E., et al. (2009). Quantitative serum glycomics of esophageal adenocarcinoma and other

esophageal disease onsets. *J. Proteome Res.* 8 (6), 2656–2666. doi:10.1021/pr8008385

Yu, H., Shu, J., and Li, Z. (2020). Lectin microarrays for glycoproteomics: an overview of their use and potential. *Expert Rev. Proteomics* 17 (1), 27–39. doi:10.1080/14789450.2020.1720512

Yu, H., Wang, J., Tang, Z., Li, X., Yin, M., Zhang, F., et al. (2020). Integrated glycomics strategy for the evaluation of glycosylation alterations in salivary proteins associated with type 2 diabetes mellitus. *RSC Adv.* 10 (65), 39739–39752. doi:10.1039/D0RA05466F

Zhang, J., Zhong, Y., Zhang, P., Du, H., Shu, J., Liu, X., et al. (2019). Identification of abnormal fucosylated-glycans recognized by LTL in saliva of HBV-induced chronic hepatitis, cirrhosis, and hepatocellular carcinoma. *Glycobiology* 29 (3), 242–259. doi:10.1093/glycob/cwy108

Zhang, W., Yang, Z., Gao, X., and Wu, Q. (2020). Advances in the discovery of novel biomarkers for cancer: spotlight on protein N-glycosylation. *Biomarkers Med.* 14 (11), 1031–1045. doi:10.2217/bmm-2020-0185

Zhang, Y., Wang, X., Cui, D., and Zhu, J. (2016). Proteomic and N-glycoproteomic quantification reveal aberrant changes in the human saliva of oral ulcer patients. *Proteomics* 16 (24), 3173–3182. doi:10.1002/pmic.201600127

An Ultrafast *N*-Glycoproteome Analysis Method Using Thermoresponsive Magnetic Fluid-Immobilized Enzymes

Zhiya Fan[1], Tong Liu[1], Fei Zheng[1], Weijie Qin[1,2]* and Xiaohong Qian[1]

[1] *State Key Laboratory of Proteomics, National Center for Protein Sciences (Beijing), Beijing Proteome Research Center, Beijing, China,* [2] *College of Basic Medicine, Anhui Medical University, Hefei, China*

N-Glycosylation is one of the most common and important post-translational modification methods, and it plays a vital role in controlling many biological processes. Increasing discovery of abnormal alterations in *N*-linked glycans associated with many diseases leads to greater demands for rapid and efficient *N*-glycosylation profiling in large-scale clinical samples. In the workflow of global *N*-glycosylation analysis, enzymatic digestion is the main rate-limiting step, and it includes both protease digestion and peptide-*N*4–(*N*-acetyl-beta-glucosaminyl) asparagine amidase (PNGase) F deglycosylation. Prolonged incubation time is generally required because of the limited digestion efficiency of the conventional in-solution digestion method. Here, we propose novel thermoresponsive magnetic fluid (TMF)-immobilized enzymes (trypsin or PNGase F) for ultrafast and highly efficient proteome digestion and deglycosylation. Unlike other magnetic material-immobilized enzymes, TMF-immobilized enzymes display a unique temperature-triggered magnetic response behavior. At room temperature, a TMF-immobilized enzyme completely dissolves in an aqueous solution and forms a homogeneous system with a protein/peptide sample for efficient digestion but cannot be separated by magnetic force because of its excellent water dispersity. Above its lower critical solution temperature (LCST), thermoflocculation of a TMF-immobilized enzyme allows it to be easily recovered by increasing the temperature and magnetic force. Taking advantage of the unique homogeneous reaction of a TMF-immobilized enzyme, both protein digestion and glycopeptide deglycosylation can be finished within 3 min, and the whole sample processing time can be reduced by more than 20 times. The application of a TMF-immobilized enzyme in large-scale profiling of protein *N*-glycosylation in urine samples led to the successful identification of 2,197 *N*-glycopeptides and further demonstrated the potential of this strategy for fast and high-throughput analysis of *N*-glycoproteome in clinical samples.

Keywords: urine proteomics, protein glycosylation, immobilized enzyme, thermo-responsiveness, magnetic fluid

INTRODUCTION

Glycosylation is one of the most prominent post-translational modification methods for proteins (Stadlmann et al., 2017; Huang et al., 2019). As a major type, *N*-glycosylation has a wide range of functions that greatly amplifies the diversity of proteins (Hart and Copeland, 2010; Yang et al., 2020). From the general biological process, such as cell adhesion and signal transduction, to specific functions of proteins like folding and stability, the complexity imparted to a proteome by *N*-glycosylation is immense (Schjoldager et al., 2020). Moreover, the *N*-glycans biosynthesis process is very sensitive to the physiological and pathological states in cells (Mereiter et al., 2019; Dong et al., 2020), and glycoproteins are a main type of current therapeutic targets and disease biomarkers (Pan et al., 2020; Zhao et al., 2020). Fueled by the increasing discovery of disease-related *N*-glycans alterations, the interest in large-scale *N*-glycosylation profiling in clinical samples is proliferating. Urine, as a reflection of body changes, is considered as an ideal source for biomarker discovery (Wu and Gao, 2015). It can be obtained in a non-invasive manner with a relatively narrower protein dynamic range and much less interference by high-abundance proteins compared with that of plasma/serum (Zhao et al., 2017). Therefore, rapid and in-depth analysis of *N*-glycoproteome from a large cohort of urine samples is highly desirable for a clinical study.

In mass spectrometry-based shotgun *N*-glycoproteome analysis, enzymatic digestion of proteins and release of *N*-glycans are regarded as key steps, which are generally performed using a solution. However, because only a small amount of enzyme is used and digestion efficiency is limited in solution-based digestion, prolonged incubation time is needed, which highly limits sample throughput (Qin et al., 2012). To address these problems, various kinds of the immobilized enzyme have been developed with the advantages of reducing enzyme self-digestion, stabilizing enzymatic activity, and allowing higher enzyme concentration for shorter digestion time (Mateo et al., 2007). In contrast with conventional in-solution digestion, an immobilized enzyme for *in situ* digestion can be easily separated from the digestion system and reused. Various supports have been proven to be feasible for enzyme immobilization, such as nanomaterials (Sharifi et al., 2020), porous silicon matrices (Létant et al., 2004), porous polymer monoliths (Krenkova et al., 2009), sol-gel supports (Yuce-Dursun et al., 2016), membranes (Luo et al., 2014), magnetic beads (Zhao et al., 2015; Fauser et al., 2020), and graphene oxide (Yuan et al., 2017). Although the required digestion time is obviously decreased using these reagents, because of the insolubility of reported supporting materials, digestion is processed under solid–liquid heterogeneous conditions that may limit reaction efficiency. Therefore, a new immobilized enzyme with homogeneous digestion and heterogeneous separation characteristics has high demand.

In this study, we developed a new type of immobilized enzyme reagent based on thermoresponsive magnetic fluid (TMF) with a lower critical solution temperature (LCST) in an aqueous solution (**Supplementary Figure 1**). To the best of our knowledge, this is the first study to immobilize trypsin or PNGase F in TMF. TMF is composed of thermoresponsive polymer chains grafted from the surface of a small magnetic Fe_3O_4 core ($\sim$10 nm), *via in situ* growth using the surface-initiated atom transfer radical polymerization (SI-ATRP) technique. Thermoresponsive polymer chains are copolymerized using *N*-isopropylacrylamide (NIPAM) (Zhang et al., 2017) and undecylenic aldehyde (UnAl), which provides TMF with thermoresponsive property, water solubility, and reactivity with enzymes. Unlike other magnetic material-immobilized enzymes, TMF-immobilized enzymes exhibit a unique temperature-triggered magnetic response behavior. When the temperature is lower than LCST, the TMF-immobilized enzyme completely dissolves in an aqueous solution, and digestion is performed under homogeneous conditions, which therefore avoids interfacial mass transfer resistance in the two-phase heterogeneous system, and enhances enzyme–substrate interaction and reaction efficiency. After digestion and raising the temperature above LCST, the TMF-immobilized enzyme can be easily recovered by heat-triggered polymer thermoflocculation and magnetic separation (see **Supplementary Video 1**). Taking advantage of homogeneous reaction and high enzyme loading, either protein digestion or glycopeptide deglycosylation can be finished within 3 min, and the whole sample processing time can be reduced by more than 20 times (**Figure 1**) with digestion efficiency comparable to that of conventional in-solution overnight digestion. The successful application of TMF-immobilized enzymes in large-scale *N*-glycosylation identification using complex protein samples further demonstrated the potential of this strategy for high-throughput analysis of *N*-glycoproteome in clinical samples.

MATERIALS AND METHODS

Preparation of TMF-Immobilized Enzyme
Synthesis of Oleic Acid-Coated Magnetic Nanoparticles (OA-Coated MNPs)

Oleic acid (OA)-coated magnetic Fe_3O_4 nanoparticles were synthesized according to the previously reported method (Sun et al., 2006) with minor modifications. $FeSO_4 \cdot 7H_2O$ (1.18 g, 4.24 mmol) and $FeCl_3 \cdot 6H_2O$ (2.05 g, 7.58 mmol) were dissolved in 50 ml deionized water with vigorous stirring. The solution was heated at 60°C, and then 25 ml 25% (w/w) $NH_3 \cdot H_2O$ was added dropwise. The solution color changed from orange to black, leading to a black precipitate. Then, 0.5 ml OA was added dropwise into the dispersion slowly at 80°C for 1 h. The whole process was carried out in a nitrogen atmosphere. Next, magnetic nanoparticles were extracted from water into toluene. Water dispersion 50 and 50 ml toluene was mixed in a 250-ml extractor. After adding a small amount of NaCl, magnetic nanoparticles transferred into the toluene phase with good dispersity under the coating of OA. Finally, the toluene dispersion was evaporated to remove the solvent under reduced pressure and dried under vacuum. The content of OA-coated MNPs was redissolved into toluene to 10 mg/ml.
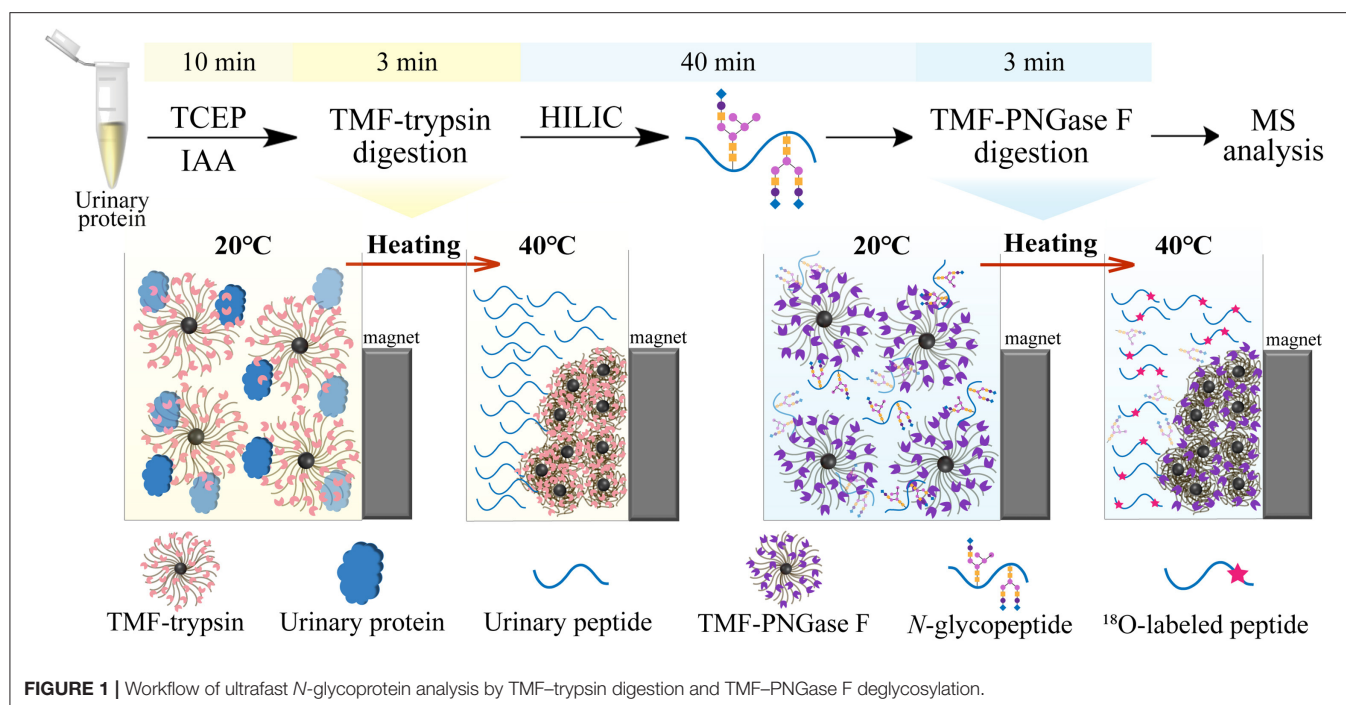
FIGURE 1 | Workflow of ultrafast *N*-glycoprotein analysis by TMF–trypsin digestion and TMF–PNGase F deglycosylation.

Synthesis of Initiator

As an initiator, 3-(2-bromoisobutyramido)propyl(triethoxy)–silane (BIBAPTES) was synthesized as follows: α-bromoisobutyryl bromide (0.1 ml, 0.8 mmol) was added dropwise to a cold solution of 3-aminopropyltriethoxysilane (ATEPS) (0.18 ml, 0.8 mmol) in dry toluene (10 ml) with triethylamine (TEA) (0.12 ml, 0.8 mmol) at 0°C. The mixture was magnetically stirred for 3 h at 0°C and further stirred for 10 h at room temperature to complete the reaction. The precipitate (triethylammonium bromide) was filtered off. The filtrate was evaporated under reduced pressure to remove unreacted TEA and dried under vacuum.

Synthesis of Initiator-Grafted Magnetic Nanoparticles

Initiator-grafted magnetic nanoparticles were synthesized by ligand exchange reaction: BIBAPTES (0.05 ml), TEA (1 ml), and OA-coated MNPs (50 mg) were added into 15 ml dry toluene in a nitrogen atmosphere. The mixture was stirred for 48 h at room temperature. Next, the mixture was centrifuged to remove salts (triethylammonium oleate). The supernatant was centrifuged at 20,000 *g* for 60 min, washed three times with toluene to remove un-grafted initiators and unreacted TEA, and dried under vacuum.

Synthesis of PNIPAM-b-PUnAl via Surface Initiated ATRP on Magnetic Nanoparticles

Initiator-grafted magnetic nanoparticles 60 mg were resuspended in 20 ml isopropanol/water (v/v: 3/1), and then *N*-isopropylacrylamide (NIPAM) (1 g, 8.85 mmol), purified Cu(I)Br (240 mg, 1.68 mmol), and Tris (2-dimethylaminoethyl)amine (Me₆TREN) (0.44 ml, 1.68 mmol) were added in a nitrogen atmosphere. The mixture was agitated at room temperature for

6–8 h, and then UnAl (1.4 ml, 7.08 mmol) was added and reacted for a further 12–18 h at room temperature. Finally, the magnetic fluid with thermoresponsive property and aldehyde groups was obtained after dialysis for 48 h.

Trypsin or PNGase F Immobilization on TMF and Measurement of Enzyme Loading Capacity

Trypsin (or PNGase F) was immobilized on the PNIPAM-*b*-PUnAl-grafted magnetic nanoparticles *via* Borch reduction between the aldehyde groups of the copolymer chains and the free amino groups of trypsin (or PNGase F). Typically, 8 mg modified magnetic nanoparticles were dispersed in 1 ml PBS (pH 7.4) containing 1 mg trypsin (or 1,500 UPNGase F) and 5 mg NaBH₃CN. Next, the mixture was agitated at 4°C for 12 h. After the reaction, the mixture was heated to 32°C to allow the magnetic nanoparticles to flocculate, and trypsin (or PNGase F)-attached magnetic nanoparticles were then collected with a magnet. The collected magnetic nanoparticles were washed three times by adding a 1 ml 50 mM NH₄HCO₃ aqueous solution and repeating dispersion and flocculation. The enzyme loading capability of TMF was determined by HPLC *via* calculation of the difference in the peak area of the bovine serum albumin (BSA) solution before and after immobilization, which is similar to that of trypsin or PNGase F.

TMF–Trypsin or Free Trypsin Digestion of Standard Protein and Urinary Proteins

This study was approved by the Institutional Review Board of the Tianjin Baodi Hospital, and all donors approved the use of their urine samples. Standard protein (BSA) was dissolved in 50 mM ammonium bicarbonate (ABC, pH = 8). Mid-stream samples of morning urine were collected and centrifuged at 12,000 g and

4°C for 30 min to remove cell debris. The urinary protein was precipitated by ice-cooled acetone and dissolved in a lysis buffer (8 M urea, 100 mM Tris-HCl, pH = 8) to a concentration of 1 mg/ml. Each sample was heated for 10 min at 95°C to denature. Then, TCEP (10 mM) reduction and CAA (40 mM) alkylation were performed. For TMF–trypsin digestion, 1 mL of the protein solution was mixed with TMF–trypsin and incubated for 3 min at room temperature. After digestion, the immobilized trypsin was retained by a heating-magnet process, and the supernatant was collected for mass spectrometry analysis. The TMF–trypsin materials can be reused after washing them three times with 50 mM ABC. For free trypsin digestion, first, the protein solution was diluted to reduce the urea concentration to 1 M. Then, free trypsin was introduced into the denatured protein solution at a substrate to enzyme ratio (w/w) of 50:1 and incubated at 37°C for 16 h. After digestion, 2 μl of formic acid was added to terminate the reaction, and the supernatant was collected for mass spectrometry analysis.

HILIC Enrichment of Urinary *N*-Glycopeptides

Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) materials (5 mg) were washed with a 200 μl binding buffer (80% ACN, 1% TFA) three times and incubated with 80 μg urinary peptides (TMF–trypsin digestion products) for 0.5 h. After that, a 600 μl binding buffer was used to remove non-specifically absorbed peptides. Finally, the *N*-glycopeptides were eluted using a 200 μl elution buffer (0.1% FA) and then vacuum-dried.

TMF–PNGase F or Free PNGase F Deglycosylation of Standard Glycoproteins and Urinary Glycopeptides

Standard glycoprotein RNase B (10 μg/μl) was dissolved in deionized water and heated at 95°C for 10 min to denature. Then, TCEP (10 mM) reduction and CAA (40 mM) alkylation were performed. Urinary *N*-glycopeptides were obtained according to *HILIC Enrichment of Urinary N-Glycopeptides* and dissolved in $H_2^{18}O$. For TMF–PNGase F deglycosylation, 4 μl of the protein solution was mixed with 1 ml TMF–PNGase F and incubated for 3 min at room temperature. The immobilized PNGase F was retained by a heating-magnet process, and the supernatant was collected for mass spectrometry analysis. The TMF–PNGase F materials can be reused after washing them three times with deionized water (or $H_2^{18}O$). For in-solution deglycosylation, PNGase F was introduced to the protein solution at a substrate to enzyme ratio (w/w) of 10:1 and incubated at 37°C for 16 h. No further pre-process was required for in-solution digestion before mass spectrometry analysis.

Mass Spectrometry Analysis
MALDI-TOF MS

For (glyco)protein analysis, equivalent volumes of the protein solution, aqueous 2% trifluoroacetic acid (TFA), and a 2,5-dihydroxyacetophenone (2, 5-DHAP) matrix were mixed by pipetting up and down until the liquid became cloudy. The

mixture (1 μl) was applied onto a matrix-assisted laser desorption ionization (MALDI) plate and air-dried. For tryptic peptides and glycopeptides analysis, equivalent volumes of the peptide solution and a 2,4-dihydroxybenzoic (DHB) matrix were mixed, deposited onto a MALDI plate (1 μl), and air-dried. A MALDI-time of flight (TOF) mass spectrometry (MS) instrument (Bruker, Bremen, Germany) performed in positive ion mode with a nitrogen pulsed laser (337 nm) was used for measurements.

LC-MS/MS

The resulting peptide mixture was analyzed using an Orbitrap Fusion Lumos Tribrid mass spectrometer coupled with an EASY-nLC 1,000 nano-LC system (ThermoFisher Scientific, Waltham, MA, USA). Peptide separation was performed on a 15-cm length reverse phase C_{18} column (150 nm id, 1.9 μm, 100 Å) using A and B buffers (buffer A:0.1% formic acid in water; buffer B:0.1% formic acid in acetonitrile) at a constant flow rate of 600 nl min^{-1}. The gradient was set as follows: 7–15% B for 7 min, 15–25% B for 37 min, 25–40% B for 20 min, and 40–100% B for 7 min. The dynamic exclusion duration of data-dependent MS2 acquisition (DDA) is 18 s. For MS1 scan, mass spectra were acquired in the positive-ion mode over the range of 300–1,400 m/z with a resolution of 120,000 and a maximum ion injection time of 50 ms. MS2 spectra were acquired with an automatic gain control target value of 5.e3 and a maximum injection time of 35 ms with higher-energy collision dissociation (HCD) with a normalized collision energy of 30%.

Data Processing

MALDI-TOF-MS spectra were analyzed using the FlexAnalysis software (version 3.4) to extract peaks and corresponding intensities. Then, Mascot search and sequence coverage calculation by allowing two missed cleavages were performed. LC-MS/MS raw data were analyzed using the MaxQuant software (version 1.6.17.0) to search against the UniPort Human database (updated on July 21, 2015). The main parameters were set as follows: (1) digestion mode was set as trypsin up to two missed cleavages allowed, (2) mass tolerances were 20 ppm and 4.5 ppm for the first search and main search, respectively, (3) fixed modification, carbamidomethyl (C), (4) variable modification, deamidation 18O (N), acetyl (protein N-term), oxidation (M), (5) false discovery rate (FDR) was set as ≤0.01 at the spectra, protein and modification levels, and (6) minimum and delta scores for the modified peptides were set as ≥40 and ≥6, respectively. For *N*-glycopeptides identification, a motif filter of NXT/S/C (where X cannot be P) in peptide sequence and a localization probability filter [deamidation 18O (N)] of ≥0.75 were applied.

Characterization of TMF Materials

Dynamic laser scattering (DLS) characterization was performed using a DynaPro NanoStar instrument (Wyatt Technology, Sta. Barbara, CA, USA). Fourier transform-infrared (FT-IR) spectra were obtained using a Tensor 27 FT-IR spectrometer (Bruker Corporation, Billerica, MA, USA). Vibrating sample magnetometry (VSM) was performed using a PPMS-9 instrument (Quantum Design, San Diego, CA, USA).

Thermogravimetric analysis (TGA) was performed using a SDT Q500 instrument (TA Instruments, New Castle, DE, USA).

RESULTS AND DISCUSSION

Preparation and Characterization of TMF

TMF was prepared *via* surface-initiated atom transfer radical polymerization (SI-ATRP) of the Fe_3O_4 nanoparticles (**Supplementary Figure 1**). First, the OA-coated Fe_3O_4 nanoparticles were synthesized according to the method, and then the initiators were immobilized on the surface of Fe_3O_4 nanoparticles through ligand exchange reaction. In SI-ATRP modification, NIPAM and UnAl were sequentially added into the reaction system to provide thermoresponsive property and aldenyde for enzyme conjugation. As a result, the polymer shell thickness, LCST, and the amounts of reactive groups can be well-defined by controlling the amounts and polymerization time of each monomer. After SI-ATRP modification, a multilayer enzyme reagent can be easily obtained by attaching enzymes on the side chains of the copolymer *via* the formation of covalent bonds with the aldehyde groups provided by poly-UnAl. More importantly, the poly-NIPAM part not only enables TMF well-dispersity in water to become magnetic fluid, which allows the digestion process under homogeneous conditions, but also offers a temperature-triggered method to separate the TMF-immobilized enzyme from the digestion medium. TMF was characterized by dynamic light scattering (DLS), FT-IR spectroscopy, VSM, and thermogravimetric analysis (TGA) to confirm the successful growth of PNIPAM-*b*-PUnAl copolymer chains on the surface of the magnetic nanoparticles and determine the content of the grafted copolymer chains.

Size and size distribution were characterized by DLS at 25°C. The size of (a) un-grafted magnetic nanoparticles was 13.77 nm, (b) that of Poly-NIPAM grafted magnetic nanoparticles was 27.99 nm, and (c) that of PNIPAM-*b*-PUnAl copolymer grafted magnetic nanoparticles was 51.12 nm. Their PDI were 0.282, 0.430, and 0.331, respectively (**Figure 2A**). The increase in size is caused by the grafting of poly-NIPAM and poly-UnAlon magnetic nanoparticles. FT-IR was performed on (a) un-grafted magnetic nanoparticles and (b) PNIPAM-*b*-PUnAl copolymer grafted magnetic nanoparticles (**Figure 2B**). The peaks at 592 cm^{-1} of both (a) and (b) are assigned to the Fe–O bonds of the Fe_3O_4 core. Two relatively strong absorption peaks at 1,550 and 1,680 cm^{-1} corresponding to the N–H and C=O bonds in the amide groups of (b) indicate successful grafting of Poly-NIPAM on magnetic nanoparticles. A relatively weak absorption peak at 1,720 cm^{-1} ascribed to the C=O bonds in the aldehyde group indicates successful growth of poly-UnAl on magnetic nanoparticles. A strong absorption peak at 2,920 cm^{-1} ascribed to the C–H bond in the methylene group is observed in (b), which indicates that the polymer continuously grows on the surface of magnetic nanoparticles. Next, magnetic hysteresis loops for the magnetic Fe_3O_4 nanoparticles before and after grafting with the poly-NIPAM-*b*-poly-UnAl copolymer were determined by VSM testing (**Figure 2C**). The saturated magnetization (Ms) of un-grafted magnetic nanoparticles is 63.5 emu/g and decreased to 37.8 emu/g after the SI-ATRP with

NIPAM and UnAl, because of the introduction of the un-magnetic copolymer. TGA analysis was carried out to further confirm the successful preparation of TMF as well as determine the content of surface-grafted polymer chains (**Figure 2D**). After thermo treatment over 700°C, the thermally stable Fe_3O_4 core remains in the residue and decomposable polymer chains contribute to weight loss. Unmodified magnetic nanoparticles (a) show <5% of total weight loss that might be attributed to the loss in water residue. In contrast, due to the decomposition of their polymer shell, modified magnetic nanoparticles showed noticeable weight loss. The weight loss is 21.7% for the magnetic nanoparticles with poly-NIPAM grafting (b) and increases to 32.85% after subsequent copolymerization with poly-UnAl (c), which indicates successful grafting of poly-NIPAM and poly-UnAl parts on the surface of the Fe_3O_4 core. The enzyme loading capability of TMF was determined by HPLC measurement (**Figure 2E**). The peak area of the BSA solution with 0.5 µg/µl (a) was 4,273 while the area decreased by 44% to 2,395 after attaching on 8-mg PNIPAM-*b*-PUnAl copolymer grafted magnetic nanoparticles; and, consequently, the concentration of the loading enzyme was 220 µg/mg.

Digestion Performance Evaluation of TMF Enzymes by Standard Protein

First, we used BSA as a standard protein to examine the performance of TMF-immobilized trypsin. Besides, in-solution free trypsin digestion (37°C, 16 h) was performed as a control. After digestion, TMF–trypsin was recovered by a heating-magnet process, and the supernatant was collected for MALDI-TOF MS analysis. A typical MALDI-TOF-MS spectrum of BSA digests by TMF–trypsin digestion showed 89% peptide coverage (**Figure 3A**). No obvious peaks exceeding 3,000 m/z could be detected, indicating that TMF–trypsin digestion was completed in such a short time. Moreover, no residual undigested BSA was observed, which demonstrated almost 100% digestion by immobilized trypsin (**Supplementary Figure 2**). Reducing the amount of BSA to 10 µg, however, can still obtain higher sequence coverage (87.67%) in TMF–trypsin system, and was relatively higher than that of free digestion (**Figure 3B**). Next, we used RNase B as a model *N*-glycoprotein to examine the performance of the TMF-immobilized PNGase F reagent. RNase B is a small glycoprotein with a molecular weight of ~15 kDa and containing a single *N*-glycosylation site at Asn34, which possesses five to nine mannose residues attached to the chitobiose core (Fu et al., 1994). RNase B was denatured at 95°C for 10 min and then digested with either TMF–PNGase F (room temperature, 3 min) or free PNGase F (37°C, 16 h). Five *N*-glycans [(Man)5–(Man)9] were identified in both conditions with similar signal intensity, indicating the deglycosylation efficiency of the immobilized PNGase F was as good as that of the free PNGase F digestion (**Figures 3C,D**), but with 320 a fold reduced digestion time. Next, the completeness of deglycosylation using TMF–PNGase F was examined by MALDI-TOF MS in the leaner model. Multiple peaks were detected within the range from 14,984.7 to 15,832.7 Da before digestion of RNase B because of the microheterogeneity of *N*-glycosylation; while only one peak

FIGURE 2 | (A) Size distribution of (a) un-grafted magnetic nanoparticles, (b) poly-NIPAM grafted magnetic nanoparticles, and (c) poly-NIPAM-*b*-poly-UnAl copolymer-grafted magnetic nanoparticles. **(B)** FT-IR spectra of (a) un-grafted magnetic nanoparticles and (b) poly-NIPAM-*b*-poly-UnAl copolymer-grafted magnetic nanoparticles. **(C)** Magnetization curves at 25°C for (a) un-grafted magnetic nanoparticles and (b) poly-NIPAM-*b*-poly-UnAl copolymer-grafted magnetic nanoparticles. **(D)** TGA curves of (a) un-grafted magnetic nanoparticles, (b) poly-NIPAM grafted magnetic nanoparticles, and (c) poly-NIPAM-*b*-poly-UnAl copolymer-grafted magnetic nanoparticles. **(E)** HPLC spectra of (a) 0.5 μg/μl BSA and (b) supernatant solution after the immobilization of BSA on poly-NIPAM-*b*-poly-UnAl copolymer-grafted magnetic nanoparticles.

at 13758.5 Da was detected after TMF–PNGase F digestion, indicating complete removal of the *N*-glycans by TMF–PNGase F (**Figure 3E**). Collectively, the TMF-immobilized enzymes (both trypsin and PNGase F) can efficiently digest (glyco)proteins using significantly reduced time compared to the conventional in-solution digestion system.

The Combinational Usage of TMF–Trypsin and TMF–PNGase F for Ultrafast Urine *N*-Glycoproteome Analysis

Long sample processing time and low throughput of conventional in-solution digestion is one of the major

limitations for the application of (glyco)proteomic technique in large-scale clinical studies. Using an immobilized enzyme is a promising solution. Unlike other reported immobilized enzymes, TMF-immobilized enzymes have the advantage of homogeneous-phase digestion and heterogeneous separation. Therefore, efficient digestion can be expected by avoiding high-mass transfer resistance in the two-phase digestion system using solid material-immobilized enzymes. Furthermore, TMF-immobilized enzymes can be easily separated from digestion products and recovered through heat-trigged flocculation and magnetic separation. We further evaluated digestion efficiency by sequential application of TMF–trypsin and TMF–PNGase F in urine glycoproteome analysis
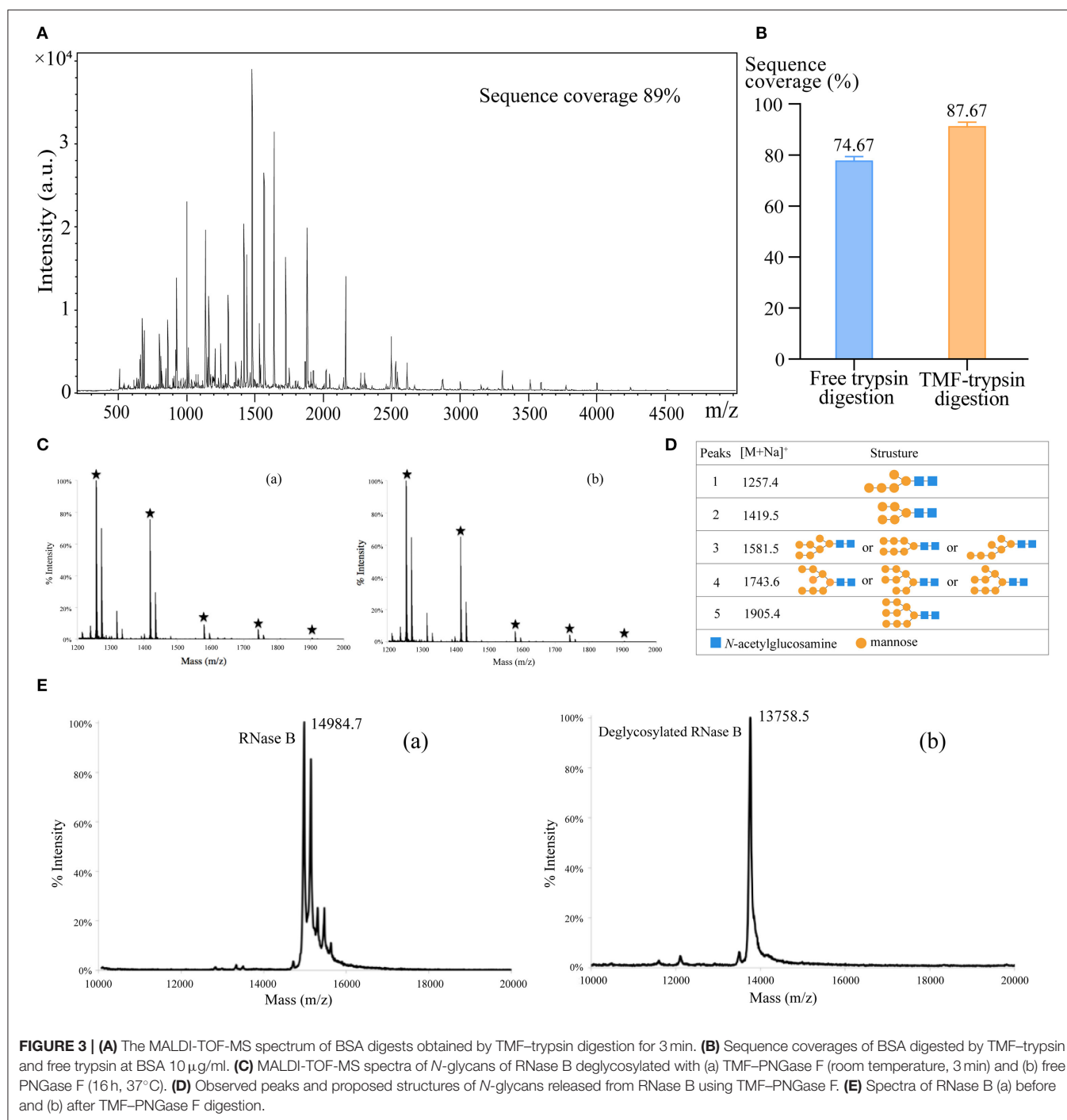
FIGURE 3 | (A) The MALDI-TOF-MS spectrum of BSA digests obtained by TMF–trypsin digestion for 3 min. **(B)** Sequence coverages of BSA digested by TMF–trypsin and free trypsin at BSA 10 μg/ml. **(C)** MALDI-TOF-MS spectra of *N*-glycans of RNase B deglycosylated with (a) TMF–PNGase F (room temperature, 3 min) and (b) free PNGase F (16 h, 37°C). **(D)** Observed peaks and proposed structures of *N*-glycans released from RNase B using TMF–PNGase F. **(E)** Spectra of RNase B (a) before and (b) after TMF–PNGase F digestion.

(**Figure 1**). After TMF–trypsin digestion of urine proteins, the peptides were subjected to *N*-glycopeptides enrichment by hydrophilic interaction liquid chromatography (HILIC) (Mysling et al., 2010; Cong et al., 2017). The *N*-glycopeptides are further subjected to deglycosylation with TMF–PNGase F in the $H_2^{18}O$ system. The ^{18}O-labeled *N*-glycopeptides were collected for mass spectrometry analysis. With TMF-enzymes, the total sample processing time from urine

protein digestion to MS analysis is about 1 h, which is at least 20 times faster than that with in-solution free enzyme digestion. Impressively, both TMF–trypsin and TMF–PNGase F digestion identified more unique proteins and peptides than in-solution digestion (**Figures 4A,D**). Compared with 793 protein groups and 4,718 peptides identified by free trypsin digestion (**Supplementary Tables 1, 2**), TMF–trypsin digestion yielded a slight increase in the number

FIGURE 4 | (A) Unique peptides and corresponding proteins identified by free trypsin and TMF–trypsin. **(B)** Identification reproducibility evaluation of TMF–trypsin by overlapping the identified peptides of three replicates. **(C)** Quantification reproducibility evaluation of TMF–trypsin. **(D)** Unique *N*-glycosylation sites and corresponding *N*-glycoproteins identified by free PNGase F and TMF–PNGase F. **(E)** Identification reproducibility evaluation of TMF–PNGase F by overlapping the identified peptides of three replicates. **(F)** Quantification reproducibility evaluation of the TMF–PNGase F. **(G)** Overlap of the proteins identified by free-trypsin and TMF–trypsin. **(H)** Overlap of the peptides of the 615 co-identified proteins by free-trypsin and TMF–trypsin. **(I)** Schematic diagram of overlapped sequence caused by missed cleavages.

of protein groups and peptides, which is 806 and 5,702, respectively (**Supplementary Tables 3, 4**). TMF–PNGase F also had a better performance, identifying 1,180 *N*-glycosylated protein groups and 1,939 *N*-glycosylation sites (**Supplementary Tables 7, 8**), while free PNGase F digestion identified 1,018 *N*-glycosylated protein groups and 1,342 *N*-glycosylation sites (**Supplementary Tables 5, 6**). Furthermore, we conducted three technical replicates to evaluate the reproducibility of the two TMF-immobilized enzymes. In TMF–trypsin digestion, 96% peptides were obtained in at least two tests. Likewise, ~88% of *N*-glycopeptides were identified in at least two tests of TMF–PNGase F digestion (**Figures 4B,E**).

Digestion products of both TMF-immobilized enzymes also have good quantitative reproducibility, with Pearson correlation coefficients higher than 0.95 for peptides and 0.88 for *N*-glycopeptides (**Figures 4C,F**), demonstrating their capability for actual application in complex samples.

We noticed that relatively more missed cleavage peptides were found in the digestion product of TMF–trypsin, which is disadvantageous for protein quantification. To further evaluate the digestion effect of TMF–trypsin, we analyzed the peptides in the 615 co-identified proteins by the two digestion methods (**Figure 4G**). Among the 2,856 peptides uniquely identified by TMF-trypsin digestion, 1,089 peptides were with partial

sequence overlapping with the corresponding peptides in the free-trypsin digestion product (**Figure 4H**) because of missed cleavages (**Figure 4I**). These partial overlapping peptides provide a little contribution to protein identification. However, for the rest of the 1,767 uniquely identified peptides obtained by TMF–trypsin, their amino acid sequences are completely different from those in the digestion product of free-trypsin, which is advantageous in improving the sequence coverage in protein identification. Analysis of amino acids sequence coverage of the co-identified proteins led to similar results. Of the 615 co-identified proteins, ∼60% have higher amino acids sequence coverage at the protein level in the TMF–trypsin digestion product. Furthermore, both free trypsin and TMF–trypsin provided a large number of uniquely identified peptides that were not covered by each other. Therefore, combined application of these two digestion methods may lead to increased protein sequence coverage and improved identification reliability, especially for the low abundant proteins, which usually only have a few or no unique peptides identified using only the free-trypsin digestion.

CONCLUSION

A novel thermoresponsive magnetic fluid (TMF)-immobilized enzyme (trypsin or PNGase F) for ultrafast and highly efficient proteome digestion and glycopeptide deglycosylation was developed in this study. The high-water dispersity and heat-triggered magnetic separation of TMF-immobilized enzymes make them capable of homogeneous digestion and heterogeneous separation. Therefore, highly efficient and rapid protein digestion and glycopeptide deglycosylation were achieved by avoiding mass transfer resistance in the heterogeneous system, and facile sample recovery was achieved by thermoflocculation and magnetic force. Taking advantage of these unique features, TMF-immobilized enzymes led to the identification of ∼2,000 *N*-glycopeptides in human urine sample and, with more than 20 times reduction in sample processing time, demonstrated the potential use of this strategy for fast and high throughput analysis of *N*-glycoproteome in clinical samples.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD024639.

AUTHOR CONTRIBUTIONS

WQ conceived the project. ZF, TL, and FZ designed the experiments and interpreted data. ZF carried out the synthesis of TMF-immobilized enzymes (trypsin or PNGase F) and performed urinary protein exaction, proteomics experiments, and MS analysis. TL and FZ performed characterization of TMF. ZF and TL performed digestion performance evaluation of TMF-immobilized enzymes. All authors commented on the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.676100/full#supplementary-material

REFERENCES

Cong, Y., Hu, L., Zhang, Z., Gao, Y., Dong, M., Qin, H., et al. (2017). Analysis of therapeutic monoclonal antibody glycoforms by mass spectrometry for pharmacokinetics study. *Talanta* 165, 664–670. doi: 10.1016/j.talanta.2017.01.023

Dong, M., Lih, T. M., Chen, S. Y., Cho, K. C., Eguez, R. V., Hoti, N., et al. (2020). Urinary glycoproteins associated with aggressive prostate cancer. *Theranostics* 10, 11892–11907. doi: 10.7150/thno.47066

Fauser, J., Savitskiy, S., Fottner, M., Trauschke, V., and Gulen, B. (2020). Sortase-mediated quantifiable enzyme immobilization on magnetic nanoparticles. *Bioconjug. Chem.* 31, 1883–1892. doi: 10.1021/acs.bioconjchem.0c00322

Fu, D., Chen, L., and O'Neill, R. A. (1994). A detailed structural characterization of ribonuclease B oligosaccharides by 1H NMR spectroscopy and mass spectrometry. *Carbohydr. Res.* 261, 173–186. doi: 10.1016/0008-6215(94)84015-6

Hart, G. W., and Copeland, R. J. (2010). Glycomics hits the big time. *Cell* 143, 672–676. doi: 10.1016/j.cell.2010.11.008

Huang, J., Dong, J., Shi, X., Chen, Z., Cui, Y., Liu, X., et al. (2019). 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.* 91, 11589–11597. doi: 10.1021/acs.analchem.9b01698

Krenkova, J., Lacher, N. A., and Svec, F. (2009). Highly efficient enzyme reactors containing trypsin and endoproteinase LysC immobilized on porous polymer monolith coupled to MS suitable for analysis of antibodies. *Anal. Chem.* 81, 2004–2012. doi: 10.1021/ac8026564

Létant, S. E., Hart, B. R., Kane, S. R., Hadi, M. Z., Shields, S. J., and Reynolds, J. G. (2004). Enzyme immobilization on porous silicon surfaces. *Adv. Mater.* 16, 689–693. doi: 10.1002/adma.200306173

Luo, J., Marpani, F., Brites, R., Frederiksen, L., Meyer, A. S., Jonsson, G., et al. (2014). Directing filtration to optimize enzyme immobilization in reactive membranes. *J. Memb. Sci.* 459, 1–11. doi: 10.1016/j.memsci.2014.01.065

Mateo, C., Palomo, J. M., Fernandez-Lorente, G., Guisan, J. M., and Fernandez-Lafuente, R. (2007). Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enzyme Microb. Technol.* 40, 1451–1463. doi: 10.1016/j.enzmictec.2007.01.018

Mereiter, S., Balmana, M., Campos, D., Gomes, J., and Reis, C. A. (2019). Glycosylation in the era of cancer-targeted therapy: where are we heading? *Cancer Cell* 36, 6–16. doi: 10.1016/j.ccell.2019.06.006

Mysling, S., Palmisano, G., Højrup, P., and Thaysen-Andersen, M. (2010). TFA- utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* 82, 5598–5609. doi: 10.1021/ac100530w

Pan, J., Hu, Y., Sun, S., Chen, L., Schnaubelt, M., Clark, D., et al. (2020). Glycoproteomics-based signatures for tumor subtyping and clinical outcome prediction of high-grade serous ovarian cancer. *Nat. Commun.* 11:6139. doi: 10.1038/s41467-020-19976-3

Qin, W., Song, Z., Fan, C., Zhang, W., Cai, Y., Zhang, Y., et al. (2012). Trypsin immobilization on hairy polymer chains hybrid magnetic nanoparticles for ultra fast, highly efficient proteome digestion, facile 18O labeling and absolute protein quantification. *Anal. Chem.* 84, 3138–3144. doi: 10.1021/ac2029216

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., and Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* 21, 729–749. doi: 10.1038/s41580-020-00294-x

Sharifi, M., Sohrabi, M. J., Hosseinali, S. H., Hasan, A., Kani, P. H., Talaei, A. J., et al. (2020). Enzyme immobilization onto the nanomaterials: application in enzyme stability and prodrug-activated cancer therapy. *Int. J. Biol. Macromol.* 143, 665–676. doi: 10.1016/j.ijbiomac.2019.12.064

Stadlmann, J., Taubenschmid, J., Wenzel, D., Gattinger, A., Durnberger, G., Dusberger, F., et al. (2017). Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* 549, 538–542. doi: 10.1038/nature24015

Sun, Y., Ding, X., Zheng, Z., Cheng, X., Hu, X., and Peng, Y. (2006). Magnetic separation of polymer hybrid iron oxide nanoparticles triggered by temperature. *Chem. Commun.* 26, 2765–2767. doi: 10.1039/b604202c

Wu, J., and Gao, Y. (2015). Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Rev. Proteomics* 12, 623–636. doi: 10.1586/14789450.2015.1094380

Yang, G., Hoti, N., Chen, S. Y., Zhou, Y., Wang, Q., Betenbaugh, M., et al. (2020). One-step enrichment of intact glycopeptides from glycoengineered Chinese hamster ovary cells. *Front. Chem.* 8:240. doi: 10.3389/fchem.2020.00240

Yuan, H., Zhang, S., Zhao, B., Weng, Y., Zhu, X., Li, S., et al. (2017). Enzymatic reactor with trypsin immobilized on graphene oxide modified polymer microspheres to achieve automated proteome quantification. *Anal. Chem.* 89, 6324–6329. doi: 10.1021/acs.analchem.7b00682

Yuce-Dursun, B., Cigil, A. B., Dongez, D., Kahraman, M. V., Ogan, A., and Demir, S. (2016). Preparation and characterization of sol–gel hybrid coating films for covalent immobilization of lipase enzyme. *J. Mol. Catal. B Enzymatic* 127, 18–25. doi: 10.1016/j.molcatb.2016.02.007

Zhang, W., Liu, T., Dong, H., Bai, H., Tian, F., Shi, Z., et al. (2017). Synthesis of a highly azide-reactive and thermosensitive biofunctional reagent for efficient enrichment and large-scale identification of O-GlcNAc proteins by mass spectrometry. *Anal. Chem.* 89, 5810–5817. doi: 10.1021/acs.analchem.6b04960

Zhao, M., Li, M., Yang, Y., Guo, Z., Sun, Y., Shao, C., et al. (2017). A comprehensive analysis and annotation of human normal urinary proteome. *Sci. Rep.* 7:3024. doi: 10.1038/s41598-017-03226-6

Zhao, M., Zhang, X., and Deng, C. (2015). Rational synthesis of novel recyclable Fe(3)O(4)@MOF nanocomposites for enzymatic digestion. *Chem. Commun.* 51, 8116–8119. doi: 10.1039/C5CC01908G

Zhao, T., Jia, L., Li, J., Ma, C., Wu, J., Shen, J., et al. (2020). Heterogeneities of site-specific N-glycosylation in HCC tumors with low and high AFP concentrations. *Front. Oncol.* 10:496. doi: 10.3389/fonc.2020.00496

Large-Scale Analysis of Apolipoprotein CIII Glycosylation by Ultrahigh Resolution Mass Spectrometry

Daniel Demus[1,2], Annemieke Naber[3], Viktoria Dotz[1], Bas C. Jansen[1,2], Marco R. Bladergroen[1], Jan Nouta[1], Eric J. G. Sijbrands[3], Mandy Van Hoek[3], Simone Nicolardi[1] and Manfred Wuhrer[1]*

[1]Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden, Netherlands, [2]Ludger Ltd., Culham Science Centre, Abingdon, United Kingdom, [3]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, Netherlands

Apolipoprotein-CIII (apo-CIII) is a glycoprotein involved in lipid metabolism and its levels are associated with cardiovascular disease risk. Apo-CIII sialylation is associated with improved plasma triglyceride levels and its glycosylation may have an effect on the clearance of triglyceride-rich lipoproteins by directing these particles to different metabolic pathways. Large-scale sample cohort studies are required to fully elucidate the role of apo-CIII glycosylation in lipid metabolism and associated cardiovascular disease. In this study, we revisited a high-throughput workflow for the analysis of intact apo-CIII by ultrahigh-resolution MALDI FT-ICR MS. The workflow includes a chemical oxidation step to reduce methionine oxidation heterogeneity and spectrum complexity. Sinapinic acid matrix was used to minimize the loss of sialic acids upon MALDI. MassyTools software was used to standardize and automate MS data processing and quality control. This method was applied on 771 plasma samples from individuals without diabetes allowing for an evaluation of the expression levels of apo-CIII glycoforms against a panel of lipid biomarkers demonstrating the validity of the method. Our study supports the hypothesis that triglyceride clearance may be regulated, or at least strongly influenced by apo-CIII sialylation. Interestingly, the association of apo-CIII glycoforms with triglyceride levels was found to be largely independent of body mass index. Due to its precision and throughput, the new workflow will allow studying the role of apo-CIII in the regulation of lipid metabolism in various disease settings.

Keywords: high-throughput, mass spectrometry, apolipoprotein-CIII, glycosylation, oxidation

INTRODUCTION

Lipid metabolism is regulated by complex biological mechanisms in which apolipoproteins – proteins embedded in lipoprotein particles – modulate the transport and availability of blood lipids (Mahley et al., 1984). Apolipoprotein-CIII (apo-CIII) is a 79 amino acid glycoprotein present on the surface of triglyceride-rich lipoproteins and is an inhibitor of lipoprotein lipase (LPL), an enzyme that hydrolyzes triglycerides into fatty acids (Shachter, 2001; Larsson et al., 2013, 2017). Apo-CIII has been associated with increased monocyte adhesion to the endothelium (Kawakami et al., 2006) and enhanced binding

of apoB-containing lipoproteins to vascular proteoglycans (Olin-Lewis et al., 2002). High apo-CIII levels are associated with hypertriglyceridemia (Kohan, 2015; Dai et al., 2019; Taskinen et al., 2019) and increased cardiovascular disease risk in the general population (Wyler Von Ballmoos et al., 2015; Rosenson et al., 2016; Dai et al., 2019) and diabetes mellitus (Juntti-Berggren and Berggren, 2017; Christopoulou et al., 2019). Recently, the clinical interest for this protein has increased due to the promising results obtained from antisense oligonucleotide-based therapies for the reduction of apo-CIII and triglyceride levels (Pollin et al., 2008; Rocha et al., 2017; Reyes-Soffer et al., 2019; Taskinen et al., 2019).

Apo-CIII exists in four major proteoforms: one non-glycosylated form (apo-CIII$_{0a}$) and three O-glycosylated variants with a core 1 (T-antigen) glycan structure, which is either non-sialylated (apo-CIII$_{0c}$), monosialylated (apo-CIII$_1$) or disialylated (apo-CIII$_2$) (Nedelkov, 2017; Ramms and Gordts, 2018). Low-abundance fucosylated, non-sialylated apo-CIII forms have also been described (Nicolardi et al., 2013a). It has been shown that not only the levels of apo-CIII but also the specific glycoforms and their relative expression control triglyceride metabolism (Yassine et al., 2015; Koska et al., 2016). For example, an inverse association between apo-CIII$_2$/apo-CIII$_1$ ratio and triglyceride levels has been confirmed by two independent studies (Koska et al., 2016; Kegulian et al., 2019). It has also been shown that sialylation modulates the apo-CIII affinity for hepatic receptors that clear lipoprotein particles (Kegulian et al., 2019) and that different proteoforms of apo-CIII may affect the inhibition of LPL (Holleboom et al., 2011) and the interaction of LDL with the vascular wall (Hiukka et al., 2009). Since the association of different apo-CIII proteoforms with specific cardiometabolic endpoints has not been fully elucidated, further research in large sample cohorts is warranted.

We have developed a high-throughput method based on magnetic-bead extraction and matrix-assisted laser desorption/ionization (MALDI) and ultrahigh-resolution Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry (MS) for the analysis of serum apo-CIII proteoforms (Nicolardi et al., 2013a; Nicolardi et al., 2013b). Apo-CIII contains methionine residues, which can be (partially) oxidized during biological processes *in vivo* (Stadtman et al., 2003), sample processing and freeze-thaw cycles (Borges et al., 2014). The presence of different oxidoforms increases mass spectra complexity, which complicates MS data processing and affects the repeatability of measurements. Although the analyte oxidation may not pose a serious challenge in MALDI MS analysis of single samples, it can seriously impact the precision and accuracy of quantitative measurements in large sample cohorts.

In the current study, we have applied a modified workflow employing a previously established MALDI FT-ICR MS method preceded by a chemical oxidation step for complete oxidation of apo-CIII methionine residues. This results in highly reproducible high-throughput measurements for relative quantification of apo-CIII proteoforms in a large number of plasma samples varying in protein oxidation levels. Furthermore, we have adopted sinapinic acid (SPA) as a MALDI matrix to minimize the loss of sialic acid induced by MALDI. The high-throughput

quantitation software, MassyTools (Jansen et al., 2015), was here further developed to facilitate semi-automated MS data processing for intact proteins. The validity of the new workflow was tested on a clinical cohort comprised of 771 plasma samples, which allowed the evaluation of the relationship between apo-CIII glycoforms and metabolic biomarkers, such as BMI, cholesterol, and triglyceride levels.

MATERIALS AND METHODS

Clinical Samples

Blood plasma samples from a group of individuals without diabetes of the DiaGene Study were used. The DiaGene Study is a case-control study comprising 1886 type-2 diabetes patients and 854 controls without diabetes, from the areas of Eindhoven and Veldhoven, in the Netherlands. The study is described in detail elsewhere (Van Herpt et al., 2017). For the current study, after quality control, apo-CIII glycosylation data were available for 771 samples, in 746 whereof, data on clinical characteristics were available. All participants gave their written informed consent. This study was approved by the Medical Ethics Committees of the Erasmus University Medical Center, Catharina Hospital and Maxima Medical Center.

Clinical information and blood samples were obtained at baseline, as described previously (Van Herpt et al., 2017). Triglycerides and cholesterol concentrations were measured using standard clinical chemistry essays and reported by the collecting clinic. Non-high-density lipoprotein (non-HDL)-cholesterol was calculated by subtracting the high-density lipoprotein (HDL)-cholesterol from the total cholesterol, body mass index (BMI) was calculated by dividing the body mass (in kg) by the square of the body length (in m). Triglyceride concentrations were logarithmically transformed before linear regression analysis, because of non-normal distribution.

Chemicals

Magnetic beads (Dynabeads RPC-18) were purchased from Invitrogen Dynal AS, Oslo, Norway. VisuCon-F plasma standard from Affinity Biologicals, Ancaster, Canada. Hydrogen peroxide 30%, ethanol and acetone were purchased from Merck, Darmstadt, Germany. Acetonitrile (ACN) was from Biosolve Chimie SARL, France. Trifluoroacetic acid (TFA) was purchased from Thermo Fisher Scientific, Tewksbury, MA. Sinapinic acid (SPA) and α-Cyano-4-hydroxycinnamic acid (HCCA) from Sigma-Aldrich. Ultrapure milliQ water (18 MΩ cm at 25°C) was used throughout.

High-Throughput RP-C18 Solid-Phase Extraction of Plasma Proteins

Plasma standards (VisuCon-F) were randomized over cohort sample plates. 10 μL of human blood plasma was transferred from the cohort sample plates into 96-well skirted PCR plates (4ti-0960/C, 4titude, Dorking, United Kingdom). 15 μL of an oxidizing solution (12% H_2O_2/0.5% TFA in water) was added to each sample. The plate was sealed with a pierce foil seal (4ti-0521,
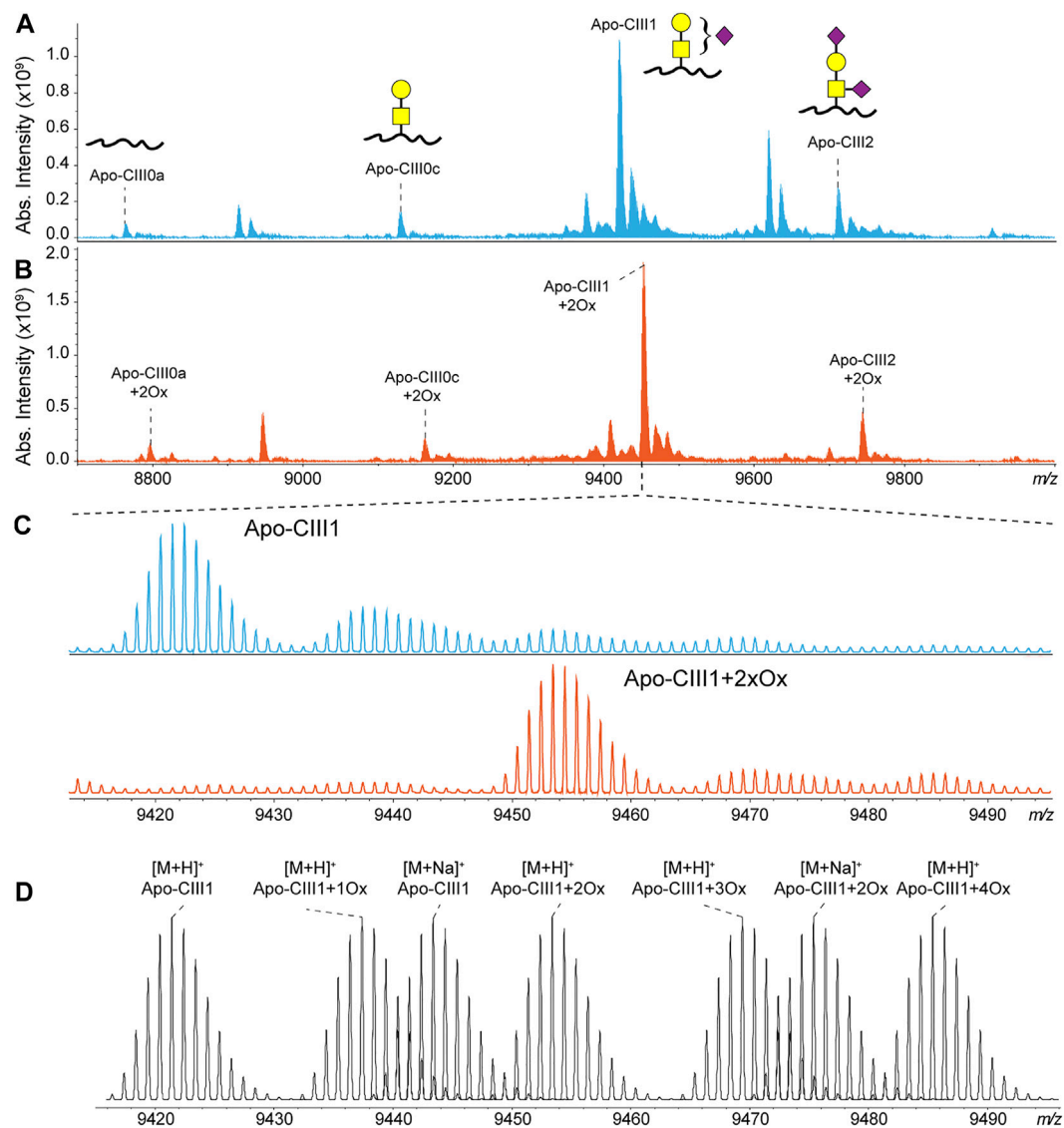
FIGURE 1 | MALDI FT-ICR mass spectra of apo-CIII proteoforms from **(A)** non-treated (blue) and **(B)** H_2O_2 treated (orange) human blood plasma samples. The four most abundant proteoforms of apo-CIII are annotated. Mass spectra from non-treated samples are characterized by the presence of oxidoforms as exemplified for apo-CIII1 in panel **(C)**. The H_2O_2 treatment allows the almost complete conversion of methionine residues to methionine sulfoxide with a minor conversion to methionine sulfone. Theoretical isotopic distribution of apo-CIII1 proteoforms are depicted in panel **(D)**. For graphical representations of glycan structures: yellow square (N-acetylgalactosamine), yellow circle (galactose), purple diamond (N-acetylneuraminic acid).

4titude Ltd., Wotton, Surrey, United Kingdom) and incubated for 1 h at 37°C. Subsequently, the plate was cooled at 4°C for 30 min and centrifuged briefly at 800 × g. The pierce foil was removed and the plate was transferred onto a liquid handling robot (Hamilton, Bonaduz, Switzerland) where solid-phase extraction (SPE) was carried out as follows: the RP-C18 beads were activated by three washes using acetonitrile (ACN) and trifluoroacetic acid (TFA) solution in water (first wash using 50% ACN/0.1% TFA followed by two washes with 0.1% TFA). Next, plasma samples were transferred to the activated beads and incubated for 10 min at room temperature. The incubation was followed by three washes: one wash using 15% ACN and two washes with 0.1% TFA. Proteins were eluted by adding 15 µL of 50% ACN/0.1% TFA in water and incubating for 5 min at room temperature. For MALDI spotting, 2 µL of sample eluates were mixed with either 16 µL of sinapinic acid solution (1.3 g/L in 2:1 v/v ethanol/acetone) or 15 µL alpha-cyano-4-hydroxycinnamic acid solution (1.4 g/L in 2:1 v/v ethanol/acetone). 1.5 µL of each sample mix was spotted in duplicate onto a MALDI AnchorChip target plate (800 µm anchor diameter; Bruker Daltonics, Bremen, Germany) and allowed to air-dry before MALDI MS analysis.
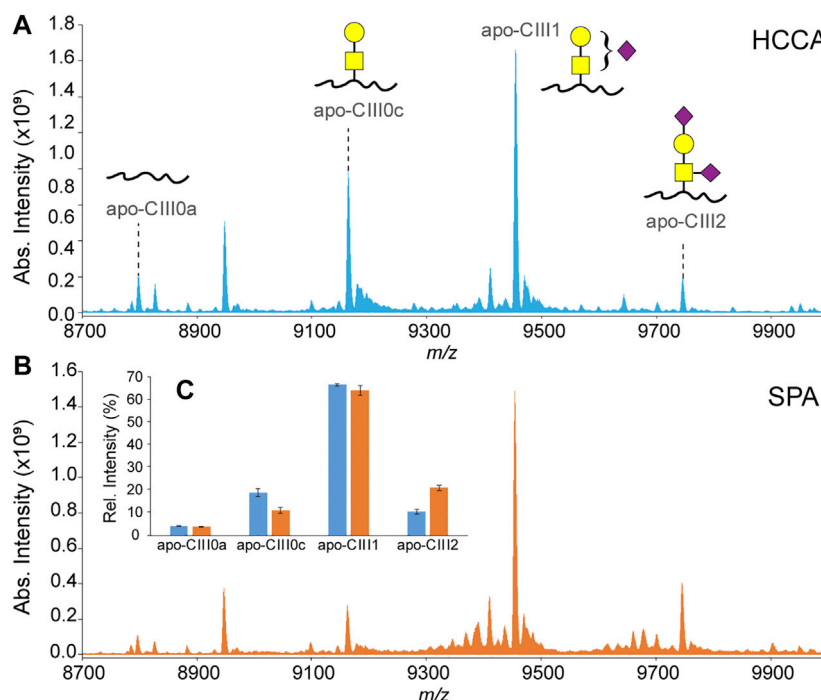
FIGURE 2 | MALDI-FT-ICR mass spectra of apo-CIII proteoforms from H_2O_2-treated human blood plasma samples obtained using either HCCA **(A)** or SPA **(B)** as MALDI matrices. Relative peak intensities of the four most abundant proteoforms of apo-CIII are shown for both sets (n = 3 in duplicate) of samples in the inset **(C)**.

MALDI FT-ICR Mass Spectrometry and MS Data Analysis

All MALDI MS experiments were performed on a 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) equipped with a Smartbeam II™ laser system (355 nm wavelength) and a ParaCell detector. All spectra were acquired in the m/z-range 3495–30,000, from the average of ten scans of 200 laser shots (at 500 Hz) each using 524,288 data points. The analyzer parameters were set as previously reported (Van Der Burgt et al., 2019). Briefly, measurements were performed with high trapping potentials (up to 8.5 V) and high ParaCell DC biases (up to 8.8 V) and with a Sweep excitation power of 57% for 13.5 µs. A laser power of 20% and "medium" laser focus was used for MS measurements using HCCA, while a laser power of 30% and "ultra-large" focus was used with SPA. Details on MS data processing and statistical analysis can be found in **Supplementary Material** (Additional experimental details: MS data processing and statistical analysis).

RESULTS AND DISCUSSION

The Controlled Oxidation of Methionine Residues Reduces the Complexity of Mass Spectra

A common event observed in proteomics studies is the oxidation of methionine residues due to biological and pathological

processes occurring *in vivo* (Stadtman et al., 2003), sample storage and processing (Borges et al., 2014). These reactions are so common that, in bottom-up studies, methionine oxidation is often included in the database search as a variable or even fixed modification. However, in general, the peptides generated by enzymatic digestion (e.g. using trypsin) are small and often do not contain methionine residues and although a (partial) oxidation of methionine residues increases the number of peptides in a digest, these unwanted reactions do not significantly affect the analysis (Lao et al., 2015; Hains and Robinson, 2017; Bettinger et al., 2020). In a clinical setting, the use of fresh samples may be an ideal approach, especially for diagnostic purposes based on profiling of intact protein. Whereas in cohort studies, the collection of large numbers of clinical samples, their storage, transfer between institutions and multiple use may lead to oxidation processes that affect the analysis of intact proteins by increasing the heterogeneity of proteoforms detected in a spectrum. The higher complexity increases the chance of overlapping signals and reduces the sensitivity of the measurements due to the spreading of the signal over a higher number of species. This results in MS spectra characterized by the presence of interfering species and very low abundant analyte peaks, which do not meet acceptable spectral quality criteria for consideration in the statistically significant quantitative analysis.

Apo-CIII contains two methionine residues which can be oxidized to form methionine sulfoxide (MetO) and methionine sulfone ($MetO_2$) although this latter form requires harsher oxidizing conditions (Kim et al., 2015; Lim et al., 2019).

TABLE 1 | Inter- and intra-plate variability. Relative peak intensities, standard deviation (SD) and coefficient of variation (CV) are given for the intra- and inter-plate variability based on 136 plasma standards.

		Apo-CIII$_{0a}$ +2Ox	Apo-CIII$_{0c}$ +2Ox	Apo-CIII$_1$ +2Ox	Apo-CIII$_2$ +2Ox
Inter-plate	Relative peak intensity	0.039	0.114	0.636	0.203
	SD	0.003	0.018	0.04	0.02
	CV	8%	16%	6%	10%
Average intra-plate	CV	7%	18%	1%	10%

Previously, MALDI-TOF MS methods have been used in analyses of apo-CIII proteoforms (Trenchevska et al., 2016). In such low-resolution methods, apo-CIII oxidoforms cannot be resolved, however, their presence results in the broadening and distortion of apo-CIII proteoforms' signals, which can eventually overlap or interfere with signals of other proteins affecting their quantification. The chance of signal interference increases when SPE is used for the enrichment of apo-CIII as it leads to the co-enrichment of other small plasma proteins. The application of more specific enrichment methods, such as immunocapture, may help to reduce signals interfering with the various apo-CIII oxidoforms, but was not implemented in the present study for simplicity reasons. Of note, apo-CIII proteoforms have been analyzed by methods employing LC systems (Kailemia et al., 2018; Olivieri et al., 2018). Despite certain advantages over MALDI-TOF MS, such as absolute quantification and enhanced resolution, the throughput of this approach remains relatively low.

Recently, we have developed a method for the analysis of apo-CIII proteoforms using ultrahigh-resolution MALDI FT-ICR MS (Nicolardi et al., 2013a). Apo-CIII proteoforms were mainly detected as singly charged ions. Thus, apo-CIII oxidoforms (1 and 2 times MetO) were detected at +15.995 Th and +31.990 Th from the non-oxidized forms (**Figure 1**). The degree of methionine oxidation of apo-CIII can vary greatly but the complete oxidation of apo-CIII (i.e. 100% conversion of the two methionine residues to MetO) is not commonly observed (**Supplementary Figure S1**). Therefore, for each of the four major proteoforms of apo-CIII, two additional oxidoforms were observed in MALDI FT-ICR MS spectra resulting in twelve proteoforms (**Figure 1**). In addition to that, we were able to detect C-terminal alanine cleaved and fucosylated proteoforms (**Supplementary Table S1**).

To reduce sample complexity, we included an oxidation step with hydrogen peroxide to perform a controlled oxidation of both apo-CIII methionine residues to MetO (**Figure 1**). While the implementation of the oxidation step added 2 h to the workflow for the analysis, it reduced the heterogeneity of the spectra and facilitated MS data processing using MassyTools software (see Implementation of MassyTools software for high-throughput MS data processing) (Jansen et al., 2015). The efficiency of the controlled oxidation was tested on 136 standard and 771 clinical plasma samples. The relative intensities between the non-, mono- and di-oxidized forms of apo-CIII$_{0a}$, apo-CIII$_{0c}$, apo-CIII$_1$, and apo-CIII$_2$ are reported in **Supplementary Tables S2, S3**. Oxidation rates over 90% were found for apo-CIII$_1$ and apo-CIII$_2$. Oxidation efficiency seemed to be lower for apo-CIII$_{0a}$, apo-CIII$_{0c}$, however, close inspection of the spectra revealed the presence of interfering species that contributed to the signal of the non- and mono-oxidized forms of apo-CIII$_{0a}$, apo-CIII$_{0c}$ thus increasing their apparent relative intensity (**Supplementary Figure S2**). Therefore, the controlled oxidation was considered efficient for all four proteoforms by providing consistent oxidation rates across standard and clinical plasma samples. These results supported our strategy of using only the signal of the di-oxidized apo-CIII proteoforms for further statistical analysis. The good efficiency and repeatability of the oxidation step allowed us to assess associations between apo-CIII glycosylation and different lipid markers using only the signal of the di-oxidized forms.

Minimizing Sialic Acid Loss Using Sinapinic Acid as MALDI Matrix

In our previously reported ultrahigh-resolution MALDI FT-ICR MS method for the analysis of apo-CIII proteoforms HCCA was used as a MALDI matrix (Nicolardi et al., 2013a; Nicolardi et al., 2013b). This compound was chosen to increase the sensitivity for other serum peptides and small proteins present in C18-SPE eluates obtained from the high-throughput enrichment step using magnetic beads. However, it is known that sialic acid loss can result from in-source decay fragmentation events of glycan structures even when linked to peptides and proteins. In fact, previous reports on the analysis of apo-CIII by MALDI-TOF MS were based on the use of a MALDI matrix colder than HCCA, namely SPA (Kegulian et al., 2019; Koska et al., 2016; Yassine et al., 2015). The use of SPA allowed to minimize the loss of sialic acid, as evidenced by an increased relative intensity of both the mono- and the disialylated apo-CIII proteoforms and leading to reproducible apo-CIII glycosylation profiles (**Figure 2**;

TABLE 2 | DiaGene cohort characteristics.

	Individuals without diabetes
Participants, n	746
Male sex, n (%)	290 (38.9)
Age, year	65.7 (6.7)
Age within males, year	66.0 (6.7)
Age within females, year	65.5 (6.7)
BMI, kg/m^2	25.4 (4.5)
HDL-cholesterol, mmol/l	1.48 (0.36)
non-HDL-cholesterol, mmol/l	4.09 (0.97)
LDL-cholesterol, mmol/l	3.56 (0.90)
Triglycerides, mmol/l	1.20 (0.68)
Total cholesterol, mmol/l	5.57 (0.99)
Use of lipid lowering therapy, n (%)	93 (12.5)

Mean (and standard deviation) for normal distribution and median (and interquartile range) for non-normal distributions (BMI and triglycerides).

TABLE 3 | Associations of apo-CIII glycosylation with clinical characteristics.

Characteristics	Apo-CIII$_{0a}$			Apo-CIII$_{0c}$			Apo-CIII$_1$		Apo-CIII$_2$	
Sex	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value
(mean ± SD)	0.05 ± 0.02	0.05 ± 0.02	0.605	0.11 ± 0.02	0.11 ± 0.02	0.558	0.63 ± 0.03	0.64 ± 0.03	**0.002**	0.21 ± 0.04	0.20 ± 0.04	**0.004**
	beta	p-value		beta	p-value		beta	p-value		beta	p-value	
Age	1.28E-05	9.10E-01		-8.84E-05	4.96E-01		-2.97E-04	5.12E-02		3.84E-04	6.47E-02	
BMI	-4.61E-04	**2.23E-02**		8.60E-04	**2.31E-04**		1.81E-03	**1.74E-11**		-2.21E-03	**2.41E-09**	
HDL cholesterol	1.52E-03	4.80E-01		-1.20E-03	6.29E-01		-7.23E-03	**1.24E-02**		6.87E-03	8.27E-02	
non-HDL cholesterol	1.36E-03	9.56E-02		1.21E-03	1.95E-01		6.53E-03	**1.46E-09**		-8.88E-03	**1.81E-09**	
LDL cholesterol	1.55E-03	7.55E-02		1.45E-03	1.49E-01		6.41E-03	**3.32E-08**		-9.15E-03	**8.51E-09**	
Triglycerides	2.61E-03	**7.45E-03**		1.60E-03	1.53E-01		1.25E-02	**9.02E-16**		-1.66E-02	**1.44E-16**	
Total cholesterol	1.51E-03	5.82E-02		9.98E-04	2.75E-01		5.28E-03	**6.77E-07**		-7.58E-03	**1.76E-07**	

Blue: negative associations, red: positive associations, bold: significant p-value. p-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations, bold values - significant p-values.

Supplementary Table S4). Importantly, compared to other matrices previously used for profiling of apo-CIII glycoforms such as DHB (Palmigiano et al., 2017; Wada and Okamoto, 2020), SPA provides more desirable matrix/analyte co-crystallization in the context of high-throughput, automated MALDI measurements.

Implementation of MassyTools Software for High-Throughput MS Data Processing

One of the advantages of using ultrahigh-resolution MS is that measurements at isotopic resolution provide more spectra information compared to broad-peak detection in linear mode MALDI-TOF MS. Previously, we showed that the goodness of the observed isotopic distributions can be used as a quality control parameter for the selection of high-quality spectra generated from the analysis of a large cohort of samples (Nicolardi et al., 2010). This concept was then implemented in a more powerful software–namely, MassyTools–developed for the high-throughput processing of MALDI mass spectra (Jansen et al., 2015). MassyTools allows the determination of a series of quality control parameters that can be used to perform a curation of MS data at different levels. Mass spectra with unacceptable internal calibration quality and low intensity were discarded at first. Then, the quality of the signal of each apo-CIII proteoform was assessed using the S/N and MME values determined for the most intense peak within an isotopic distribution. Additionally, the quality of such distribution (i.e. IPQ value) was taken into account. The distributions of values of these parameters over 136 standard and 771 clinical plasma samples are reported in **Supplementary Figures S3, S4**; **Supplementary Table S5**. The analytes passing the curation process were then used for statistical analysis.

As assessed on 136 standard plasma samples, which were distributed over 17 MALDI target plates measured over 28 days, the method provided good repeatability for relative quantitation of all four proteoforms with CVs in a range of 1–18% for average intra-plate and 6–16% for inter-plate variability (**Table 1**). While we reduced in-source decay by selecting SPA as a MALDI matrix,

we expect that partial sialic acid loss from apo-CIII$_1$ during MS analysis may lead to a slight, artificial increase in the apo-CIII$_{0c}$ glycoform abundance. Hence, fluctuations in the extent of sialic acid loss may contribute to the larger CVs for apo-CIII$_{0c}$.

Associations Between Apo-CIII Sialylation and Lipid Markers

We used this approach to determine non-glycosylated and the glycosylated non-sialylated, mono-sialylated and disialylated apo-CIII glycoforms within a cohort of 746 individuals without diabetes (cohort characteristics in **Table 2**) and test their association with a range of metabolic biomarkers. We found the association of disialylated apo-CIII$_2$ with overall improved lipid profiles and decreased BMI (**Table 3**; **Supplementary Table S6**), which is in accordance with some of the previous reports (Koska et al., 2016; Kegulian et al., 2019). A subgroup analysis in participants not using statins or fibrates, did not change these associations (**Supplementary Tables S7A,B**).

So far, the inhibitory effect of apo-CIII on LPL has been linked to total apo-CIII concentration, but not to the relative proportion of apo-CIII glycoforms (Olivieri et al., 2018). Recent studies proposed that the presence of apo-CIII on triglyceride-rich lipoproteins (TRLs) alters the affinity between TRLs and their receptors in the liver. Kegulian et al. demonstrated that the degree of apo-CIII sialylation directs TRLs to different hepatic clearance pathways, as shown in mice (Kegulian et al., 2019). In detail, apo-CIII$_1$-enriched very low-density lipoproteins (VLDLs) are preferentially cleared by faster-acting low-density lipoprotein (LDL) receptor (LDLR) and LDL receptor-related protein 1 (LRP1), whereas apo-CIII$_2$ directs VLDLs to syndecan 1 (SDC1) receptors that are characterized by a slower but larger capacity metabolism of TRLs. The same study also showed that a 13 weeks antisense oligonucleotide treatment for apo-CIII, which, as expected, reduced plasma TG levels, also altered relative abundances of these two glycoforms leading to an increase of apo-CIII$_2$ and a
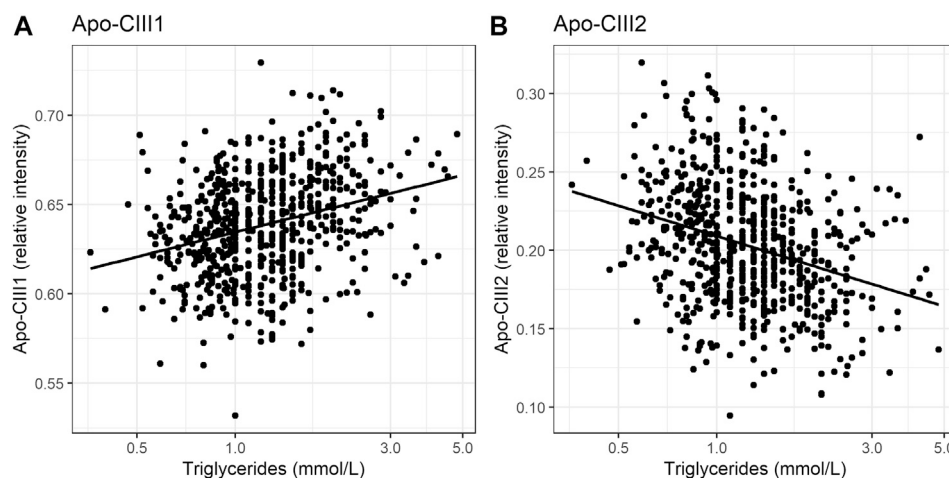
FIGURE 3 | Relationship of apo-CIII1 [panel **(A)**] and apo-CIII2 [panel **(B)**] with triglycerides.

decrease of apo-CIII$_1$. The increase of the apo-CIII$_2$/apo-CIII$_1$ ratio in a response to the antisense oligonucleotide therapy was explained by a differing capacity and clearance speed of the hepatic TRL receptors. In support of this, we observed in our cohort study of individuals without diabetes a positive association of the relative abundance of apo-CIII$_1$ glycoform with TG levels, and a negative association for apo-CIII$_2$ (**Figure 3**; **Table 3**).

Our study supports the hypothesis that triglyceride clearance may be regulated, or at least strongly influenced, by apo-CIII glycosylation, specifically sialylation. However, other aspects have to be considered. For instance, defective LDLR/LRP1-driven metabolic pathways might lead to decreased clearance of TGs. Expression and stability of LDLR and LRP1 in the liver might be affected by naturally occurring genetic variants (Oldoni et al., 2018; Paththinige et al., 2018; Reyes-Soffer et al., 2019; Xu et al., 2020). Moreover, it has been shown in mice that a high-fat diet can lead to the down-regulated expression of hepatic LRP1 by causing hyperglycemia with a high level of plasma triglycerides (Kim et al., 2014). In humans, obesity is associated with increases in plasma triglycerides (Howard et al., 2003; Franssen et al., 2011). We hypothesized that the association of apo-CIII glycoforms with triglycerides could be confounded by BMI. Surprisingly, after adjustment for BMI, the direction of effect and goodness-of-fit did not evidently change (**Supplementary Tables S8A,B**). This indicates that the association of apo-CIII sialylation with triglycerides is largely independent of BMI, and that it is not obesity-associated physiological changes that determine apo-CIII sialylation and its association with triglycerides.

Expression levels of apo-CIII were not investigated in this study. The differences in apo-CIII glycosylation profiles observed between individuals may be caused by varying expression levels of apo-CIII (Olivieri et al., 2018) or apo-CIII glycoforms (Holleboom et al., 2011), or the accumulation of certain glycoforms due to dysfunctional clearance pathways, based on recent findings by Kegulian et al. (Kegulian et al., 2019). It may also be a combination of the listed factors, which should be explored in further research. Nevertheless, from the results of this

study, we cannot determine whether apo-CIII sialylation influences triglyceride levels, or vice versa. Further studies are needed to elucidate the genetic and environmental factors that determine apo-CIII sialylation in health and disease.

CONCLUSION

Apo-CIII is a novel potential drug target in the management of cardiovascular disease driven by multiple studies demonstrating that plasma levels of apo-CIII are predictive of coronary heart disease and the risk of disease-related events (Borén et al., 2020). Previously, it has been shown that sialylated apo-CIII glycoforms are differentially cleared by hepatic receptors and that a higher apo-CIII$_2$/apo-CIII$_1$ ratio is associated with improved triglyceride levels (Kegulian et al., 2019). In humans, the production rates of these two glycoforms are comparable (Mauger et al., 2006), therefore varying apo-CIII2/apo-CIII1 ratios between individuals in healthy and disease groups might suggest various dysfunctional mechanisms involved in their production and clearance. This is the first large-scale study of apo-CIII glycosylation by ultrahigh resolution mass spectrometry. Clinical cohort studies employing large numbers of individuals will provide more insight into this topic, and the development of highly robust and accurate analytical methods enabling such large-scale studies is warranted.

Here, we present a workflow for high-throughput MALDI FT-ICR MS analysis of apo-CIII glycosylation in human plasma samples varying in protein oxidation levels. The controlled oxidation of apo-CIII methionine residues, the use of sinapinic acid as a MALDI matrix, and the use of MassyTools software for semi-automated, standardized spectra processing have been implemented to achieve highly repeatable measurements of intact apo-CIII proteoforms. The new analytical workflow allowed us to overcome the problem of the high spectral heterogeneity produced by methionine oxidation thus allowing the robust screening of a large cohort of plasma samples for the

relative quantitation of apo-CIII proteoforms. Importantly, the evaluation of MS spectra-derived quality parameters was implemented to minimize biases and ensure accuracy of collected data.

The cohort analysis confirmed that the level of apo-CIII sialylation is strongly associated with lipid biomarkers, especially with triglyceride levels. The relation between relative abundances of apo-CIII glycoforms and cardiovascular disease development should be further explored. More insight into the role of apo-CIII glycosylation in disease pathophysiology could provide new drug targets. Also, understanding of the mechanisms of existing drugs might increase by considering apo-CIII glycosylation. The methods presented, will enable such large-scale studies.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committees of the Erasmus University Medical Center and Catharina Hospital and Maxima Medical Center. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DD carried out all experiments and sample cohort analysis, partially assisted by SN, MB, and JN. DD and VD performed MS data processing. BJ provided a modified MassyTools script. AN performed association analysis assisted by DD. MW, MH and VD designed and supervised the study. ES and MH provided the study sample cohort. DD, SN and MW prepared the manuscript with significant contributions from all authors.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.678883/full#supplementary-material

REFERENCES

Bettinger, J. Q., Welle, K. A., Hryhorenko, J. R., and Ghaemmaghami, S. (2020). Quantitative Analysis of In Vivo Methionine Oxidation of the Human Proteome. *J. Proteome Res.* 19 (2), 624–633. doi:10.1021/acs.jproteome.9b00505

Borén, J., Packard, C. J., and Taskinen, M. R. (2020). The Roles of ApoC-III on the Metabolism of Triglyceride-Rich Lipoproteins in Humans. *Front. Endocrinol.* 11 (July), 1–10. doi:10.3389/fendo.2020.00474

Borges, C. R., Rehder, D. S., Jensen, S., Schaab, M. R., Sherma, N. D., Yassine, H., et al. (2014). Elevated Plasma Albumin and Apolipoprotein A-I Oxidation under Suboptimal Specimen Storage Conditions. *Mol. Cell Proteomics* 13 (7), 1890–1899. doi:10.1074/mcp.m114.038455

Christopoulou, E., Tsimihodimos, V., Filippatos, T., and Elisaf, M. (2019). Apolipoprotein CIII and Diabetes. Is There a Link? *Diabetes/Metabolism Res. Rev.* 35 (3), e3118. doi:10.1002/dmrr.3118

Dai, W., Zhang, Z., Yao, C., and Zhao, S. (2019). Emerging Evidences for the Opposite Role of Apolipoprotein C3 and Apolipoprotein A5 in Lipid Metabolism and Coronary Artery Disease. *Lipids Health Dis.* 18 (1), 1–7. doi:10.1186/s12944-019-1166-5

Franssen, R., Monajemi, H., Stroes, E. S. G., and Kastelein, J. J. P. (2011). Obesity and Dyslipidemia. *Med. Clin. North America* 95 (5), 893–902. doi:10.1016/j.mcna.2011.06.003

Fritz, B. V., Ruotolo, G., and Robbins, D. C. (2003). Obesity and Dyslipidemia. *Endocrinol. Metab. Clin. of North Am* 32 (4), 855–867. doi:10.1016/s0889-8529(03)00073-2

Hains, P. G., and Robinson, P. J. (2017). The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification. *J. Proteome Res.* 16 (9), 3443–3447. doi:10.1021/acs.jproteome.7b00022

Hiukka, A., Ståhlman, M., Pettersson, C., Levin, M., Adiels, M., Teneberg, S., et al. (2009). ApoCIII-enriched LDL in Type 2 Diabetes Displays Altered Lipid Composition, Increased Susceptibility for Sphingomyelinase, and Increased Binding to Biglycan. *Diabetes* 58 (9), 2018–2026. doi:10.2337/db09-0206

Holleboom, A. G., Karlsson, H., Lin, R.-S., Beres, T. M., Sierts, J. A., Herman, D. S., et al. (2011). Heterozygosity for a Loss-Of-Function Mutation in GALNT2 Improves Plasma Triglyceride Clearance in Man. *Cel Metab.* 14 (6), 811–818. doi:10.1016/j.cmet.2011.11.005

Jansen, B. C., Reiding, K. R., Bondt, A., Hipgrave Ederveen, A. L., Palmblad, M., Falck, D., et al. (2015). MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality Control Developed for Glycomic and Glycoproteomic MALDI-MS. *J. Proteome Res.* 14 (12), 5088–5098. doi:10.1021/acs.jproteome.5b00658

Juntti-Berggren, L., and Berggren, P. O. (2017). Apolipoprotein CIII Is a New Player in Diabetes. *Curr. Opin. Lipidol.* 28 (1), 27–31. doi:10.1097/MOL.0000000000000372

Kailemia, M. J., Wei, W., Nguyen, K., Beals, E., Sawrey-Kubicek, L., Rhodes, C., et al. (2018). Targeted Measurements of O- and N-Glycopeptides Show that Proteins in High Density Lipoprotein Particles Are Enriched with Specific Glycosylation Compared to Plasma. *J. Proteome Res.* 17 (2), 834–845. doi:10.1021/acs.jproteome.7b00604

Kawakami, A., Aikawa, M., Libby, P., Alcaide, P., Luscinskas, F. W., and Sacks, F. M. (2006). Apolipoprotein CIII in Apolipoprotein B Lipoproteins Enhances the Adhesion of Human Monocytic Cells to Endothelial Cells. *Circulation* 113 (5), 691–700. doi:10.1161/circulationaha.105.591743

Kegulian, N. C., Ramms, B., Horton, S., Trenchevska, O., Nedelkov, D., Graham, M. J., et al. (2019). ApoC-III Glycoforms Are Differentially Cleared by Hepatic TRL (Triglyceride-Rich Lipoprotein) Receptors. *Atvb* 39 (10), 2145–2156. doi:10.1161/atvbaha.119.312723

Kim, G., Weiss, S., Weiss, J., and Levine, R. (2015). Methionine Oxidation and Reduction in Proteins. *Biochim. Biophys. Acta* 1840 (2), 901–905. doi:10.1016/j.bbagen.2013.04.038

Kim, H. J., Moon, J. H., Kim, H. M., Yun, M. R., Jeon, B. H., Lee, B., et al. (2014). The Hypolipidemic Effect of Cilostazol Can Be Mediated by Regulation of Hepatic Low-Density Lipoprotein Receptor-Related Protein 1 (LRP1) Expression. *Metabolism* 63 (1), 112–119. doi:10.1016/j.metabol.2013.09.006

Kohan, A. B. (2015). Apolipoprotein C-III. *Curr. Opin. Endocrinol. Diabetes Obes.* 22 (2), 119–125. doi:10.1097/med.0000000000000136

Koska, J., Yassine, H., Trenchevska, O., Sinari, S., Schwenke, D. C., Yen, F. T., et al. (2016). Disialylated Apolipoprotein C-III Proteoform Is Associated with Improved Lipids in Prediabetes and Type 2 Diabetes. *J. Lipid Res.* 57 (5), 894–905. doi:10.1194/jlr.p064816

Lao, Y. W., Gungormusler-Yilmaz, M., Shuvo, S., Verbeke, T., Spicer, V., and Krokhin, O. V. (2015). Chromatographic Behavior of Peptides Containing Oxidized Methionine Residues in Proteomic LC-MS Experiments: Complex Tale of a Simple Modification. *J. Proteomics* 125, 131–139. doi:10.1016/j.jprot.2015.05.018

Larsson, M., Allan, C. M., Jung, R. S., Heizer, P. J., Beigneux, A. P., Young, S. G., et al. (2017). Apolipoprotein C-III Inhibits Triglyceride Hydrolysis by GPIHBP1-Bound LPL. *J. Lipid Res.* 58 (9), 1893–1902. doi:10.1194/jlr.m078220

Larsson, M., Vorrsjö, E., Talmud, P., Lookene, A., and Olivecrona, G. (2013). Apolipoproteins C-I and C-III Inhibit Lipoprotein Lipase Activity by Displacement of the Enzyme from Lipid Droplets. *J. Biol. Chem.* 288 (47), 33997–34008. doi:10.1074/jbc.m113.495366

Lim, J. M., Kim, G., and Levine, R. L. (2019). Methionine in Proteins: It's Not Just for Protein Initiation Anymore. *Neurochem. Res.* 44 (1), 247–257. doi:10.1007/s11064-017-2460-0

Mahley, R. W., Innerarity, T. L., Rall, S. C., and Weisgraber, K. H. (1984). Plasma Lipoproteins: Apolipoprotein Structure and Function. *J. Lipid Res.* 25 (12), 1277–1294. doi:10.1016/s0022-2275(20)34443-6

Mauger, J.-F., Couture, P., Bergeron, N., and Lamarche, B. (2006). Apolipoprotein C-III Isoforms: Kinetics and Relative Implication in Lipid Metabolism. *J. Lipid Res.* 47 (6), 1212–1218. doi:10.1194/jlr.m500455-jlr200

Nedelkov, D. (2017). Mass Spectrometric Studies of Apolipoprotein Proteoforms and Their Role in Lipid Metabolism and Type 2 Diabetes. *Proteomes* 5 (4), 27. doi:10.3390/proteomes5040027

Nicolardi, S., Palmblad, M., Dalebout, H., Bladergroen, M., Tollenaar, R. A. E. M., Deelder, A. M., et al. (2010). Quality Control Based on Isotopic Distributions for High-Throughput MALDI-TOF and MALDI-FTICR Serum Peptide Profiling. *J. Am. Soc. Mass. Spectrom.* 21 (9), 1515–1525. doi:10.1016/j.jasms.2010.05.004

Nicolardi, S., Van Der Burgt, Y. E. M., Dragan, I., Hensbergen, P. J., and Deelder, A. M. (2013a). Identification of New Apolipoprotein-CIII Glycoforms with Ultrahigh Resolution MALDI-FTICR Mass Spectrometry of Human Sera. *J. Proteome Res.* 12 (5), 2260–2268. doi:10.1021/pr400136p

Nicolardi, S., van der Burgt, Y. E. M., Wuhrer, M., and Deelder, A. M. (2013b). Mapping O -glycosylation of Apolipoprotein C-III in MALDI-FT-ICR Protein Profiles. *PROTEOMICS* 13 (6), 992–1001. doi:10.1002/pmic.201200293

Oldoni, F., van Capelleveen, J. C., Dalila, N., Wolters, J. C., Heeren, J., Sinke, R. J., et al. (2018). Naturally Occurring Variants in LRP1 (Low-Density Lipoprotein Receptor-Related Protein 1) Affect HDL (High-Density Lipoprotein) Metabolism through ABCA1 (ATP-Binding Cassette A1) and SR-B1 (Scavenger Receptor Class B Type 1) in Humans. *Arterioscler. Thromb. Vasc. Biol.* 38 (7), 1440–1453. doi:10.1161/atvbaha.117.310309

Olin-Lewis, K., Krauss, R. M., La Belle, M., Blanche, P. J., Barrett, P. H. R., Wight, T. N., et al. (2002). ApoC-III Content of apoB-Containing Lipoproteins Is Associated with Binding to the Vascular Proteoglycan Biglycan. *J. Lipid Res.* 43 (11), 1969–1977. doi:10.1194/jlr.m200322-jlr200

Olivieri, O., Chiariello, C., Martinelli, N., Castagna, A., Speziali, G., Girelli, D., et al. (2018). Sialylated Isoforms of Apolipoprotein C-III and Plasma Lipids in Subjects with Coronary Artery Disease. *Clin. Chem. Lab. Med.* 56 (9), 1542–1550. doi:10.1515/cclm-2017-1099

Palmigiano, A., Bua, R. O., Barone, R., Rymen, D., Régal, L., Deconinck, N., et al. (2017). MALDI-MS Profiling of Serum O -glycosylation and N -glycosylation in COG5-CDG. *J. Mass. Spectrom.* 52 (6), 372–377. doi:10.1002/jms.3936

Paththinige, C. S., Rajapakse, J. R. D. K., Constantine, G. R., Sem, K. P., Singaraja, R. R., Jayasekara, R. W., et al. (2018). Spectrum of Low-Density Lipoprotein Receptor (LDLR) Mutations in a Cohort of Sri Lankan Patients with Familial Hypercholesterolemia - A Preliminary Report. *Lipids Health Dis.* 17 (1), 1–7. doi:10.1186/s12944-018-0763-z

Pollin, T. I., Damcott, C. M., Shen, H., Ott, S. H., Shelton, J., Horenstein, R. B., et al. (2008). A Null Mutation in Human APOC3 Confers a Favorable Plasma Lipid Profile and Apparent Cardioprotection. *Science* 322 (5908), 1702–1705. doi:10.1126/science.1161524

Ramms, B., and Gordts, P. L. S. M. (2018). Apolipoprotein C-III in Triglyceride-Rich Lipoprotein Metabolism. *Curr. Opin. Lipidol.* 29 (3), 171–179. doi:10.1097/mol.0000000000000502

Reyes-Soffer, G., Sztalryd, C., Horenstein, R. B., Holleran, S., Matveyenko, A., Thomas, T., et al. (2019). Effects of APOC3 Heterozygous Deficiency on Plasma Lipid and Lipoprotein Metabolism. *Atvb* 39 (1), 63–72. doi:10.1161/atvbaha.118.311476

Rocha, N. A., East, C., Zhang, J., and McCullough, P. A. (2017). ApoCIII as a Cardiovascular Risk Factor and Modulation by the Novel Lipid-Lowering Agent Volanesorsen. *Curr. Atheroscler. Rep.* 19 (12), 62. doi:10.1007/s11883-017-0697-3

Rosenson, R. S., Brewer, H. B., Ansell, B. J., Barter, P., Chapman, M. J., Heinecke, J. W., et al. (2016). Dysfunctional HDL and Atherosclerotic Cardiovascular Disease. *Nat. Rev. Cardiol.* 13 (1), 48–60. doi:10.1038/nrcardio.2015.124

Shachter, N. S. (2001). Apolipoproteins C-I and C-III as Important Modulators of Lipoprotein Metabolism. *Curr. Opin. Lipidol.* 12 (3), 297–304. doi:10.1097/00041433-200106000-00009

Stadtman, E. R., Moskovitz, J., and Levine, R. L. (2003). Oxidation of Methionine Residues of Proteins: Biological Consequences. *Antioxid. Redox Signaling* 5 (5), 577–582. doi:10.1089/152308603770310239

Taskinen, M. R., Packard, C. J., and Borén, J. (2019). Emerging Evidence that ApoC-III Inhibitors Provide Novel Options to Reduce the Residual CVD. *Curr. Atheroscler. Rep.* 21 (8), 27. doi:10.1007/s11883-019-0791-9

Trenchevska, O., Nelson, R. W., and Nedelkov, D. (2016). Mass Spectrometric Immunoassays in Characterization of Clinically Significant Proteoforms. *Proteomes* 4 (1), 1623–1633. doi:10.3390/proteomes4010013

Van Der Burgt, Y. E. M., Kilgour, D. P. A., Tsybin, Y. O., Srzentić, K., Fornelli, L., Beck, A., et al. (2019). Structural Analysis of Monoclonal Antibodies by Ultrahigh Resolution MALDI In-Source Decay FT-ICR Mass Spectrometry. *Anal. Chem.* 91 (3), 2079–2085. doi:10.1021/acs.analchem.8b04515

Van Herpt, T. T. W., Lemmers, R. F. H., Van Hoek, M., Langendonk, J. G., Erdtsieck, R. J., Bravenboer, B., et al. (2017). Introduction of the DiaGene Study: Clinical Characteristics, Pathophysiology and Determinants of Vascular Complications of Type 2 Diabetes. *Diabetology Metab. Syndr.* 9 (1), 1–10. doi:10.1186/s13098-017-0245-x

Wada, Y., and Okamoto, N. (2020). Apolipoprotein C-III O-Glycoform Profiling of 500 Serum Samples by Matrix-Assisted Laser Desorption/ionization Mass Spectrometry for Diagnosis of Congenital Disorders of Glycosylation. *J. Mass. Spectrom.* 56 (4), e4597. doi:10.1002/jms.4597

Wyler Von Ballmoos, M. C., Haring, B., and Sacks, F. M. (2015). The Risk of Cardiovascular Events with Increased Apolipoprotein CIII: A Systematic Review and Meta-Analysis. *J. Clin. Lipidol.* 9 (4), 498–510. doi:10.1016/j.jacl.2015.05.002

Xu, Y.-X., Peloso, G. M., Nagai, T. H., Mizoguchi, T., Deik, A., Bullock, K., et al. (2020). EDEM3 Modulates Plasma Triglyceride Level through its Regulation of LRP1 Expression. *IScience* 23 (4), 100973. doi:10.1016/j.isci.2020.100973

Yassine, H. N., Trenchevska, O., Ramrakhiani, A., Parekh, A., Koska, J., Walker, R. W., et al. (2015). The Association of Human Apolipoprotein C-III Sialylation Proteoforms with Plasma Triglycerides. *PLoS ONE* 10 (12), 1–14. doi:10.1371/journal.pone.0144138

Identification of Dysregulated Complement Activation Pathways Driven by N-Glycosylation Alterations in T2D Patients

Yang Zhao[1†], Man Wang[1,2†], Bo Meng[1], Ying Gao[1], Zhichao Xue[1], Minjun He[1], You Jiang[1], Xinhua Dai[1]*, Dan Yan[2,3,4]* and Xiang Fang[1]*

[1]Center for Advanced Measurement Science, National Institute of Metrology, Beijing, China, [2]College of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [3]Department of Pharmacy, Beijing Friendship Hospital, Capital Medical University, Beijing, China, [4]Beijing Key Laboratory of Bio-characteristic Profiling for Evaluation of Rational Drug Use, Beijing Shijitan Hospital, Capital Medical University, Beijing, China

Diabetes has become a major public health concern worldwide, most of which are type 2 diabetes (T2D). The diagnosis of T2D is commonly based on plasma glucose levels, and there are no reliable clinical biomarkers available for early detection. Recent advances in proteome technologies offer new opportunity for the understanding of T2D; however, the underlying proteomic characteristics of T2D have not been thoroughly investigated yet. Here, using proteomic and glycoproteomic profiling, we provided a comprehensive landscape of molecular alterations in the fasting plasma of the 24 Chinese participants, including eight T2D patients, eight prediabetic (PDB) subjects, and eight healthy control (HC) individuals. Our analyses identified a diverse set of potential biomarkers that might enhance the efficiency and accuracy based on current existing biological indicators of (pre) diabetes. Through integrative omics analysis, we showed the capability of glycoproteomics as a complement to proteomics or metabolomics, to provide additional insights into the pathogenesis of (pre)diabetes. We have newly identified systemic site-specific N-glycosylation alterations underlying T2D patients in the complement activation pathways, including decreased levels of N-glycopeptides from C1s, MASP1, and CFP proteins, and increased levels of N-glycopeptides from C2, C4, C4BPA, C4BPB, and CFH. These alterations were not observed at proteomic levels, suggesting new opportunities for the diagnosis and treatment of this disease. Our results demonstrate a great potential role of glycoproteomics in understanding (pre)diabetes and present a new direction for diabetes research which deserves more attention.

Keywords: proteomics, glycoproteomics, N-glycopeptides, diabetes, complement

INTRODUCTION

Diabetes mellitus is a group of chronic diseases that could cause severe damage to various organs in human body, leading to disabling and life-threatening health complications (American Diabetes Association, 2019). It is estimated that 463 million individuals have diabetes mellitus globally in 2019, 90% of which are type 2 diabetes (T2D). By the year 2045, this number is projected to increase to 700 million (Saeedi et al., 2019). Due to its chronic nature, diabetes mellitus causes devastating

personal suffering and socioeconomic costs (Hernández-Jiménez et al., 2019; Williams et al., 2020). Early detection may allow the early medical interventions and lifestyle modifications that can largely delay or even prevent the onset of diabetes and its complications (Federation, 2019). Therefore, new biomarkers that enhance early detection are in urgent need.

The routine clinical assessment of T2D is mainly based on the measurement of fasting plasma glucose (FPG), 2-h plasma glucose (2 h PG) from oral glucose tolerance test (OGTT), or glycosylated hemoglobin A1c (HbA1c) concentrations (the Association A.D., 2019). 2 h PG is a time-consuming and expensive test which is unpopular with both patients and physicians in recent years (Pippitt et al., 2016). Due to the great convenience of FPG measurement, FPG test has now been widely used for the detection of T2D, but with the limitations of lower sensitivity, repeatability, and reproducibility (Waugh et al., 2013; Gar et al., 2018). Compared with FPG and 2 h PG tests, HbA1c tests often detect T2D at later stages with less interpersonal variability when repeated. However, it is reported that many factors could influence the hemoglobin glycation independently of glycemia, including HIV treatment, race/ethnicity, pregnancy status, genetic background, and anemia/hemoglobinopathies (Kim et al., 2009; Eckhardt et al., 2012). Given the above limitations of current assessments and the strong demand for early medical intervention, current research efforts are sought to identify a serial of reliable, sensitive, and noninvasive biomarkers from human plasma to improve the accuracy of the early diagnosis for T2D patients and to help elucidate the underlying pathogenesis.

Recent advances in proteome technologies have now enabled the large-scale studies of proteins in tissues and body fluids, and identified many candidate protein biomarkers to diagnose T2D (Zhao et al., 2017; Nowak et al., 2018; Sohail et al., 2018; Jiang et al., 2019; Zhang et al., 2019; Zheng et al., 2020). However, few of those potential biomarkers have been successfully translated into clinical use. N-glycosylation is a heterogeneous posttranslational modification of proteins that plays a critical role in disease pathologies (Haltiwanger and Lowe, 2004; Ohtsubo and Marth, 2006; Vajaria and Patel, 2017; Pan et al., 2020). For example, we recently revealed that the alterations of N-glycopeptides were significantly associated with tumor progression in prostate and papillary thyroid carcinoma (Zhang et al., 2019). But until now, only a limited number of high-throughput glycoproteomic analyses have been carried out specifically for T2D (Miura and Endo, 2016; Sharma et al., 2020). Here, we propose that integrated proteomic and glycoproteomic analyses of human plasma will identify novel biomarkers that could provide additional insight into the pathogenesis of T2D and the assessments of which could be more accurate and sensitive for disease diagnosis than currently available test methods.

Therefore, we conducted comprehensive proteomic and glycoproteomic analyses in the plasma samples of 24 Chinese individuals, including eight T2D patients, eight prediabetic (PDB) subjects, and eight healthy control (HC) individuals. Our goal was to investigate the proteomic and glycoproteomic profile changes in the plasma of T2D patients, and also to provide new clues for exploring molecular and pathological mechanisms of T2D.

MATERIALS AND METHODS

Sample Assembly

The study protocol was approved by the internal review boards and ethical committee boards of participating institutions. All patients have provided written informed consent. The study is registered at the Chinese Clinical Trial Registry (number: ChiCTR 1800014301).

Twenty-four plasma samples were collected for proteomic and glycoproteomic analyses, including eight T2D patients, eight PDB subjects, and eight HC individuals. All the patients were diagnosed as per the World Health Organization fasting blood glucose criteria (The World Health Organization, 2006).

Sample Preparation

The 12 highest abundance proteins in plasma were first removed by the Pierce™ TOP 12 Abundant Protein Depletion Spin Columns (Thermo) according to the user manual. For proteomic analysis, the depleted samples were digested and prepared by the filter-aided sample preparation (FASP) method with optimization for plasma (Wiśniewski et al., 2009). In brief, 10 μl of plasma were depleted by the top 12 depletion kit, and then the depleted samples were diluted with UA buffer into 200 μl (8 M urea in 0.1 M Tris-HCl, pH 8.5) and centrifuged on a 30-kDa filter for 15 min. After centrifugation, 200 μl UA solution with 10 mM dithiothreitol (DTT) was added, and the reduction reaction was kept for 4 h at 37°C. The solution was removed by centrifugation at 14,000 g, and 200 μl UA solution with 50 mM iodoacetamide (IAA) was added. The mixture was incubated in the dark for 30 min at room temperature. The mixture was then washed three times with 200 μl UA and 200 μl ABC (50 mM ammonium bicarbonate) by centrifugation at 14,000 g for 15 min at room temperature. Then, 100 μl of ABC containing 0.1 μg/μl trypsin was added to each filter tube and incubated at 37°C for 12 h. The digested peptides were collected by washing the filter tubes with 100 μl water followed by 15 min of 14,000 g centrifugation twice. The peptide concentration was measured using a NanoDrop OneC (Thermo) at 280 nm absorbance (Jiang et al., 2019). The digested peptides were completely dried by SpeedVac centrifuge at 45°C (Eppendorf, Concentrator plus) and then stored at −80°C for further LC-MS/MS analysis.

For glycoproteomic analysis, 50 μl of 80% ACN/0.2% TFA solution was used to resuspend the digested peptides, and the concentration was determined by measuring 2 μl of suspension through NanoDrop OneC (Thermo) using the ε205 = 31 method. Meanwhile, Venusil HILIC (5 μm, 100 Å) was activated by washing with 0.1% TFA and 80% ACN/0.2%TFA three times for 10 min. 50 μg of each peptide suspension was mixed with 5 mg activated Venusil HILIC (5 μm, 100 Å) and subjected to a 2-h rotation at room temperature. Then, the mixtures were loaded onto a Pipet Tip (Axygen, Inc., Union City, CA, United States) packed with C8 membrane and washed twice with 80%

ACN/0.2% TFA. Intact N-glycopeptides bound to HILIC column were collected by eluting with 70 µl of 0.1% TFA for three times. The eluents were then pooled and dried with SpeedVac centrifuge at 45°C (Eppendorf, Concentrator plus) and stored at −80°C for further LC-MS/MS analysis.

LC-MS/MS Analysis

Samples were measured using LC-MS instrumentation consisting of an EASY-nLC 1,200 ultrahigh-pressure system (Thermo Fisher Scientific) coupled *via* a nano-electrospray ion source (Thermo Fisher Scientific) to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). For proteomic analysis, 0.5 µg of peptide mixture resolved in buffer A [0.1% formic acid (FA)] were loaded onto a 2-cm self-packed trap column (100-µm inner diameter, ReproSil-Pur C18-AQ, 3 µm; Dr. Maisch) using buffer A and separated on a 75-µm inner-diameter column with a length of 25 cm (ReproSil-Pur C18-AQ, 1.9 µm; Dr. Maisch) over a 120-min gradient (buffer A, 0.1% FA in water; buffer B, 0.1% FA in 80% ACN) at a flow rate of 600 nl/min (0–16 min, 3–10% B; 16–76 min, 10–22% B; 76–106 min, 22–30% B; 106–118 min, 30–90% B; 118–120 min, 90% B). The Orbitrap Fusion Lumos was set to the OT–OT mode. For a full mass spectrometry survey scan, the target value was 5×10^5, and the scan ranged from 300 to 1,500 m/z at a resolution of 120,000 and a maximum injection time of 50 ms. For the MS2 scan, a duty cycle of 3 s was set with the top-speed mode. Only spectra with a charge state of 2–7 were selected for fragmentation by higher-energy collision dissociation with a normalized collision energy of 35%. The MS2 spectra were acquired in the Obitrap with an AGC target of 50,000 and a maximum injection time of 30 ms (Jiang et al., 2019).

For glycoproteomic analysis, 0.5 µg of N-glycopeptides reconstituted in 0.1% FA were separated over a gradient of 78 min at a flow rate of 400 nl/min (0–5 min, 6–10% B; 5–58 min, 10–22% B; 58–70 min, 22–50% B; 70–73 min, 50–90% B; 73–75 min, 90% B; 75–78 min, 90–10% B). For a full MS scan, the Orbitrap resolution was set to 120,000, with an AGC target value of 2×10^5 for a scan range of 800–2,000 m/z and a maximum injection time of 100 ms. For MS2 scan, the higher-energy collision dissociation fragmentation was performed at the isolation width of 2 m/z and a 30% normalized collision energy. The MS2 spectra were collected in the Orbitrap detector with a 5×10^5 AGC target, a maximum injection time of 250 ms, and a dynamic exclusion duration of 15 s (Zhang et al., 2019).

Data Analysis

For proteomic analysis, the tandem mass spectra were searched against the human UniProt database (version 20200911, 20,375 sequences) using MaxQuant (version 1.6.12.0). Trypsin was selected as the proteolytic enzyme, and two missed cleavage sites were allowed. Cysteine carbamidomethylation was set as the fixed modification. The oxidation of M and acetylation of the protein N-terminal were set as the variable modifications. The first search mass tolerance was 20 ppm, and the main search peptide tolerance was 4.5 ppm. The false discovery rates of the peptide–spectrum matches (PSMs) and proteins were set to less than 1% (Cox and Mann, 2008; Jiang et al., 2019).

For glycoproteomic analysis, the raw MS files of enriched N-glycopeptides were searched against the human Swiss-Prot database (version 20200911, 20,375 sequences) with pGlyco v. 2.2.2 (Liu et al., 2017), as previously described by Zhang et al. (2019). The following parameters were used. Mass tolerances for the precursors and fragment ions were set as ±5 and ±20 ppm, respectively. Two missed cleavages sites were allowed for trypsin digestion. The fixed modification was carbamidomethylation of all cysteine residues (+57.02 Da). Variable modifications included oxidation of methionine (+15.99 Da), deamidation of asparagine (+0.98 Da), and acetylation of the protein N-terminal (+42.01 Da). The N-glycosylation sequon (N-X-S/T/C; X ≠ P) was modified by changing "N" to "J." Both of these had the same mass. Quality control methods for intact glycopeptide identification were set to the 1% glycopeptide–spectrum matche (GPSM) false discovery rate (FDR). Quantification information (MS1 peak intensity) of N-glycopeptides spectra was acquired from the "allpeptide.txt" file in MaxQuant (Max Planck Gesellschaft, Munich, Germany) based on their unique MS/MS scan numbers from pGlyco 2.0 results.

Bioinformatic Analysis

The bioinformatic analysis was mainly performed in the statistical analysis environment R (version 3.5.4) (The R Core Team, 2012). The expression matrix was normalized based on the median values of each sample and then subjected to quantitative analyses. In detail, the abundance of each protein or each N-glycopeptide was divided by the median value of the corresponding sample, respectively, and then subjected to log2 transformation for further analysis (Ge et al., 2018). Missing values were imputed using the minimum value across the proteome or glycoproteome data. Cellular localization and molecular function of identified proteins in proteomics and glycoproteomics were analyzed by using ClueGo app (v.2.3.3) plugin Cytoscape software (v.3.5.1), based on Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome data resources (Ashburner et al., 2000; Kanehisa et al., 2002; Shannon et al., 2003; Bindea et al., 2009; Croft et al., 2011). A simple linear model and moderated t-statistics were used to determine the significantly changed proteins and N-glycopeptides using R/Bioconductor package limma (Ritchie et al., 2015). The p values were corrected for multiple testing using the Benjamini–Hochberg (BH) procedure. In proteomics, proteins with a *p*-value < 0.05 and an absolute log2-fold change ≥1 were considered as differential expressed signatures. In glycoproteomics, the signature N-glycopeptides were determined based on a corrected *p*-value < 0.01 and an absolute log2-fold change ≥1. Functional characterization of identified signatures was performed by using the hypergeometric test (Falcon and Gentleman, 2008), based on three categories of GO cellular component (CC), molecular function (MF), and biological process (BP) gene sets obtained from the MSigDB database v.7.2 (Subramanian et al., 2005, http://software.broadinstitute.org/gsea/msigdb/index.jsp). All the interaction networks were constructed by using Cytoscape application.
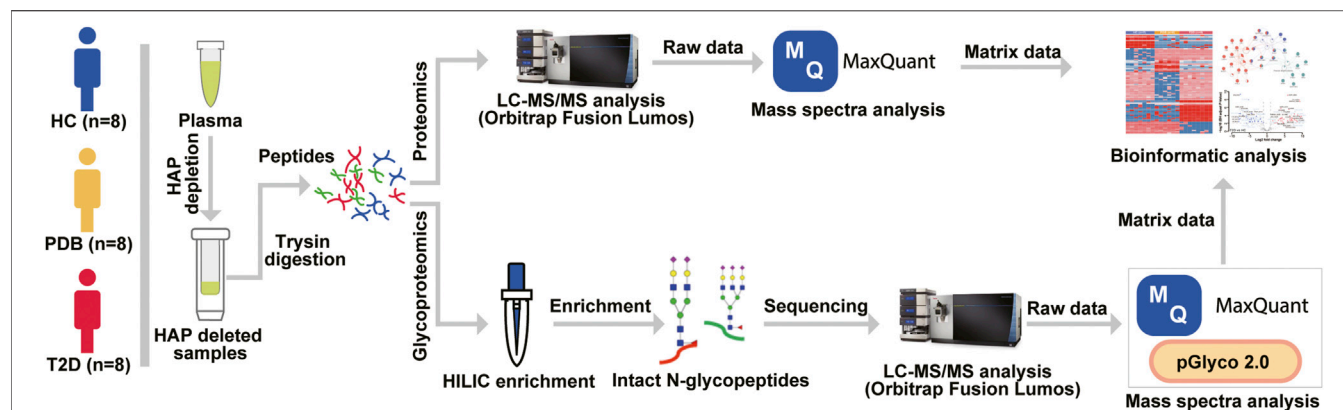
FIGURE 1 | Mass spectrometry–based proteomic and glycoproteomic workflow. The 12 highest abundance proteins (HAPs) in the plasma were removed by Pierce™ TOP 12 Abundant Protein Depletion Spin Columns according to the user manual, and then the HAP depleted samples were digested by a modified FASP protocol. For proteomic analysis, the digested peptides were directly analyzed by an Orbitrap Fusion Lumos LC-MS/MS instrument, and mass spectra were searched against the human UniProt database using MaxQuant software before bioinformatic analysis. For glycoproteomic analysis, a Hilic enrichment analysis was conducted for digested peptides, and then the enriched N-glycopeptides were analyzed by LC-MS/MS instrument. The raw data were processed by MaxQuant and pGlyco software before bioinformatic analysis.

Data Availability Statement

The thermo .RAW files of proteomic and glycoproteomic datasets have been deposited to the ProteomeXchange Consortium *via* the iProX partner repository (Ma et al., 2019) with the dataset identifier PXD024105. Supplementary data are available online at http://bib.oxfordjournals.org/.

RESULTS

Proteomic and Glycoproteomic Profiles of Plasma Samples

We analyzed a panel of 24 plasma samples using a high-throughput integrated proteomic and glycoproteomic analysis protocol described in our recent study (Zhang et al. 2019; **Figure 1** and *Methods*). The clinical information was provided in **Supplementary Table S1**. In total, we identified 346 proteins from the plasma samples that ranged from 228 to 274 proteins per individual, and identified 586 intact N-linked glycopeptides from 145 proteins, with an average of 406 N-glycopeptides per individual (**Figures 2A,B**). As shown in the Venn diagram, twenty proteins were exclusively identified in glycoproteomic profiles (**Figure 2C**).

To obtain a general overview of the cellular localizations and molecular functions of the identified proteins, the two protein groups were uploaded into ClueGo app separately as two clusters for pathway enrichment analysis and interaction network module analysis (Methods). For cellular localization analysis, a total of 39 component terms linked to nine network modules were enriched within all 366 proteins (**Figure 2C**, **Supplementary Table S2**). Most proteins were annotated as extracellular matrix, membrane attack complex, fibrinogen complex, lipoprotein particles, and blood microparticle components, or localized in the vesicle, lumen, and extracellular spaces. On the other hand, N-glycosylated proteins were preferentially found in the components of lipoprotein particle, membrane attack complex,

platelet alpha granule, blood microparticle, endoplasmic reticulum, and lysosomal lumens, suggesting a specific cellular localization of N-glycosylated proteins.

For molecular function analysis, fourteen network modules related to 217 enriched pathways were notably identified within all the proteins, including pathways of plasma lipoprotein particle remodeling, humoral immune response, fibrinolysis, regulation of cholesterol transport, positive regulation of protein secretion, and cell substrate adhesion (**Figure 3**, **Supplementary Table S2**). However, N-glycosylated proteins were exclusively enriched in pathways of plasma lipoprotein particle remodeling, complement cascade, complement and coagulation cascades, binding and uptake of ligands by scavenger receptor, serine endopeptidase activity, platelet degranulation, and insulin growth factor (IGF) regulation, many of which are known to be associated with T2D (Fateh-Moghadam et al., 2005; Fadini et al., 2014; Haywood et al., 2019). The distinct aspects of the enrichment pathways suggested that glycoproteomic analysis can capture some new clues as a complementary to proteomic analysis, and highlighted that integrative of these two analyses will enable new advances in diabetes biology, diagnostics, and therapeutics.

Signature Proteins and N-Glycopeptides for Type 2 Diabetic Patients

To better understand the proteomic changes in the plasma of T2D patients, supervised pairwise comparisons were performed to identify signature proteins and N-glycopeptides using moderated t-statistics (Methods). Comparing with HC individuals, a total of 32 and 20 differential expressed proteins were identified in T2D and PDB groups, respectively, with an overlapping of eight proteins (p-value < 0.05 and absolute log2-fold change ≥1; **Supplementary Table S3**, **Figures 4A,B**). There were 19 proteins that distinguished T2D from PDB (**Supplementary Table S3**). Among these proteins, multiple previously described (pre)diabetic biomarkers were found,
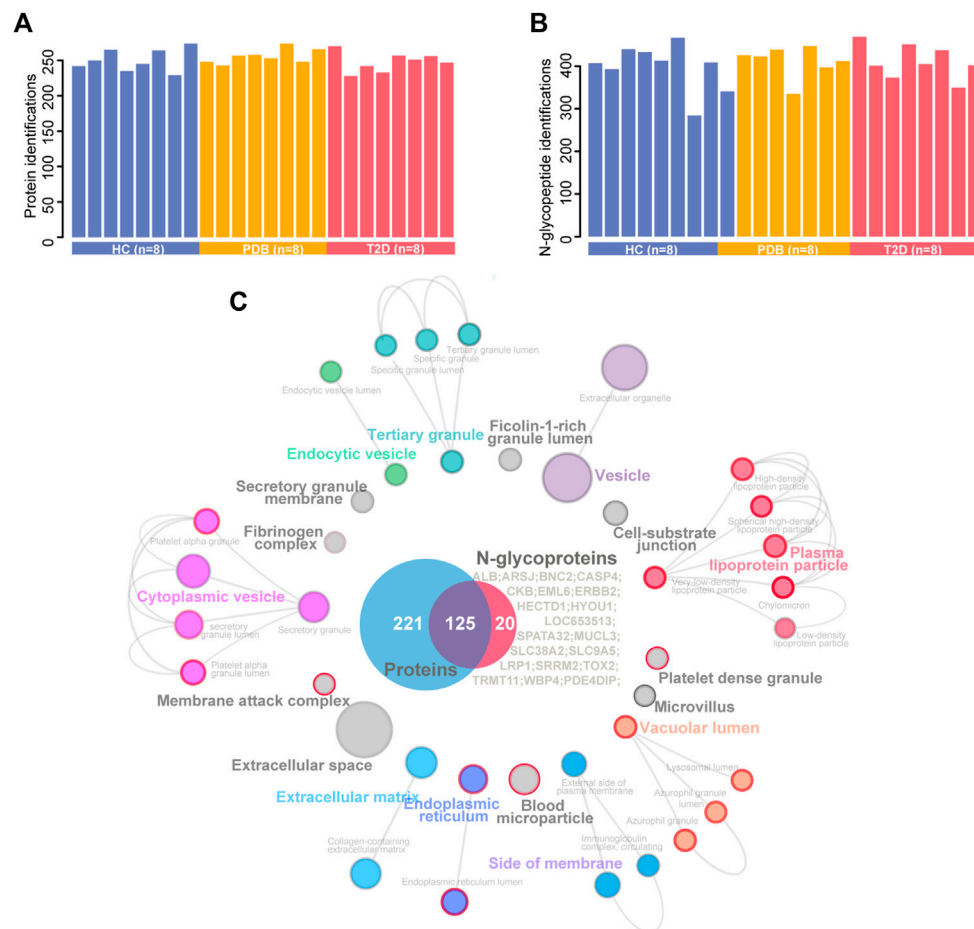
FIGURE 2 | A summary of plasma proteomic and glycoproteomic analysis of our data. **(A, B)** Number of protein (**A**) and N-glycopeptide **(B)** identifications in HC (blue), PDB (yellow), and T2D (red) samples. **(C)** Cellular localization landscape of all proteins identified in proteomic and glycoproteomic analyses. A Venn diagram of proteins identified in proteomics or glycoproteomics is shown in the center of the picture, and the proteins exclusively identified in glycoproteomics are listed on the right of the Venn diagram. Cellular component interaction network of all proteins was constructed by the ClueGO app in Cytoscape software, based on Gene Ontology (GO) Cellular Component data resource. The component terms are connected based on the kappa score. The network modules are defined using the kappa score and annotated with different colors. For each module, the most significant pathway is highlighted by a colored name label. The size of the nodes indicates the number of identified proteins with this component. The components specifically enriched in glycoproteomics are labeled with a red node border.

such as ADIPOQ, SAA1, TF, and VNN1, suggesting the robustness of our proteomic analysis (Memişoğulları and Bakan, 2004; Siitonen et al., 2011; Anderberg et al., 2015; Kang et al., 2016). Gene ontology (GO) enrichment analysis was then performed to localize these signature proteins and determine their molecular functions and biological process. As shown in **Figure 4C**, the protein alterations of T2D occurred mainly in blood microparticle, immunoglobulin complex, the external side of the plasma membrane, high-density lipoprotein particle, protein–lipid complex endocytic vesicle lumen, and vesicle lumen. Their major molecular functions included antigen binding, immunoglobulin receptor binding, serine hydrolase activity, endopeptidase activity, and glycolipid binding. These protein alterations participated in protein activation cascade, complement activation, lymphocyte-mediated immunity, phagocytosis recognition, tissue homeostasis, negative regulation of cell adhesion, and acute

inflammatory response processes, whereass the differential proteins in PDB were found mostly in blood microparticle, vesicle lumen, high-density lipoprotein particle, immunoglobulin complex, smooth endoplasmic reticulum, and protein–lipid complex (**Figure 4D**). Their main molecular functions were scavenger receptor activity, antigen binding, CAR receptor activity, serine-type endopeptidase inhibitor activity, and peptidase inhibitor activity. Differential proteins in PDB were involved in the receptor-mediated endocytosis, complement activation, and phagocytosis processes.

In glycoproteomics, 75 N-glycopeptides from 40 proteins and 67 N-glycopeptides from 30 proteins were differentially expressed in PDB subjects and T2D patients as compared to HC individuals (BH p-value < 0.01 and absolute log2-fold change ≥1), with an overlapping of 25 N-glycopeptides from 14 proteins (**Supplementary Table S4**, **Figures 5A,B**). Using the same cutoff, 68 differential expressed N-glycopeptides from 45

FIGURE 3 | Molecular function landscape of all proteins identified in proteomics and glycoproteomics. Pathway interaction network of all proteins was constructed by the ClueGO app in Cytoscape software, based on Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome data resources. The pathways are connected based on the kappa score. The functional network modules are defined using the kappa score and annotated with different colors. For each module, the most significant pathway is highlighted by a colored name label. The size of the nodes indicates the number of identified proteins with this pathway. The pathways specifically enriched in glycoproteomics are labeled with a red node border.

proteins were identified between PDB and T2D groups (**Supplementary Table S4**). Among them, 78 N-glycopeptides from 42 proteins were specifically identified from one of the three groups, many of which had been reported as biomarkers for T1D or T2D at protein levels, but not at N-glycosylation levels yet, including AZGP1, CLU, C2, C4, F2, KNG1, and SERPING1 proteins (Zhang et al., 2013; Das and Kamalden, 2014; **Figure 5C**). Gene ontology (GO) enrichment analysis of the mapped proteins showed a high concordance with the cellular localizations and molecular functions of signature N-glycopeptides between PDB and T2D. As shown in **Figures 5D,E**, the alteration N-glycopeptides were mainly identified from the proteins that localized at blood microparticle, vesicle lumen, secretory granule lumen, platelet alpha granule lumen, and protein–lipid complex components, which had molecular functions of peptidase inhibitor activity, heparin binding, serine hydrolase activity, glycosaminoglycan binding, glycoprotein binding, and enzyme inhibitor activity, and involved in protein activation cascade, negative regulation of proteolysis, complement activation, and platelet degranulation processes.

Next, we performed unsupervised PCA analysis to investigate whether the signature panels can be served as indicators to predict T2D. As shown in **Figure 5F**, patients with T2D could be separated from HC individuals, but could hardly be separated

from PDB subjects due to the overlapping at the protein level. However, a clear separation and a more tightened distribution of PCA plots were observed at N-glycoprotein levels (**Figure 5G**). In this view, glycoproteomic profiles and the signatures derived from them may be served as better predictive biomarkers for screening T2D patients. Depending on the connection of the pathogenesis of T2D development, these markers could be potential therapeutic targets in the future.

Glycosylation Alterations of Complement Activation Pathways in Type 2 Diabetes

To further investigate molecular changes of T2D, the mapped proteins of differential N-glycopeptides identified from T2D vs. HC comparison were uploaded to ClueGo app in Cytoscape software for enriched pathway–protein interaction network module analysis. Four network modules were enriched, including the proteins involved in platelet degranulation, protein activation cascade, complement activation, and peptidase inhibitor activity (**Figure 6A**). In particular, a dysregulation of N-glycosylation of proteins involved in complement activation module were found in T2D patients. Normally, the first step to initiate complement system is the activation of three pathways, including classical pathways, lectin pathways, and alterative pathways (Flyvbjerg, 2017).

FIGURE 4 | Proteomic alterations in T2D patients and PDB subjects. **(A, B)** Volcano plots of differential expressed proteins identified in T2D patients **(A)** or PDB subjects **(B)** compared to HC individuals. Red presents upregulation and blue indicates downregulation. Signature proteins were determined based on a *p*-value < 0.05 and an absolute log2-fold change ≥1. **(C, D)** GO enrichment results of the signature proteins identified in T2D **(C)** and PDB **(D)** groups.

In classical and lectin pathways, C1 complex or the MASP1 protein recruits C2 and C4 to generate C4b2a C3 convertase, and in the alternative pathway, CFB binds to the spontaneous production of C3b to form C3bBb C3 convertase, facilitating the cleavage of C3 to generate C4b2a3b or C3bBbC3b C5 convertase and then activating the complement system through membrane attack complex (MAC) proteins (Chen et al., 2014; Ghosh et al., 2015). As shown in **Figure 6B**, the N-glycopeptides NCGVN_{174}CSGDVFTALIGEIASPNYPKPYPENSR-HexNAc4Hex5-NeuAc1Fuc2 from the C1s protein, NN_{385}LTTYK-HexNAc4Hex5NeuAc2Fuc1 from the MASP1 protein, and IVLDPSGSMNIYLVLDGSDSIGASN_{285}FTGAK-HexNAc4Hex5NeuAc1Fuc2 from the CFB protein, all of which modified by fucosylated glycans, were clearly down-regulated in the plasma of T2D patients. By contrast, the N-glycopeptides TMFPN_{651}LTDVR-HexNAc4Hex5NeuAc1Fuc2–carrying fucosylated glycans from the C2 protein and GLN_{1328}VTLSSTGR-HexNAc4Hex5NeuAc2 from the C4 protein were significantly up-regulated in T2D patients, which are downstream targets of C1 complex and the MASP1 protein, as well as TVLTPATNHMGN_{85}VTFTIPANR-HexNAc2Hex7 from the C3 protein, which is the downstream target of C3 convertase. In addition, the N-glycopeptides that

were derived from C3 convertase inhibitors were of higher abundance in T2D patients than in HC individuals. These included FSLLGHASISCTVEN_{221}ETIGVWRPSPPTCEK-HexNAc4Hex5NeuAc1Fuc2–carrying fucosylated glycans, TLFCN_{64}ASK-HexNAc5Hex6NeuAc1Fuc2–carrying fucosylated glycans, and KAFITN_{81}FSMIIDGMTYPGIIK-HexNAc4Hex5NeuAc2 from C4BPA, C4BPB, and CFH proteins, respectively. Suppose that N-glycopeptides abundance is positively associated with molecular function, our results above suggested that the complement activation pathways were inhibited in T2D from the perspective of glycoproteomics, and *vice versa*. Further studies will be needed to assess the biological relevance of the findings.

DISCUSSION

Diabetes is a global health problem that affects more than 400 million peoples each year (Saeedi et al., 2019). The routine clinical tests for the diagnosis and follow-up of diabetes have certain limitations; for example, they may not be sufficiently sensitive at early disease stages (the Association A.D., 2019). Recent advances in proteomics and glycoproteomics technologies hold great potential in providing additional

FIGURE 5 | Glycoproteomic alterations in T2D patients and PDB subjects. **(A, B)** Volcano plots of differential expressed N-glycopeptides identified in PDB subjects **(A)** or T2D patients **(B)** or compared to HC individuals. Red presents upregulation and blue indicates downregulation. Signature proteins were determined based on a Benjamin–Hochberg corrected p-value < 0.01 and an absolute log2-fold change ≥1. **(C)** Heatmap depicting the relative abundance of signature N-glycopeptides. Samples are in columns and peptides are in rows. The gene names and glycan composition numbers are annotated on the left panel, and the peptide information is annotated on the right panel. **(D, E)** GO enrichment results of the mapped proteins of signature N-glycopeptides for PDB **(D)** and T2D **(E)** groups. **(F, G)** PCA plots of first two components of signature proteins **(F)** and N-glycopeptides **(G)**. The ellipse presents the 0.95 confidence intervals for each type.

insights into the disease biology and discovering new biomarkers (Vajaria and Patel, 2017; Zheng et al., 2020). Herein, we presented the global proteomic and glycoproteomic landscape in the fasting plasma of the 24

Chinese participants, and provided some clues for diagnosis and treatment of T2D.

Our integrated analysis suggested a specific cellular localization and molecular function of N-glycosylated proteins
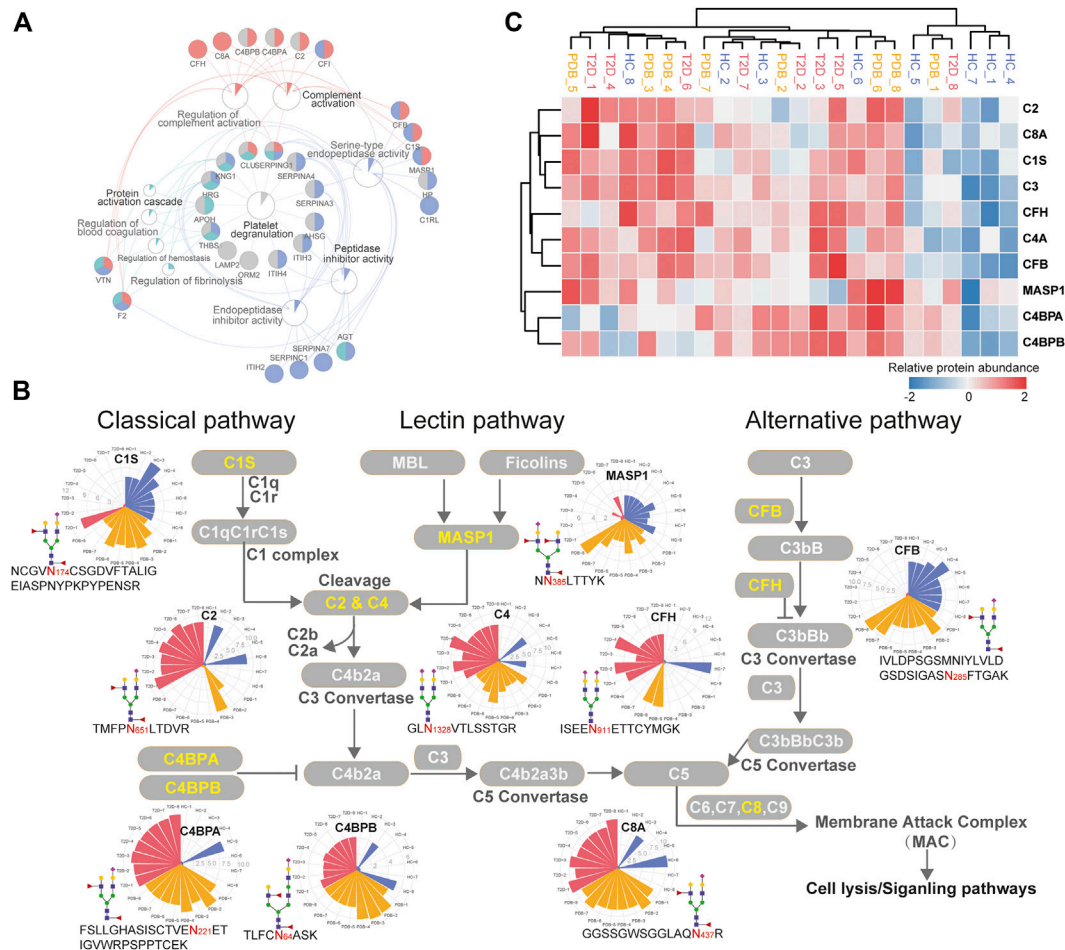
FIGURE 6 | Dysregulation of N-linked glycosylation of complement activation pathways in T2D. **(A)** Pathway and protein interaction network of the mapped proteins from T2D signature N-glycopeptides. Distinct functional modules are annotated with different colors. The node size is determined by the number of proteins identified in that pathway, and the size of colored pie chart reflects the proportion of identified proteins of that pathway. **(B)** Network diagram summarizes relevant signature N-glycopeptides and signaling cascades involved in complement activation pathways. The proteins with N-glycosylation alterations are annotated as yellow characters, and their altered glycopeptides, glycosites, glycan compositions, and N-glycopeptide abundance are displayed. The relative abundance of N-glycopeptides is depicted as radar circles. Blue, HC individuals; yellow, PDB subjects; red, T2D patients. **(C)** Heatmap depicting the corresponding expression levels of the proteins discussed in complement activation pathways.

and highlighted the fundamental discrepancy between proteomic and glycoproteomic profiles. In proteomics, a total of 32 and 20 differential expressed proteins were identified in T2D and PDB groups compared to the HC group, which were mainly localized in blood microparticle, immunoglobulin complex, protein–lipid complex, and vesicle lumen components, and involved in antigen binding, immunoglobulin receptor binding, protein activation cascade, complement activation, receptor-mediated endocytosis, and phagocytosis pathways. Within our selected markers, there are actually several proteins that have already been reported in previous studies, such as ADIPOQ, SAA1, TF, and VNN1 proteins. For instance, VNN1 is a pantetheinase that plays a crucial role in gluconeogenesis and lipogenesis, affecting multiple metabolic pathways (Chen et al., 2014; Kang et al., 2016). Clinical investigations showed that the level of VNN1 was increased in the blood of diabetic patients (Chen et al., 2014). Our results further

supported the involvement of VNN1 in the pathological process of diabetes, thus making VNN1 as an important biomarker candidate for the diagnosis of this disease. Moreover, our results agreed with previous findings that the expression of TF and ADIPOQ was decreased and the expression of SAA1 was increased in the plasma of T2D patients, which enhanced the robustness of our proteomic findings and highlighted the potential of our signature proteins serving as biomarkers for the early detection of (pre)diabetes (Memişoğulları and Bakan, 2004; Siitonen et al., 2011; Anderberg et al., 2015).

In glycoproteomics, 75 and 67 differential expressed N-glycopeptides were identified from 40 proteins in PDB subjects and 30 proteins in T2D patients, respectively. The mapped proteins of these N-glycopeptides are known to be involved in platelet degranulation, protein activation cascade, peptidase inhibitor activity, and complement activation, some of

which have been identified as potential biomarkers for T1D or T2D, including AZGP1, CLU, C2, C4, F2, KNG1, and SERPING1 (Zhang et al., 2013; Das and Kamalden, 2014). In addition, some family members of SERPING1, including SERPINA3, SERPINA4, SERPINA7, and SERPINC1, were also found to be dysregulated in T2D patients through glycoproteomics in our study, which were involved in platelet degranulation and peptidase inhibitor pathways (**Figure 6A**). Within their gene families SERPINA6, SERPINB2, and SERPINB8 have long been known as biomarkers for diabetes (Wang et al., 2008; Zhang et al., 2013). However, all the biomarkers that were reported previously were identified from the proteomic level, and the changes of their N-glycosylation are first time reported in the plasma of T2D patients by our study. The potential role of these N-glycosylation changes and their contribution to the development of T2D warrant further mechanistic investigation.

The major glycoproteomic changes of T2D patients were characterized by the dysregulation of N-glycopeptides involved in the activation of complement pathways. The complement system plays an important role in innate immune defense and humoral immunity, which is typically thought to be activated through classical, lectin, and alternative pathways, and increasing evidence suggests their pathological roles in the development of diabetes and its complications (Acosta et al., 2000; Yang et al., 2013). The classical pathway is activated by binding of the C1 complex (comprising C1q, C1r, and C1s) to antibodies, while the lectin pathway is activated by MBL-associated serine proteases (MASP1 or MASP2) after recognition of carbohydrates by mannose-binding lectin (MBL) or recognition of N-acetylglucosamine residues by ficolins, facilitating the cleavage of C2 and C4 to form C4b2b C3 convertase (Ghosh et al., 2015; Flyvbjerg, 2017). The alternative pathway is triggered through the spontaneous activation of C3, and the produced C3b forms C3bBb C3 convertase under the action of complement factor B (CFB) and complement factor D (CFD). All these pathways lead to the cleavage of C3 and converge into one pathway, eventually resulting in the formation of membrane attack complex (MAC), the main effector of complement-mediated tissue damage, which lyses, damages, or activates target cells (Ghosh et al., 2015; Flyvbjerg, 2017).

In the present study, we found that several N-glycopeptides from C1s, MASP1, and CFP proteins were clearly down-regulated in T2D patients, while N-glycopeptides from C2, C4, C4BPA, C4PBB, and CFH proteins were inversely up-regulated. According to these results, we speculate that at glycoproteomic levels, the decreased N-glycopeptides of positive regulator proteins C1s and MASP1 may hinder the cleavage of C2 and C4 to generate C4b2a C3 convertase, and the decreased N-glycopeptide of CFB may inhibit the formation of C3Bb C3 convertase, resulting in increased N-glycopeptide levels of C2, C4, and C3 proteins, which might suppress the complement activation system in one way. Meanwhile, the increased levels of the C3 convertase inhibitors C4BPA, C4BPB, and CFH would result in the failure of C3 cleavage and further suppress the complement activation system (**Figure 6B**). However, these alterations were not observed at proteomic levels (**Figure 6C**). Moreover, many immunoglobulin biomarkers associated with the complement activation pathway are found to be lost in T2D patients instantly, suggesting a suppressing status of complement system. In specific, we found

that most N-glycosylation alterations occurred in the complement pathway are modified with fucosylated glycans, indicating a potential role of this glycan modification in the development of diabetes. Overall, these results suggest that N-glycosylation dysregulation in complement activation pathways may be closely associated with the pathogenesis of T2D, thereby requiring further analyzing and exploring in future studies.

In addition to the complement system we focused on, we also observed other interesting N-glycosylation changes in other pathways. For instance, zinc-alpha-2-glycoprotein (ZAG) is a soluble protein that can be identified in plasma. As an adipokine, ZAG is mainly secreted by subcutaneous and visceral adipocytes and is involved in stimulating adipocytes to induce lipid degradation (Yang et al., 2013). It has been reported that ZAG is related to the markers of insulin resistance or obesity and may play a role in the pathogenesis of insulin resistance (IR) (Bouchara et al., 2018). ZAG index can be served as an indicator to identify IR (Qu et al., 2016). The alteration of intact N-glycopeptide may play a vital role in the development of T2D and can be used as a potential indicator. Ceruloplasmin (Cp), another one worth mentioning, is a copper-binding glycoprotein with ferroxidase activity. As a key protein participated in iron metabolism and antioxidant defense, the level of Cp during the acute phase is regulated in response to inflammation, infection, copper and iron metabolism disorders, and vascular injury (Daimon et al., 1998; Golizeh et al., 2017). It has been found that the increase in the plasma Cp level in diabetic patients may be a signal of increased oxidative stress (Cunningham et al., 1995). The progression of T2D is associated with oxidative stress, iron metabolism disorders, and inflammation (Golizeh et al., 2017). Therefore, the change of the Cp N-glycopeptide level may also predict the occurrence and development of diabetes. Briefly, the alternation of N-glycopeptide prior to protein change may help us diagnose diseases in advance.

In summary, this work provides a comprehensive plasma proteomic and glycoproteomic resource for pathogenic investigation of diabetes and identifies a set of predictive biomarkers that might enhance the efficiency and accuracy based on current existing biological indicators of (pre)diabetes. To our knowledge, the systemic dysregulation of the glycoproteomic level in the complement activation system is first time identified to be a characteristic of T2D, which sheds new light on the pathogenesis of this disease and suggests new possibilities for diagnosis and treatment. As we see it, although the number of samples presented here is too few to make robust conclusions, this work illustrates the capability of glycoproteomics which serves as complement to proteomics or metabolomics and could provide additional detailed insights into pathways and processes that drive the development of diabetes disease. It thereby presents a new direction for understanding diabetic pathogenesis and deserves more attention.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Scientific Research, Beijing Shijitan Hospital, Capital Medical University (ChiCTR 1800014301). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XF, XD, and DY directed and designed the research. MW, BM, YG, and YJ conducted the experiments. YZ, MW, BM, and MH analyzed data. YZ and MW interpreted the data in the context of diabetes biology and wrote the manuscript. YZ and ZX reviewed the manuscript and prepared tables and figures. All authors reviewed, edited, and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.677621/full#supplementary-material

REFERENCES

Acosta, J., Hettinga, J., Flückiger, R., Krumrei, N., Goldfine, A., Angarita, L., et al. (2000). Molecular Basis for a Link between Complement and the Vascular Complications of Diabetes. *Proc. Natl. Acad. Sci.* 97, 5450–5455. doi:10.1073/pnas.97.10.5450

American Diabetes Association (2019). 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. *Diabetes Care* 42, S13–S28. doi:10.2337/dc19-S002

Anderberg, R. J., Meek, R. L., Hudkins, K. L., Cooney, S. K., Alpers, C. E., Leboeuf, R. C., et al. (2015). Serum Amyloid A and Inflammation in Diabetic Kidney Disease and Podocytes. *Lab. Invest.* 95, 250–262. doi:10.1038/labinvest.2015.38

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Bouchara, A., Yi, D., Pastural, M., Granjon, S., Selag, J.-C., Laville, M., et al. (2018). Serum Levels of the Adipokine Zinc-Alpha2-Glycoprotein (ZAG) Predict Mortality in Hemodialysis Patients. *Kidney Int.* 94, 983–992. doi:10.1016/j.kint.2018.07.019

Chen, S., Zhang, W., Tang, C., Tang, X., Liu, L., and Liu, C. (2014). Vanin-1 Is a Key Activator for Hepatic Gluconeogenesis. *Diabetes* 63, 2073–2085. doi:10.2337/db13-0788

Cox, J., and Mann, M. (2008). MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-range Mass Accuracies and Proteome-wide Protein Quantification. *Nat. Biotechnol.* 26, 1367–1372. doi:10.1038/nbt.1511

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: A Database of Reactions, Pathways and Biological Processes. *Nucleic Acids Res.* 39, D691–D697. doi:10.1093/nar/gkq1018

Cunningham, J., Leffell, M., Mearkle, P., and Harmatz, P. (1995). Elevated Plasma Ceruloplasmin in Insulin-dependent Diabetes Mellitus: Evidence for Increased Oxidative Stress as a Variable Complication. *Metabolism* 44, 996–999. doi:10.1016/0026-0495(95)90095-0

Daimon, M., Susa, S., Yamatani, K., Manaka, H., Hama, K., Kimura, M., et al. (1998). Hyperglycemia is a Factor For An Increase in Serum Ceruloplasmin in Type 2 Diabetes. *Diabetes care* 21, 1525–1528. doi:10.2337/diacare.21.9.1525

Das, S., and Kamalden, T. A. (2014). "The Next Generation of Diagnostic Biomarkers for Type 2 Diabetes," in *Diabetic Cardiomyopathy* (Springer), 313–321. doi:10.1007/978-1-4614-9317-4_19

Eckhardt, B. J., Holzman, R. S., Kwan, C. K., Baghdadi, J., and Aberg, J. A. (2012). Glycated Hemoglobin A1c as Screening for Diabetes Mellitus in HIV-Infected Individuals. *AIDS Patient Care and STDs* 26, 197–201. doi:10.1089/apc.2011.0379

Fadini, G. P., Albiero, M., Millioni, R., Poncina, N., Rigato, M., Scotton, R., et al. (2014). The Molecular Signature of Impaired Diabetic Wound Healing Identifies serpinB3 as a Healing Biomarker. *Diabetologia* 57, 1947–1956. doi:10.1007/s00125-014-3300-2

Falcon, S., and Gentleman, R. (2008). "Hypergeometric Testing Used for Gene Set Enrichment Analysis," in *Bioconductor Case Studies. Use R!* (New York, NY: Springer). doi:10.1007/978-0-387-77240-0-14

Fateh-Moghadam, S., Li, Z., Ersel, S., Reuter, T., Htun, P., Plöckinger, U., et al. (2005). Platelet Degranulation Is Associated with Progression of Intima-media Thickness of the Common Carotid Artery in Patients with Diabetes Mellitus Type 2. *Atvb* 25, 1299–1303. doi:10.1161/01.ATV.0000165699.41301.c5

Federation, I. D. (2019). *IDF Diabetes Atlas-9th Edition*.

Flyvbjerg, A. (2017). The Role of the Complement System in Diabetic Nephropathy. *Nat. Rev. Nephrol.* 13, 311–318. doi:10.1038/nrneph.2017.31

Gar, C., Rottenkolber, M., Prehn, C., Adamski, J., Seissler, J., and Lechner, A. (2018). Serum and Plasma Amino Acids as Markers of Prediabetes, Insulin Resistance, and Incident Diabetes. *Crit. Rev. Clin. Lab. Sci.* 55, 21–32. doi:10.1080/10408363.2017.1414143

Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., et al. (2018). A Proteomic Landscape of Diffuse-type Gastric Cancer. *Nat. Commun.* 9, 1–16. doi:10.1038/s41467-018-03121-2

Ghosh, P., Sahoo, R., Vaidya, A., Chorev, M., and Halperin, J. A. (2015). Role of Complement and Complement Regulatory Proteins in the Complications of Diabetes. *Endocr. Rev.* 36, 272–288. doi:10.1210/er.2014-1099

Golizeh, M., Lee, K., Ilchenko, S., Ösme, A., Bena, J., Sadygov, R. G., et al. (2017). Increased Serotransferrin and Ceruloplasmin Turnover in Diet-Controlled Patients with Type 2 Diabetes. *Free Radic. Biol. Med.* 113, 461–469. doi:10.1016/j.freeradbiomed.2017.10.373

Haltiwanger, R. S., and Lowe, J. B. (2004). Role of Glycosylation in Development. *Annu. Rev. Biochem.* 73, 491–537. doi:10.1146/annurev.biochem.73.011303.074043

Haywood, N. J., Slater, T. A., Matthews, C. J., and Wheatcroft, S. B. (2019). The Insulin like Growth Factor and Binding Protein Family: Novel Therapeutic Targets in Obesity & Diabetes. *Mol. Metab.* 19, 86–96. doi:10.1016/j.molmet.2018.10.008

Hernández-Jiménez, S., García-Ulloa, A. C., Bello-Chavolla, O. Y., Aguilar-Salinas, C. A., and Kershenobich-Stalnikowitz, D. (2019). Long-term Effectiveness of a Type 2 Diabetes Comprehensive Care Program. The CAIPaDi Model. *Diabetes Res. Clin. Pract.* 151, 128–137. doi:10.1016/j.diabres.2019.04.009

Jiang, Y., Sun, A., Sun, A., Zhao, Y., Ying, W., Sun, H., et al. (2019). Proteomics Identifies New Therapeutic Targets of Early-Stage Hepatocellular Carcinoma. *Nature* 567, 257–261. doi:10.1038/s41586-019-0987-8

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG Databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. doi:10.1093/nar/30.1.42

Kang, M., Qin, W., Buya, M., Dong, X., Zheng, W., Lu, W., et al. (2016). VNN1, a Potential Biomarker for Pancreatic Cancer-Associated New-Onset Diabetes, Aggravates Paraneoplastic Islet Dysfunction by Increasing Oxidative Stress. *Cancer Lett.* 373, 241–250. doi:10.1016/j.canlet.2015.12.031

Kim, P. S., Woods, C., Georgoff, P., Crum, D., Rosenberg, A., Smith, M., et al. (2009). A1C Underestimates Glycemia in HIV Infection. *Diabetes Care* 32, 1591–1593. doi:10.2337/dc09-0177

Liu, M.-Q., Zeng, W.-F., Fang, P., Cao, W.-Q., Liu, C., Yan, G.-Q., et al. (2017). pGlyco 2.0 Enables Precision N-Glycoproteomics with Comprehensive Quality Control and One-step Mass Spectrometry for Intact Glycopeptide Identification. *Nat. Commun.* 8, 1–14. doi:10.1038/s41467-017-00535-2

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., et al. (2019). iProX: an Integrated Proteome Resource. *Nucleic Acids Res.* 47, D1211–D1217. doi:10.1093/nar/gky869

Memişoğulları, R., and Bakan, E. (2004). Levels of Ceruloplasmin, Transferrin, and Lipid Peroxidation in the Serum of Patients with Type 2 Diabetes Mellitus. *J. Diabetes Complications* 18, 193–197. doi:10.1016/S1056-8727(03)00032-1

Miura, Y., and Endo, T. (2016). Glycomics and Glycoproteomics Focused on Aging and Age-Related Diseases - Glycans as a Potential Biomarker for Physiological Alterations. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1860, 1608–1614. doi:10.1016/j.bbagen.2016.01.013

Nowak, C., Carlsson, A. C., Östgren, C. J., Nyström, F. H., Alam, M., Feldreich, T., et al. (2018). Multiplex Proteomics for Prediction of Major Cardiovascular Events in Type 2 Diabetes. *Diabetologia* 61, 1748–1757. doi:10.1007/s00125-018-4641-z

Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126, 855–867. doi:10.1016/j.cell.2006.08.019

Pan, J., Hu, Y., Sun, S., Chen, L., Schnaubelt, M., Clark, D., et al. (2020). Glycoproteomics-based Signatures for Tumor Subtyping and Clinical Outcome Prediction of High-Grade Serous Ovarian Cancer. *Nat. Commun.* 11, 1–13. doi:10.1038/s41467-020-19976-3

Pippitt, K., Li, M., and Gurgle, H. E. (2016). Diabetes Mellitus: Screening and Diagnosis. *Am. Fam. Physician* 93 (7), 103–109.

Qu, C., Zhou, X., Yang, G., Li, L., Liu, H., and Liang, Z. (2016). The Natural Logarithm of Zinc-A2-Glycoprotein/HOMA-IR Is a Better Predictor of Insulin Sensitivity Than the Product of Triglycerides and Glucose and the Other Lipid Ratios. *Cytokine* 79, 96–102. doi:10.1016/j.cyto.2015.12.024

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007

Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., et al. (2019). Global and Regional Diabetes Prevalence Estimates for 2019 and Projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th Edition. *Diabetes Res. Clin. Pract.* 157, 107843. doi:10.1016/j.diabres.2019.107843

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Sharma, A., Cox, J., Glass, J., Lee, T. J., Kodeboyina, S. K., Zhi, W., et al. (2020). Serum Glycoproteomic Alterations in Patients with Diabetic Retinopathy. *Proteomes* 8, 25. doi:10.3390/proteomes8030025

Siitonen, N., Pulkkinen, L., Lindström, J., Kolehmainen, M., Eriksson, J. G., Venojärvi, M., et al. (2011). Association of ADIPOQ Gene Variants with Body Weight, Type 2 Diabetes and Serum Adiponectin Concentrations: the Finnish Diabetes Prevention Study. *BMC Med. Genet.* 12, 1–13. doi:10.1186/1471-2350-12-5

Sohail, W., Majeed, F., and Afroz, A. (2018). Differential Proteome Analysis of Diabetes Mellitus Type 2 and its Pathophysiological Complications.

Diabetes Metab. Syndr. Clin. Res. Rev. 12, 1125–1131. doi:10.1016/j.dsx.2018.06.009

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

The R Core Team (2012). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing, 10. Available at www.R-project.org/.

The World Health Organization (2006). *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia: Report of a WHO/IDF Consultation.*

Vajaria, B. N., and Patel, P. S. (2017). Glycosylation: a Hallmark of Cancer? *Glycoconj. J.* 34, 147–156. doi:10.1007/s10719-016-9755-2

Wang, X., Jia, S., Geoffrey, R., Alemzadeh, R., Ghosh, S., and Hessner, M. J. (2008). Identification of a Molecular Signature in Human Type 1 Diabetes Mellitus Using Serum and Functional Genomics. *J. Immunol.* 180, 1929–1937. doi:10.4049/jimmunol.180.3.1929

Waugh, N., Shyangdan, D., Taylor-Phillips, S., Suri, G., and Hall, B. (2013). Screening for Type 2 Diabetes: a Short Report for the National Screening Committee. *Health Technol. Assess.* 17, 1–90. doi:10.3310/hta17350

Williams, R., Karuranga, S., Malanda, B., Saeedi, P., Basit, A., Besançon, S., et al. (2020). Global and Regional Estimates and Projections of Diabetes-Related Health Expenditure: Results from the International Diabetes Federation Diabetes Atlas, 9th Edition. *Diabetes Res. Clin. Pract.* 162, 108072. doi:10.1016/j.diabres.2020.108072

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* 6, 359–362. doi:10.1038/nmeth.1322

Yang, M., Liu, R., Li, S., Luo, Y., Zhang, Y., Zhang, L., et al. (2013). Zinc- 2-Glycoprotein Is Associated with Insulin Resistance in Humans and Is Regulated by Hyperglycemia, Hyperinsulinemia, or Liraglutide Administration: Cross-Sectional and Interventional Studies in normal Subjects, Insulin-Resistant Subjects, and Subjects with Newly Diagnosed Diabetes. *Diabetes Care* 36, 1074–1082. doi:10.2337/dc12-0940

Zhang, Q., Fillmore, T. L., Schepmoes, A. A., Clauss, T. R. W., Gritsenko, M. A., Mueller, P. W., et al. (2013). Serum Proteomics Reveals Systemic Dysregulation of Innate Immunity in Type 1 Diabetes. *J. Exp. Med.* 210, 191–203. doi:10.1084/jem.20111843

Zhang, Y., Zhao, W., Zhao, Y., Mao, Y., Su, T., Zhong, Y., et al. (2019). Comparative Glycoproteomic Profiling of Human Body Fluid between Healthy Controls and Patients with Papillary Thyroid Carcinoma. *J. Proteome Res.* 19, 2539–2552. doi:10.1021/acs.jproteome.9b00672

Zhao, D., Shen, L., Wei, Y., Xie, J., Chen, S., Liang, Y., et al. (2017). Identification of Candidate Biomarkers for the Prediction of Gestational Diabetes Mellitus in the Early Stages of Pregnancy Using iTRAQ Quantitative Proteomics. *Prot. Clin. Appl.* 11, 1600152. doi:10.1002/prca.201600152

Zheng, Z., Ao, X., Li, P., Lian, Z., Jiang, T., Zhang, Z., et al. (2020). CRLF1 Is a Key Regulator in the Ligamentum Flavum Hypertrophy. *Front. Cel Dev. Biol.* 8, 858. doi:10.3389/fcell.2020.00858

GlycanGUI: Automated Glycan Annotation and Quantification Using Glucose Unit Index

Rui Zhang[1], Wenjing Peng[2], Sakshi Gautam[2], Yifan Huang[2], Yehia Mechref[2] and Haixu Tang[1]*

[1]Department of Computer Science, Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, Bloomington, IN, United States, [2]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, United States

The retention time provides critical information for glycan annotation and quantification from the Liquid Chromatography Mass Spectrometry (LC-MS) data. However, the variation of the precise retention time of glycans is highly dependent on the experimental conditions such as the specific separating columns, MS instruments and/ or the buffer used. This variation hampers the exploitation of retention time for the glycan annotation from LC-MS data, especially when inter-laboratory data are compared. To incorporate the retention time of glycan across experiments, Glucose Unit Index (GUI) can be computed using the dextrin ladder as internal standard. The retention time of glycans are then calibrated with respect to glucose units derived from dextrin ladders. Despite the successful application of the GUI approach, the manual calibration process is quite tedious and often error prone. In this work, we present a standalone software tool GlycanGUI, with a graphic user interface to automatically carry out the GUI-based glycan annotation/ quantification and subsequent data analysis. When tested on experimental data, GlycanGUI reported accurate GUI values compared with manual calibration, and thus is ready to be used for automated glycan annotation and quantification using GUI.

Keywords: GUI, glycan, annotation, quantification, mass spectrometry

1 INTRODUCTION

Glycosylation is a post-translational modification that plays critical roles in important biological processes such as immune response, cellular differentiation/adhesion and host-pathogen interactions (Varki et al., 2009). The aberrant alteration of glycan structure is implicit with malfunction of cells and possesses potential significance for the medical diagnosis of complex human diseases such as cancer (Ohtsubo and Marth, 2006; Pinho and Reis, 2015; Stowell et al., 2015). The Liquid Chromatography coupled Mass Spectrometry (LC-MS) is one of the most widely used techniques for glycan analysis due to its high sensitivity and throughput (Pabst and Altmann, 2011). For a chromatography, the elution order of glycans is consistent and reproducible among experiments. Hence, the retention time is often used to assess the separation of glycans and to assist glycan identification (Pabst and Altmann, 2011). However, the precise retention time of a specific glycan may vary widely depending on the experimental conditions such as the separating columns, MS instruments and/or the buffer used. This variation hampers the interpretation of LC-MS/MS data, especially when inter-laboratory data are compared without any calibrations.

To address this issue, the Glucose Unit Index (GUI) can be utilized to normalize the retention time that eliminates the variation (Mellis and Baenziger, 1981). This method employs the dextrin

ladder as the internal standard and measures the retention time of glycans with respect to the glucose units derived from the dextrin ladder. The retention time is assigned to a glucose unit (GU) value, which serves as a calibration of different experiments. The GUI approach was proposed for glycomics (Campbell et al., 2008; Stockmann et al., 2013; Abrahams et al., 2018) and has been extensively used for the normalization of retention time of glycans (Ashwood et al., 2020; Gautam et al., 2020; Fabini et al., 2001; Van den Steen et al., 2006; Royle et al., 2008). In our previous study, we examined the use of permethylated dextrin for the annotation of permethylated N-glycans and their isomers derived from standard glycoproteins and human blood serum (Gautam et al., 2020). The calibrated GUI was proved to be reproducible across inter- and intra-laboratory analyses (Gautam et al., 2020).

Despite the successful application of the GUI approach, the calibration process is quite tedious, which requires manual data processing such as peak finding and curve fitting (Ashwood et al., 2019). In this paper, we present a standalone software tool GlycanGUI to automate the whole calibration process. GlycanGUI offers a graphic user interface for users to choose logarithmic or polynomial fitting that will be used in the automatic calibration of the retention time of glycan ions. Furthermore, GlycanGUI implemented the computational procedures improved from those previously developed in GlycoHybridSeq (Zhang et al., 2021) for automated glycan annotation and label-free quantification based on the total peak area of glycan ions with various charges and abducts. Therefore, GlycanGUI is ready to be used for large-scale comparative analyses of glycan abundances across many glycomic samples.

2 METHOD

2.1 Experimental Data

The data were obtained from the dextrin spiked human blood serum using LC-MS according to the previously published experimental protocols (Gautam et al., 2020). Briefly, 1 μg of dextrin was spiked as an internal standard in initial 10 μl serum before permethylation. For each injection, the released N-glycans from 1 μl of the initialserum sample were resuspended in 6 μl of 20% ACN and 0.1%FA. For C18 columns, solution A was 98% water, 2% ACN, and 0.1% FA and solution B was 100% ACN and 0.1% FA. The gradient started at 20% solution B and increased to 42% in 11 min. After 48 min, it reached 55% and increased to 90% in 1 min. The organic phase remained at 90% for 54 min and decreased to 20% for 6 min. The LC-MS data were acquired by using LTQ Orbitrap Velos (Thermo Scientific) instrument.

2.2 GlycoGUI Software

The software GlycoGUI was implemented in C# using the WPF framework for graphic user interface. After GlycoGUI reads the input LC-MS data (in thermo. raw format) using the MSFileReader library, a user can calibrate the retention time of any annotated glycans into the corresponding GUI value. The abundances (peak areas) of the glycans are computed over

retention time and specifically for the detected retention range of major isomers. The source code of GlycanGUI can be found on Github at https://github.com/ruizhang84/GlycanGUIApp.

2.3 Glucose Units Identification

The ions corresponding to *Glucose Units* are extracted directly from the full MS spectra based on their expected mass-to-charge-ratio (m/z). Due to the instrumental noise and overlapping peaks, multiple putative glucose units (of different m/z) may be extracted at a specific retention time. To correctly label glucose units and avoid mis-identified glucose units, a dynamic programming algorithm is implemented to obtain the most likely sequence of glucose units. Briefly, a score is computed recursively for all glucose units at each retention time where a putative glucose unit (i.e., dextrin) is observed, which is based on the intensity of the peak matching the dextrin:

$$score[i][u] = max\big(score[i-1][u], max_{j\in[2,u]}\big(score[i-1][j]\big)\big) \tag{1}$$

where i is the index for the observed retention time, u is the index of the glucose units. This equation is derived based on the fact that the higher glucose units (dextrin) always elute at a latter time than the lower units, as well as the assumption that the true peaks of the glucose units are likely more intensive than those false peaks. After identifying peaks corresponding to the glucose units in each experimental MS spectrum, the retention time with highest peak intensity for each glucose unit is used for curve fitting.

2.4 Curve Fitting

To calibrate the retention time of ion species into a glucose unit index (GUI), we adopt a polynomial regression,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_n x_i^n \tag{2}$$

where for an ion species ion i, y_i is the (target) glucose unit index, x_i is its retention time, and β_j is jth coefficient in the polynomial regression. For a logarithmic fitting, the regression can be solved similarly by converting x_i into $log(x_i)$. The linear regression can be expressed in terms of matrix multiplication for a total of m glucose units (i.e., dextrin ladders),

$$\begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & x_2^3 & \ldots & x_2^m \\ \ldots & & & & & \\ 1 & x_n & x_n^2 & x_n^3 & \ldots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_m \end{bmatrix} \tag{3}$$

or simply in the vector form as,

$$\overrightarrow{y} = X\overrightarrow{\beta} \tag{4}$$

Using the least square estimation, the coefficients are computed as,

$$\overrightarrow{\beta} = \big(X^T X\big)^{-1} X \overrightarrow{y} \tag{5}$$

where X^{-1} is the inverse of the matrix X, which can be computed by using the Gauss-Jordan method (Althoen and Mclaughlin, 1987). Once the coefficients are estimated using the elution time

and the (target) GUI values of the dextrin ladders that are injected and observed in a LC-MS experiment, they can be used in the polynomial regression for calibrating the elution time of other ion species of interests (e.g., those annotated as glycans; see below), into GUI values.

2.5 Glycan Annotation and Label-free Quantification

We adopted the algorithm used for glycan annotation similar to one implemented in GlycoHybridSeq (Zhang et al., 2021). Briefly, the potential N-glycans are pre-computed with up to a certain maximum number of monosaccharide residues (by default, #HexNAc ≤ 12, #Hex ≤ 12, #Fuc ≤ 5, and #NeuAc ≤ 4) according to the biosynthesis rule (Zhang et al., 2021). The theoretical isotopic distribution of these N-glycans are subsequently derived using the BRAIN algorithm (Dittwald et al., 2013) based on their chemical formulas. To search for a particular N-glycan, the accurate mass value corresponding to its most abundant isotopic ion is searched against the peaks in a full MS spectrum with a given mass tolerance. The most abundant isotopic ion is expected to be most intensive and thus is most likely to be observed in an experimental spectrum if the N-glycan is indeed present in the sample. We adopted the same spectrum pre-processing procedure including charge derivation and peak picking as that used in GlycoHybridSeq. To speed up searching of peaks within a given mass tolerance, a bucket search algorithm is employed as employed in GlycoHybridSeq. After the most abundant isotopic ion is matched, we extend the peak matching of other isotopic ions to generate a matched isotopic envelope, which is then scored against the theoretical isotopic distribution as Pearson Coefficient (r)

$$score_r = \frac{\sum (t_i - \bar{t}) \sum (o_i - \bar{o})}{\sqrt{\sum (t_i - \bar{t})^2 (o_i - \bar{o})^2}} \qquad (6)$$

where t_i and o_i are the intensities of the matched theoretical and observed isotopic ions, and $\bar{t}$ ad $\bar{o}$ are the average intensities of all peaks in the isotopic envelope. Pearson Coefficient measures the similarity of two isotopic clusters, which offer similar measures such as Cosine Similarity, but are invariant to linear transformation of data (Van Dongen and Enright, 2012). A user-defined cutoff is used (default 0.9) to reduce likely false N-glycan annotation.

To assess the abundance of a glycan, the extracted-ion chromatogram (XIC) of glycan ion is generated by GlycanGUI by searching theoretical m/z over all full-MS spectra (as described above) and measuring the corresponding peak area at each retention time. The peak area is computed from the combined intensity of top three isotope peaks, which is adopted from the concept of top three-isotopes quantification (3TIQ) algorithm (Park et al., 2016). It is reported that considering top three isotope peaks as in 3TIQ algorithm offers more sensitive results with better signal-to-noise ratio (Park et al., 2016). To detect individual XIC of glycan isomers (that have the same composition and thus theoretical m/z), the XIC peak detection algorithm reported by Aoshima et al. Aoshima et al. (2014) is adopted with customized

modification to suit for glycan detection. Briefly, the point of highest intensity (i.e., apex) is first located by searching the local maximum over XIC peaks. The leftmost and rightmost neighbor of the apex are then tracked starting from the center of apex, which are defined as the peak boundary higher than (or equal to) a given cutoff (by default 50%) of apex intensity. The left and right bounds of XIC are discovered by extending the leftmost and rightmost neighbor using a local minimum algorithm. The peak area of the detected glycan isomers were summed

$$Area = \sum_{left < i < right} A_i \qquad (7)$$

where A_i is the peak area over the detected range within left and right bound in XIC.

3 RESULTS AND DISCUSSION

3.1 GUI Calibration

GlycanGUI provides a graphic user interface that takes mass spectra data containing dextran ladder as the internal standard (as shown in **Supplementary Figure S1**). It determines Glucose Unit (GU) values by fitting to 3rd order polynomial (by default) or logarithmic function (**Supplementary Figure S1**), which allows for calibrating the retention time to the corresponding Glucose Unit Index (GUI). This calibration process involves the determination of retention time of dextrin ladders with different number of GUs (i.e., GU 2–12) followed by the polynomial (or logarithmic) curve fitting.

To assign the peaks of the dextrin ladders in LC-MS data, each peak in a full MS spectrum was first compared against the theoretical mass-to-charge-ratios of dextrin ladders with putative ion charges (by default up to +3) and abducts (by default the proton) using a user-defined mass tolerance. For each dextrin of a specific GU, only the most abundant isotopic ion among all matched peaks in the spectrum is recorded, along with its GU, retention time and the maximum intensity of matched peaks. However, sometimes multiple dextrin ladders with different number of GUs may be extracted from the same spectrum (**Figure 1** for an example). We employed the dynamic programming algorithm as described in the Methods section to resolve the ambiguity. As shown in **Figure 2**, the retention time of the extracted peaks are consistent with the manual assignment in the previous study (Gautam et al., 2020) (**Supplementary Figure S2**), indicating GlycoGUI assigned the internal standards (dextrin ladders) accurately. Notably, for dextrins with large GUI (e.g., for *GUI* > 10), their peak intensities become so low that are hard to be distinguished from the baseline peaks, which is difficult for manual assignment. On the other hand, the automated peak assignment by GlycoGUI can avoid such potential errors in manual assignment.

To calibrate the retention time into GUI, the value of GUs (i.e., GUIs) are regressed against the retention times of the dextrin ladders (corresponding to ithe most abundant isotopic peak) using a polynomial function as described in Method. **Figure 3** shows the standard plot of the GUI against the retention times for the dextrin ladders observed in a human serum sample. The coefficient of determination (R^2) is greater than 0.99 from the

FIGURE 1 | The peak assignment of dextrin ladders in a full MS spectrum.

FIGURE 2 | The elution profile of dextrin ladders extracted by using peak matching and the dynamic programming algorithm.

FIGURE 3 | The curve fitting of GUI as a polynomial function of the retention time on dextrin ladders with different number of GUs.
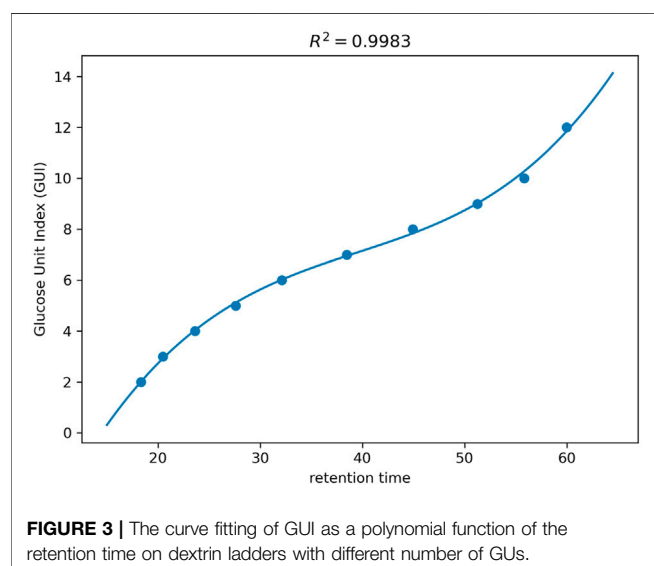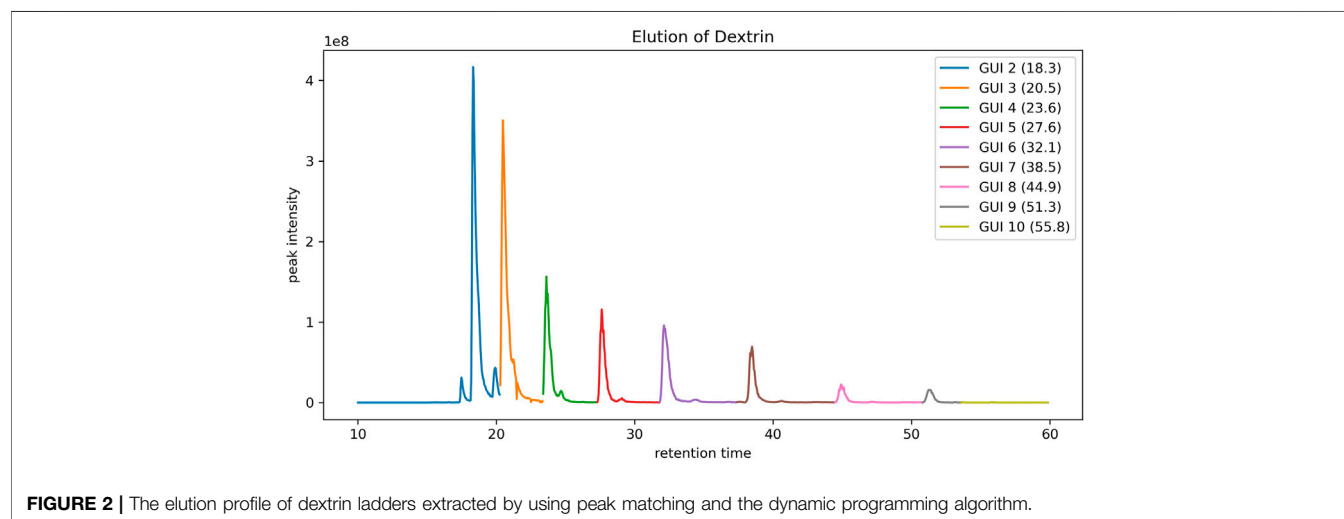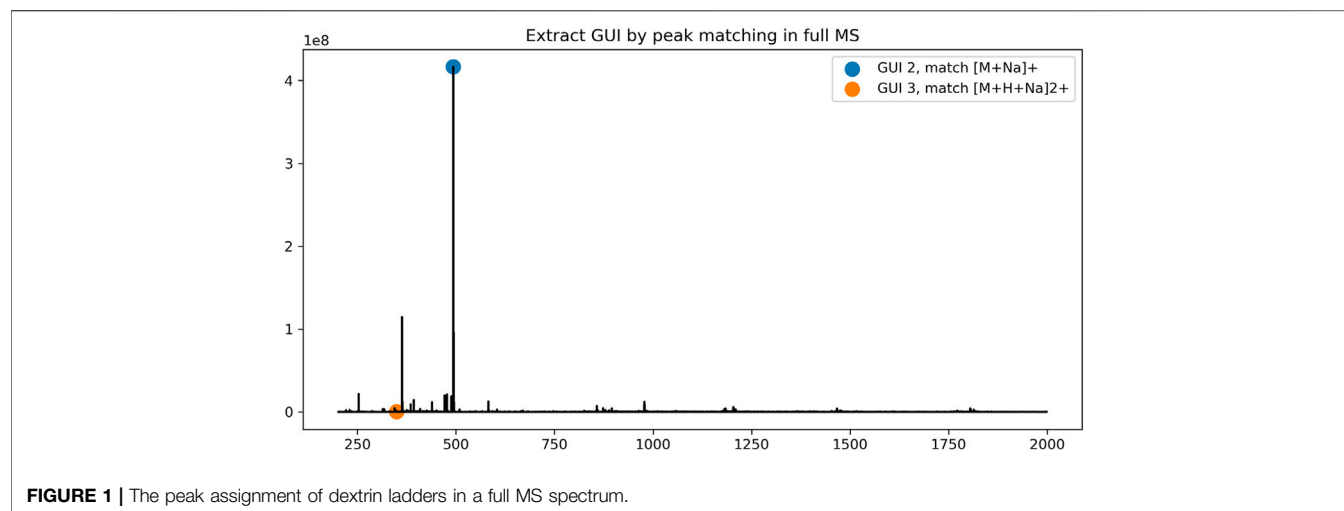
curve fitting, indicating the high correlation between the retention time and GUI. Alternatively, a logarithmic function can also be used for fitting the standard curve of GUIs against the retention time (as shown in **Supplementary Figure S3**). We implemented both curve fitting functionalities, since our previous study showed the polynomial fitting gave the best fitting results (Gautam et al., 2020), while Ashwood et al. concluded that the logarithmic fitting best described the retention profile of the dextran ladders (Ashwood et al., 2019). In our experiments, both methods received high correlation ($R^2 > 0.99$), which implies that the calibration can achieve satisfactory results by using either method, as long as the same method is consistently used for the calibrations of all samples being compared. For the results presented below, the polynomial fitting is used unless otherwise stated.

3.2 Glycan Annotation

To facilitate glycan identifications, glycanGUI provides a graphic user interface that allows user to annotate glycans with GUI (as
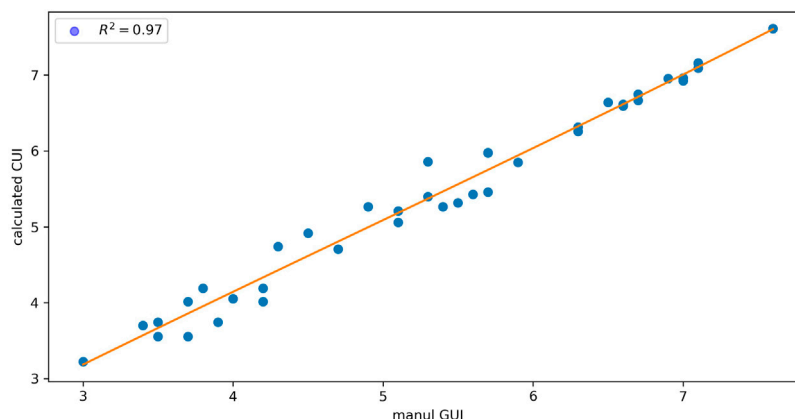
FIGURE 4 | The comparison of GUI values calibrated by using GlycanGUI and by manual calibration for the N-glycans in human serum glycomic data.
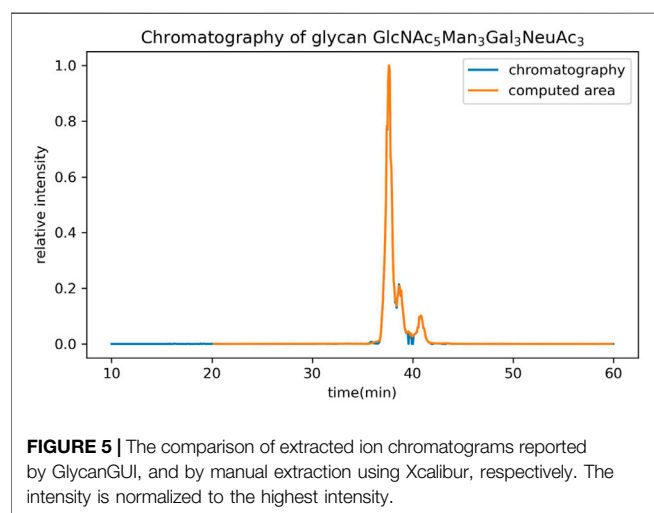
FIGURE 5 | The comparison of extracted ion chromatograms reported by GlycanGUI, and by manual extraction using Xcalibur, respectively. The intensity is normalized to the highest intensity.
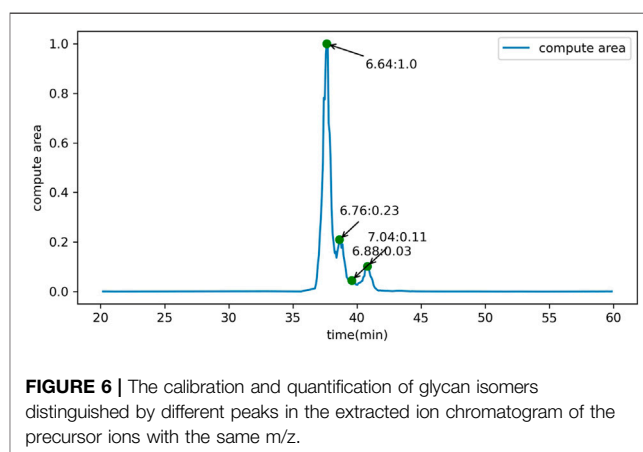
FIGURE 6 | The calibration and quantification of glycan isomers distinguished by different peaks in the extracted ion chromatogram of the precursor ions with the same m/z.
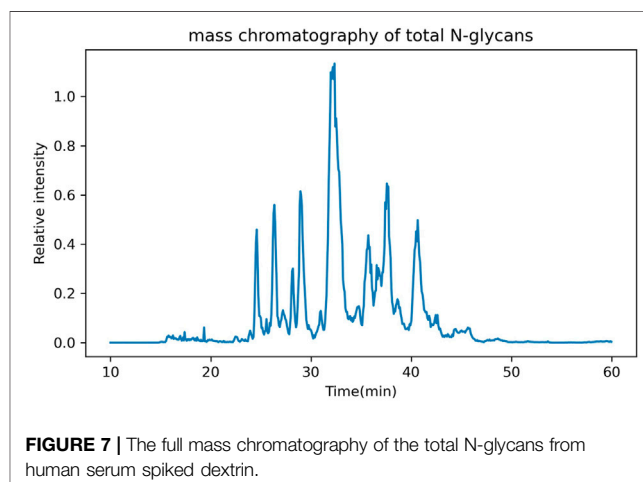
FIGURE 7 | The full mass chromatography of the total N-glycans from human serum spiked dextrin.

shown in **Supplementary Figure S4**). With reported GUI of annotated glycans, the annotated glycans can be easily verified according to calibrated retention time (Ashwood et al., 2019; Gautam et al., 2020). Moreover, the validation of annotated glycans can be conducted with GUI library of known glycans (Gautam et al., 2020) to filter putative false glycan annotations.

To evaluate the glycan annotation and GUI calibration by GlycanGUI, the GUI values reported by GlycanGUI for the annotated N-glycans in human serum glycomic data were compared with our manual calibrations reported previously (Gautam et al., 2020). As shown in **Figure 4**, the calibrated GUI values by GlycanGUI is highly correlated with the manually calibrated values ($R^2 = 0.97$), indicating the automated glycan annotation and GUI calibration by GlycanGUI achieved satisfactory results. Notably, a total of 275 N-glycans were annotated by GlycanGUI, while only 51 of them were manually annotated (Gautam et al., 2020); the above comparison is performed on the N-glycans that were annotated by both methods. Given the 1,046 full MS spectra acquired in this LC-MS study, the full annotation of all N-glycans

involves tremendous manual efforts. Hence, GlycanGUI serves as a reliable software tool to automate the data processing that will drastically reduces the time for data analyses.

In addition to the GUI values, GlycanGUI performs label-free quantification of the glycans with calibrated GUI values based on

the peak areas in the extracted ion chromatograms (XIC). As shown in **Figure 5**, the XIC reported by GlycanGUI is highly correlated ($R^2 = 0.9995$) to that obtained manually using Xcalibur (see **Supplementary Figure S6**), indicating quantitative measurement by GlycanGUI is reliable. Moreover, GlycanGUI can automatically distinguish the glycan isomers based on the peaks in the XIC of the precursor ions with the same m/z (within the mass tolerance). The peak areas and the corresponding GUI values of the isomers are reported, respectively, as shown in **Figure 6**. A full mass chromatography of all the N-glycans automatically identified from human serum spiked dextrin is shown as in **Figure 7**. A large portion of N-glycans (60.4%) elutes during 30–40 min, while several major elution peaks are observed over 20–50 min. It is worth to mention that identified N-glycans may requires human validation, especially when interference are present. To improve glycan identification, tandem mass spectrometry (MS/MS) can be performed along with full MS so that N-glycans can be identified according to ion fragments. This improvement will be pursued in our future endeavour.

4 CONCLUSION

Glycan annotation and label-free quantification from LC-MS data often involve time consuming manual efforts. Here, we report a open-source software tool GlycanGUI for automated calibration of the elution time of glycans into GUI values and the quantification of the corresponding ions based on peak intensities. The source code of GlycanGUI is released on Github at https://github.com/ruizhang84/GlycanGUIApp. We note that although GlycanGUI has been tested extensively on glycomic datasets, the glycan annotation results should be used with caution. To produce the more reliable results for glycan annotation, GlycanGUI is preferred to be used in combination with a library of glycan GUIs (Gautam et al., 2020) or with glycan annotation software tools (Yu et al., 2013; Hu et al., 2015). In the future, we plan to implement glycan identification algorithm in GlycanGUI, for example, by exploiting the fragment ion patterns in tandem mass spectra.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

RZ and HT designed the study; RZ, SG, WP and YH investigated the study; RZ and HT drafted the manuscript; RZ, SG, WP, YH, YM, and HT reviewed and edited the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.707382/full#supplementary-material

REFERENCES

Abrahams, J. L., Campbell, M. P., and Packer, N. H. (2018). Building a Pgc-Lc-Ms N-Glycan Retention Library and Elution Mapping Resource. *Glycoconj J.* 35, 15–29. doi:10.1007/s10719-017-9793-4

Althoen, S. C., and Mclaughlin, R. (1987). Gauss-jordan Reduction: A Brief History. *The Am. Math. monthly* 94, 130–142. doi:10.2307/2322413

Aoshima, K., Takahashi, K., Ikawa, M., Kimura, T., Fukuda, M., Tanaka, S., et al. (2014). A Simple Peak Detection and Label-free Quantitation Algorithm for Chromatography-Mass Spectrometry. *BMC bioinformatics* 15, 1–14. doi:10.1186/s12859-014-0376-0

Ashwood, C., Pratt, B., MacLean, B. X., Gundry, R. L., and Packer, N. H. (2019). Standardization of Pgc-Lc-Ms-Based Glycomics for Sample Specific Glycotyping. *Analyst* 144, 3601–3612. doi:10.1039/c9an00486f

Ashwood, C., Waas, M., Weerasekera, R., and Gundry, R. L. (2020). Reference Glycan Structure Libraries of Primary Human Cardiomyocytes and Pluripotent Stem Cell-Derived Cardiomyocytes Reveal Cell-type and Culture Stage-specific Glycan Phenotypes. *J. Mol. Cell Cardiol.* 139, 33–46. doi:10.1016/j.yjmcc.2019.12.012

Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M. (2008). Glycobase and Autogu: Tools for Hplc-Based Glycan Analysis. *Bioinformatics* 24, 1214–1216. doi:10.1093/bioinformatics/btn090

Dittwald, P., Claesen, J., Burzykowski, T., Valkenborg, D., and Gambin, A. (2013). Brain: a Universal Tool for High-Throughput Calculations of the Isotopic Distribution for Mass Spectrometry. *Anal. Chem.* 85, 1991–1994. doi:10.1021/ac303439m

Fabini, G., Freilinger, A., Altmann, F., and Wilson, I. B. H. (2001). Identification of Core α1,3-Fucosylated Glycans and Cloning of the Requisite Fucosyltransferase cDNA from *Drosophila melanogaster*. *J. Biol. Chem.* 276, 28058–28067. doi:10.1074/jbc.m100573200

Gautam, S., Peng, W., Cho, B. G., Huang, Y., Banazadeh, A., Yu, A., et al. (2020). Glucose Unit index (Gui) of Permethylated Glycans for Effective Identification of Glycans and Glycan Isomers. *Analyst* 145, 6656–6667. doi:10.1039/d0an00314j

Hu, Y., Zhou, S., Yu, C.-Y., Tang, H., and Mechref, Y. (2015). Automated Annotation and Quantitation of Glycans by Liquid Chromatography/electrospray Ionization Mass Spectrometric Analysis Using the Multiglycan-Esi Computational Tool. *Rapid Commun. Mass. Spectrom.* 29, 135–142. doi:10.1002/rcm.7093

Mellis, S. J., and Baenziger, J. U. (1981). Separation of Neutral Oligosaccharides by High-Performance Liquid Chromatography. *Anal. Biochem.* 114, 276–280. doi:10.1016/0003-2697(81)90480-2

Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126, 855–867. doi:10.1016/j.cell.2006.08.019

Pabst, M., and Altmann, F. (2011). Glycan Analysis by Modern Instrumental Methods. *Proteomics* 11, 631–643. doi:10.1002/pmic.201000517

Park, G. W., Kim, J. Y., Hwang, H., Lee, J. Y., Ahn, Y. H., Lee, H. K., et al. (2016). Integrated Glycoproteome Analyzer (I-gpa) for Automated Identification and Quantitation of Site-specific N-Glycosylation. *Sci. Rep.* 6, 1–12. doi:10.1038/srep21175

Pinho, S. S., and Reis, C. A. (2015). Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* 15, 540–555. doi:10.1038/nrc3982

Royle, L., Campbell, M. P., Radcliffe, C. M., White, D. M., Harvey, D. J., Abrahams, J. L., et al. (2008). Hplc-based Analysis of Serum N-Glycans on a 96-well Plate Platform with Dedicated Database Software. *Anal. Biochem.* 376, 1–12. doi:10.1016/j.ab.2007.12.012

Stöckmann, H., Adamczyk, B., Hayes, J., and Rudd, P. M. (2013). Automated, High-Throughput Igg-Antibody Glycoprofiling Platform. *Anal. Chem.* 85, 8841–8849. doi:10.1021/ac402068r

Stowell, S. R., Ju, T., and Cummings, R. D. (2015). Protein Glycosylation in Cancer. *Annu. Rev. Pathol. Mech. Dis.* 10, 473–510. doi:10.1146/annurev-pathol-012414-040438

Van den Steen, P. E., Van Aelst, I., Hvidberg, V., Piccard, H., Fiten, P., Jacobsen, C., et al. (2006). The Hemopexin and O-Glycosylated Domains Tune Gelatinase B/mmp-9 Bioavailability via Inhibition and Binding to Cargo Receptors. *J. Biol. Chem.* 281, 18626–18637. doi:10.1074/jbc.m512308200

Van Dongen, S., and Enright, A. J. (2012). *Metric Distances Derived from Cosine Similarity and pearson and spearman Correlations.*arXiv preprint arXiv: 1208.3145

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., et al. (2009). *Essentials of glycobiology.* 3rd Edn, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. Chapter 9

Yu, C.-Y., Mayampurath, A., Hu, Y., Zhou, S., Mechref, Y., and Tang, H. (2013). Automated Annotation and Quantification of Glycans Using Liquid Chromatography-Mass Spectrometry. *Bioinformatics* 29, 1706–1707. doi:10.1093/bioinformatics/btt190

Zhang, R., Zhu, J., Lubman, D., Mechref, Y., and Tang, H. (2021). Glycohybridseq: Automated Identification of N-Linked Glycopeptides Using Electron Transfer/ high-Energy Collision Dissociation (Ethcd). *J. Proteome Res.* 20, 3345–3352. doi:10.1021/acs.jproteome.1c00245

Site-Specific N- and O-Glycosylation Analysis of Human Plasma Fibronectin

Ding Liu[1], Shuaishuai Wang[1], Junping Zhang[2], Weidong Xiao[2], Carol H. Miao[3], Barbara A. Konkle[4], Xiu-Feng Wan[5,6,7,8] and Lei Li[1]*

[1]Department of Chemistry, Georgia State University, Atlanta, GA, United States, [2]School of Medicine, Indiana University, Indianapolis, IN, United States, [3]Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, WA, United States, [4]Bloodworks Northwest, Seattle, WA, United States, [5]Center for Influenza and Emerging Infectious Diseases, University of Missouri, Columbia, MO, United States, [6]Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, MO, United States, [7]Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, [8]Department of Electrical Engineering & Computer Science, College of Engineering, University of Missouri, Columbia, MO, United States

Human plasma fibronectin is an adhesive protein that plays a crucial role in wound healing. Many studies had indicated that glycans might mediate the expression and functions of fibronectin, yet a comprehensive understanding of its glycosylation is still missing. Here, we performed a comprehensive N- and O-glycosylation mapping of human plasma fibronectin and quantified the occurrence of each glycoform in a site-specific manner. Intact N-glycopeptides were enriched by zwitterionic hydrophilic interaction chromatography, and N-glycosite sites were localized by the ^{18}O-labeling method. O-glycopeptide enrichment and O-glycosite identification were achieved by an enzyme-assisted site-specific extraction method. An RP–LC–MS/MS system functionalized with collision-induced dissociation and stepped normalized collision energy (sNCE)-HCD tandem mass was applied to analyze the glycoforms of fibronectin. A total of 6 N-glycosites and 53 O-glycosites were identified, which were occupied by 38 N-glycoforms and 16 O-glycoforms, respectively. Furthermore, 77.31% of N-glycans were sialylated, and O-glycosylation was dominated by the sialyl-T antigen. These site-specific glycosylation patterns on human fibronectin can facilitate functional analyses of fibronectin and therapeutics development.

Keywords: fibronectin, glycosylation, mass spectrometer, operator, stepped normalized collision energy

OPEN ACCESS

INTRODUCTION

As a major post-translational modification (PTM) of proteins, glycosylation has been reported to play critical roles in protein folding, stability, macromolecular interactions, functions, and activity (Mitra et al., 2003; Shental-Bechor and Levy, 2008). Alteration of glycosylation can also affect the immunogenicity of therapeutic proteins, as demonstrated in the case of interferon and others (Kuriakose et al., 2016). N-linked glycosylation and O-linked glycosylation are two major types of protein glycosylation (Varki et al., 2015). The most common O-linked glycosylation refers to the attachment of an α-linked GalNAc residue (or extended structures) to serine (Ser) or threonine (Thr) residues of protein by an O-glycosidic bond, whereas N-linked glycosylation refers to the attachment of a β-linked N-glycan to asparagine (Asn) residues within a consensus peptide sequence of Asn-Xxx (not Pro)-Ser/Thr *via* an N-glycosidic bond.

Human plasma fibronectin is a large glycoprotein that plays a crucial role in wound healing (Furie and Furie, 2005; Patten and Wang, 2020). It is a dimer consisting of two nearly identical monomers

linked covalently at their C-termini by a pair of disulfide bonds. Each monomer has an approximate molecular weight of ~250 kDa, consisting of a linear arrangement of three types (types I, II, and III) of repeating units (Erickson, 2002). Plasma fibronectin is a substrate of thrombin-activated coagulation factor XIII (FXIIIa, plasma transglutaminase), which can also be crosslinked to fibrin leading to structural alterations of the fibrin network (Jara et al., 2020). In the clot retraction process, fibronectin may interact with platelet $\alpha_{IIb}\beta_3$, thereby regulating the interaction between platelets and fibrin (Huang et al., 2019). Adhesion of circulating platelets to matrix proteins including fibronectin and to fibronectin–fibrin clot can activate platelets, leading to the formation of more platelet thrombi and enhancement of platelet cohesion (Furie and Furie, 2005).

The influence of glycosylation on fibronectin has attracted continuous attention among immunologists (Sano et al., 2008; Freire-De-Lima et al., 2011; Park et al., 2011; Hsiao et al., 2017). It is reported that glycosylation can significantly affect the ligand recognition of fibronectin, suggesting a modulation role of glycans (Sano et al., 2008). O-GalNAc glycosylation of fibronectin is found to be associated with the epithelial–mesenchymal transition (EMT) process, where an O-glycan at a specific Thr of fibronectin (inside the type III homolog connective segment) induces the reactivity of monoclonal antibody (mAb) FDC6 (Freire-De-Lima et al., 2011). In addition, it was reported that O-GalNAc glycosylation may interfere with the intracellular degradation process of fibronectin after endocytosis, thus stabilizing fibronectin (Park et al., 2011). The study further suggested that the fibronectin O-glycosylation pathway may be an important factor in breast cancer development and progression (Park et al., 2011). Furthermore, another study indicated that poly-N-acetyllactosamine–containing glycans on fibronectin could decrease its binding affinity to gelatin (Sauerzapfe et al., 2009). A comprehensive glycosylation mapping of fibronectin can facilitate our understanding of the molecular details of these functions.

The initial glycoproteomics study of fibronectin was conducted 16 years ago by MALDI-TOF (Tajiri et al., 2005) glycopeptide analysis which was limited by conventional enrichment and MS fragmentation techniques. Recent advances in chromatography and mass spectrometry have greatly improved the efficiency of glycopeptide enrichment enabling high-resolution glycan analysis. Particularly, zwitterionic-hydrophilic interaction chromatography (ZIC-HILIC) has been applied extensively to improve glycopeptide enrichment with TFA added to the ACN/H_2O mobile phase system as an ion-pairing reagent (Wohlgemuth et al., 2009; Alagesan et al., 2017). Orbitrap brings higher collision dissociation HCD, thus enabling high resolution and more fragmentation in tandem MS, which is now routinely used in modern peptide analysis (Michalski et al., 2012). In addition, stepped normalized collision energy (sNCE)-HCD was recently developed to further enhance glycopeptide fragmentation (Liu et al., 2017). Moreover, an O-glycan–specific protease-assisted method named "site-specific extraction of O-linked glycopeptides" (EXoO) was recently developed, which allows for the comprehensive analysis of O-glycosites and O-glycans (Yang et al., 2020). We have been using these advanced glycoproteomic techniques to elaborate glycosylation of key proteins in thrombosis, achieving site-specific glycan mapping of human factor FVIII (Qu et al., 2020) and factor FV (Ma et al., 2020).

In this study, we performed systematic glycan analysis of human plasma fibronectin by optimizing and applying an integrated approach we recently developed (Ma et al., 2020). Briefly, Glu-C and trypsin were used for peptide mapping, N-glycopeptides were enriched by ZIC-HILIC, and O-glycopeptides were enriched through EXoO (Yang et al., 2020). N-glycosites and O-glycosites were localized by ^{18}O-labeling and EXoO (with or without sialidase), respectively. Optimized sNCE-HCD fragmentation (Zhu et al., 2020) was utilized to annotate N- and O-glycopeptide sequences. As a result, 308 unique glycopeptides comprising 6 N-glycosites and 53 O-glycosites were identified with simultaneous determination of peptide sequences and glycoform compositions.

MATERIALS AND METHODS

Materials

Reagents, solvents, and chemicals were all from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Human plasma fibronectin with a purity of a minimum 90% was purchased from Haematologic Technologies (Essex Junction, VT). Trypsin (sequencing grade modified) and endoproteinase Glu-C (Glu-C) were from Promega (Madison, WI). PNGase F was obtained from New England Biolabs (Beverly, MA). The 10-kDa Microcon centrifugal filter devices were purchased from Millipore (Bedford, MA). The ZIC-HILIC material was from SeQuant (Umea, Sweden). The OpeRATOR/SialEXO kit was from Genovis, Inc. (Cambridge, MA).

Sample Preparation and Enzymatic Digestion

Filter-aided sample preparation (FASP) was used to prepare samples for MS analysis. Briefly, the fibronectin protein was dissolved in 0.4% SDS, 50 mM dithiothreitol (DTT) in 50 mM NH_4HCO_3 (pH 7.8), which was incubated at 95°C for 10 min. The resulting solution was diluted by 200 μL of 8 M urea in 100 mM Tris-HCl buffer, pH 8.5 (UA solution), and then transferred to a 10-kDa ultracentrifuge filter for centrifugation at 13,500 g for 20 min. The concentrate was then mixed with another 200 μL of the UA solution and centrifuged twice. Subsequently, 50 μL of 50 mM iodoacetamide (IAA) in the UA solution was added, and the mixture was incubated in darkness at room temperature for 30 min, followed by brief centrifugation for 30 min. Then, the concentrate was diluted with 200 μL of UA solution and centrifuged twice to remove excess amounts of IAA. Finally, the sample was diluted with 100 μL of 40 mM NH_4HCO_3 and concentrated twice. After concentrating, the sample was digested with 1:50 trypsin:sample (w/w) in 40 mM NH_4HCO_3 (pH 7.8)

overnight at 37°C. In protein ID identification, Glu-C was added at a 1:10 ratio (w/w) in the same buffer and incubated for 4 h at 37°C before trypsin digestion. Fibronectin peptides were eluted with 50 μL NH_4HCO_3 (pH 7.8) by centrifuging the filter units for 20 min. This step was repeated three times. The final concentration of peptides was determined by a UV spectrophotometer (Nanodrop, Thermo) using an extinction coefficient of 1.1 for 0.1% (g/L) solution at 280 nm.

N-Glycopeptide Enrichment

N-glycopeptides were enriched by an in-house packed ZIC-HILIC micro-tip as previously described (Ma et al., 2020). The HILIC micro-tip was washed with 500 μL of ACN, water, and binding buffer (80% ACN/1% TFA) twice sequentially. For sample loading, the dried fibronectin peptides were dissolved in 500 μL of binding buffer and loaded onto the tip three times. The tip was then washed with 1.5 ml of binding buffer, and glycopeptides were eluted with 1 ml of elution buffer (0.1% TFA). The eluted solution was lyophilized and stored at -20°C until use. Enriched glycopeptides were directly injected into an LTQ-Orbitrap Elite mass spectrometer (MS) for intact glycopeptide analysis. For N-glycosite analysis, HILIC-enriched N-glycopeptides were incubated with PNGase F in 20 μL of 50 mM NH_4HCO_3 (pH 7.5) in $H_2^{18}O$ at 37°C overnight. Deglycosylated peptides were desalted, dried, and stored at -20°C for mass spectrometry analysis.

O-Glycopeptide Analysis

The EXoO method was used to identify O-glycoforms and O-glycosites as previously reported (Yang et al., 2020). Briefly, 200 μL of AminoLink resin (Pierce, Rockford, IL) was incubated with 500 μg of digested peptides in 50 mM $NaCNBH_3$ and 50 mM NaH_2PO_4 (pH 7.5) overnight at room temperature. The resin was then washed in a spin column and blocked by 1 M Tris-HCl (pH 7.4) and 50 mM $NaCNBH_3$ at room temperature for 30 min. After washing, the O-glycopeptides were released from the resin by the O-glycan–specific protease OpeRATOR (1 unit per 1 μg peptides) in 20 mM Tris-HCl (pH 6.8) at 37°C for 15 h. The released O-glycopeptides were desalted and dried for MS analysis. For O-glycosite analysis, sialidase was added together with OpeRATOR, so O-glycoforms would be unified and thus enhance the signal of O-glycopeptides.

LC–MS/MS Analysis of Intact Glycopeptides

Experiments were performed on an LTQ-Orbitrap Elite MS equipped with an EASY-Spray source and a nano-LC UltiMate 3000 high-performance liquid chromatography system (Thermo Fisher). An EASY-Spray PepMap C18 column (length, 15 cm; particle size, 3 μm; pore size, 100 Å; Thermo Fisher) was used for separation. The separation was achieved under a linear gradient elution condition from 3 to 40% solvent B for 30 min at a flow rate of 300 nL/min (mobile phase A, 2% ACN, 98% H_2O, and 0.1% FA; mobile phase B, 80% ACN, 20% H_2O, and 0.1% FA). LTQ-Orbitrap Elite was operated in the data-dependent mode, and the ten most intense ions in MS^1 were subjected to HCD 30 in the HCD collision cell for deglycosylated peptide analysis, or stepped normalized collision energy (sNCE)-HCD 15–30–45 fragmentation for intact glycopeptide analysis. The Orbitrap MS acquired a full-scan survey (m/z range from 375 to 1,500; automatic gain control target, 10^6 ions; resolution of 60,000 at m/z 400; maximum ion accumulation time, 50 ms). For sNCE-HCD, the Orbitrap analyzer acquired HCD fragment ion spectra with a resolution of 15,000 at m/z 400 (automatic gain control target, 10,000 ions; maximum ion accumulation time, 200 ms). The MS/MS scan model was set as the centroid. Other conditions used include an S-lens RF level of ~60% and an ion selection threshold of 50,000 counts for HCD.

Data Analysis

Data analysis of N-glycosite mapping was performed by pFind 3.0 (Chi et al., 2018). Peptide fragments were matched against the fibronectin protein sequence (UniProtKB entry P02751), where iodoacetamide on Cys was set as a static modification and oxidation of Met and ^{18}O-labeling of Asn (m = 2.9848) were set as a dynamic modification. Trypsin and Glu-C were chosen as the enzyme, and two missed cleavages were allowed. A false discovery rate (FDR) of 1% was estimated and applied at the peptide level. pGlyco 2.0 was applied for intact N-glycopeptide analysis (Liu et al., 2017). O-glycosylation results were manually interpreted with the assistance of GPQuest (Toghi Eshghi et al., 2015). Briefly, the MS/MS spectra containing at least two of the oxonium ions of HexNAc, including m/z = 126.05, 138.05, 144.06, 168.06, 186.08, and 204.08, were selected as glycopeptide spectra. The presence of at least 30% of b or y ions and three intact glycopeptide ions was required (Toghi Eshghi et al., 2015). The theoretical peptide database was constructed by using Lys/Arg on the C-terminal side (trypsin digestion) followed by Ser/Thr (OpeRATOR digestion) on the N-terminal side with four miss-cleavage sites allowed. In all analysis, the mass tolerance was set at 10 ppm for precursor ions and 50 ppm for product ions. Relative quantitative analysis was performed by comparing the peak area of ion chromatograms extracted from Xcalibur. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaíno et al., 2014) with the dataset identifier PXD025886.

RESULTS AND DISCUSSION

Protein and Glycopeptide Identification

Proteomic analysis of Glu-C-tryptic–digested fibronectin confirmed the identity of plasma fibronectin. The sequence coverage of the full-length glycoprotein (fibronectin, 2475 AAs, ~500 kDa) is 90.35% with minimal protein impurities. ZIC-HILIC enabled the enrichment of most N-glycopeptides, and sNCE-HCD fragmentations enabled the determination of both peptide sequences and glycoform compositions in one spectrum. From our previous optimization, HCD 15–30–45 was the most suitable condition for Orbitrap Elite (Zhu et al., 2020). For example, the N-glycopeptide
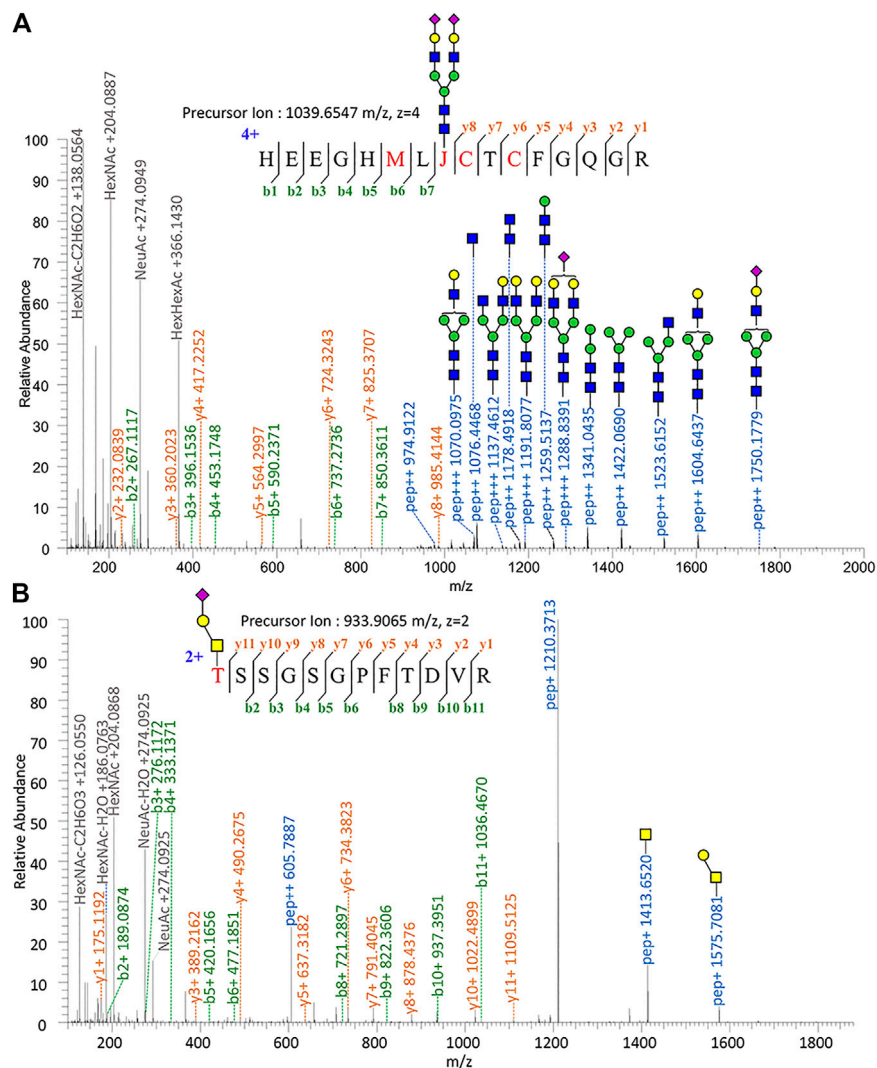
FIGURE 1 | Tandem mass spectrum annotation of fibronectin glycopeptides. **(A)** N-glycopeptide HEEGHMLJ[542]CTCFGQG R and **(B)** O-glycopeptide T[279]SSGSGPFTDVR.

HEEGHMLJ*CTCFGQGR (Asn542) of fibronectin with disialylated complex type N-glycan was fully annotated from the fragment information (**Figure 1A**) and Asn with glycosylation was replaced by "J" in data analysis. Oxonium ions of HexNAc (m/z = 204.08) could be detected under our fragmentation method, including m/z = 186.08, 144.07, 138.05, and 126.05 (**Figure 1**) (Halim et al., 2014). The oxonium ion of Neu5Ac (m/z = 274.09 or 292.10) was used to determine the presence of sialic acid. Diagnostic ions of HexHexNAcFuc (m/z = 512.19 Da) were applied to define the fucose branch (**Supplementary Figure S3**). A core-fucose structure was determined by the PepHexNAcFuc (+) ion and its neutral loss of fucose ion (146.06 Da) PepHexNAc (+) (**Supplementary Figure S1**) (Zhou et al., 2017).

Since OpeRATOR digests O-glycopeptides at the N-terminus of O-glycosylated Ser or Thr, the EXoO method enables the enrichment of O-glycopeptides as well as the identification of

O-glycan localization (O-glycosites). Similarly, sNCE-HCD was used to profile fibronectin O-glycopeptides, as shown in **Figure 1B** showing the spectrum of O-glycopeptide T*SSGSGPFTDVR (Thr279). Tandem MS annotations of other O-glycopeptides are provided in **Supplementary Figures S14–S21**.

N-Glycosylation Profiling and Relative Quantification

A total of 82 unique *N*-glycopeptides were identified from 302 MS spectra of HILIC-enriched samples, which contain 38 site-specific N-glycoforms on 6 *N*-glycosites as evidenced by HCD fragmentation (**Figure 2**, **Supplementary Table S1**). Three *N*-glycosites were identified on type I domains, two *N*-glycosites were identified on type III domains, and only one *N*-glycosite was found on type II domains. In addition, four or

FIGURE 2 | Microheterogeneity and relative abundance of *N*-glycoforms at six fibronectin *N*-glycosites. Only two glycoforms were identified at Asn1904. Two N-glycosites (red color) were newly identified sites. ▪: GlcNAc; ●: Man; ○: Gal; ◆: Neu5Ac; ▼: Fuc.
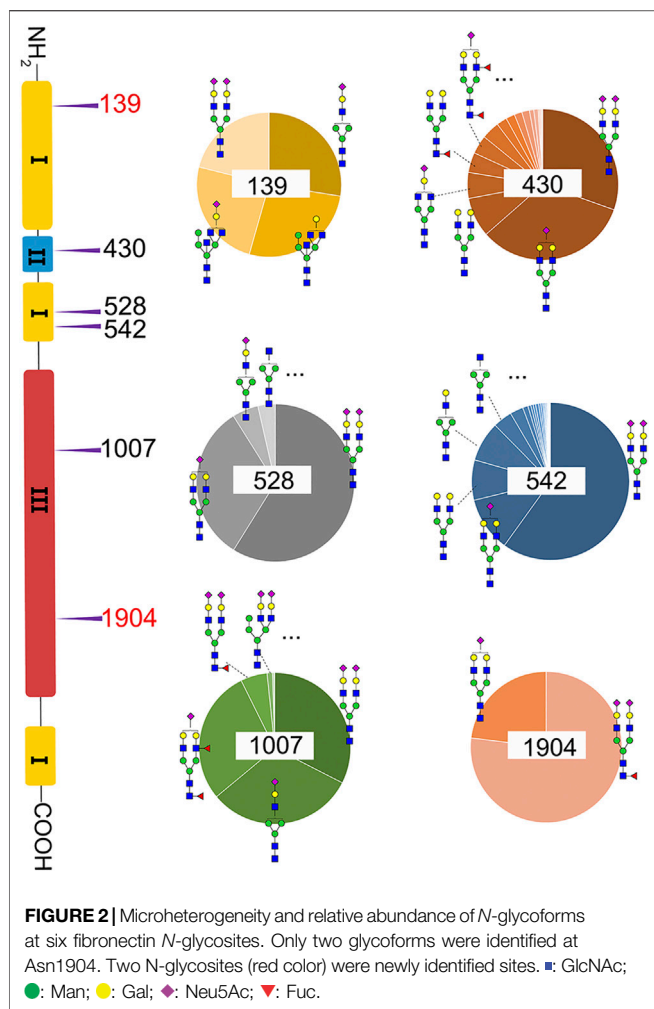
more glycoforms were identified on five *N*-glycosites, suggesting a glycan microheterogeneity of fibronectin. Among the six *N*-glycosites, Asn542 located within type I domain 8 was the most heterogeneous site with 27 different N-glycoforms. Sialyated glycans are predominant on this site. The second most heterogenous N-glycosite is Asn430 within type II domain 2, which presents 17 different glycoforms. Additionally, eight, four, and four glycoforms were identified on N-glycosites Asn1007, Asn528, and Asn139, respectively. Interestingly, Asn1904 contained only two glycoforms with structural similarity (N4H5S1, N4H5F1S2).

We then quantified the overall relative abundance of all *N*-glycoforms as well as the relative abundance of *N*-glycoforms at each *N*-glycosite. Ion chromatograms of identified peptides extracted from Xcalibur were used to quantify different glycopeptides. The mass-to-charge ratio (m/z) and retention time (RT) of glycopeptide precursor ions identified above were used to extract ion chromatograms and calculate the peak area of individual glycopeptides. Peak areas were then normalized to obtain site-specific relative abundances of each *N*-glycoform as shown in pie charts (**Figure 2**). Among 27 *N*-glycoforms identified at Asn542, disialylated bi-antennary

N4H5S2 (**21**) and monosialylated bi-antennary N4H5S1 (**20**) are highly abundant, with 59.93 and 11.23%, respectively. The top three abundant *N*-glycans on site Asn430 are similar to those on Asn542 with different abundance, including N4H5S1 (**20**, 33.11%), N4H5S2 (**21**, 30.42%), and N4H5 (**13**, 8.48%). Asn1007 also contains a high percentage of N4H5S2 (**21**, 32.51%) but also other glycoforms such as N3H4S1 (**10**, 31.47%) and N4H5F2S1 (**19**, 28.70%). **Figures 3A and C** illustrate the overall distribution of fibronectin *N*-glycans. The top two abundant *N*-glycans are all complex types, including N4H5S2 (**21**, 57.09%) found on most sites and N4H5S1 (**20**, 13.43%) identified on Asn430, Asn528, Asn542, and Asn1904.

Another observation is that most multi-antennary *N*-glycans are in relatively low abundance. The most abundant tri-antennary hybrid type *N*-glycan N5H5 (**23**, 26.87%) and its monosialylated form N5H5S1 (**27**, 24.52%) are all attached on Asn139, and both carry the bisecting GlcNAc. It was reported that fibronectin binds to transmembrane receptor protein integrins (Pankov and Yamada, 2002), and the existence of the bisecting GlcNAc on the α5 subunit could considerably diminish the adhesion of integrin α5β1 to fibronectin (Takahashi et al., 2009). It is thus reasonable to speculate that Asn139 within type I domain 3 and the glycans on this site could play a critical role in the integrin–fibronectin interaction, as no bisecting glycans were identified on other glycosites.

FIGURE 3 | A) N-glycan percentage of each N-glycosite; **(B)** O-glycan percentage of each O-glycosite; **(C)** overall N-glycan percentage; **(D)** overall O-glycan percentage. ▪: GlcNAc; ▪: GalNAc; ○: Gal; ◆: Neu5Ac; ▼: Fuc.
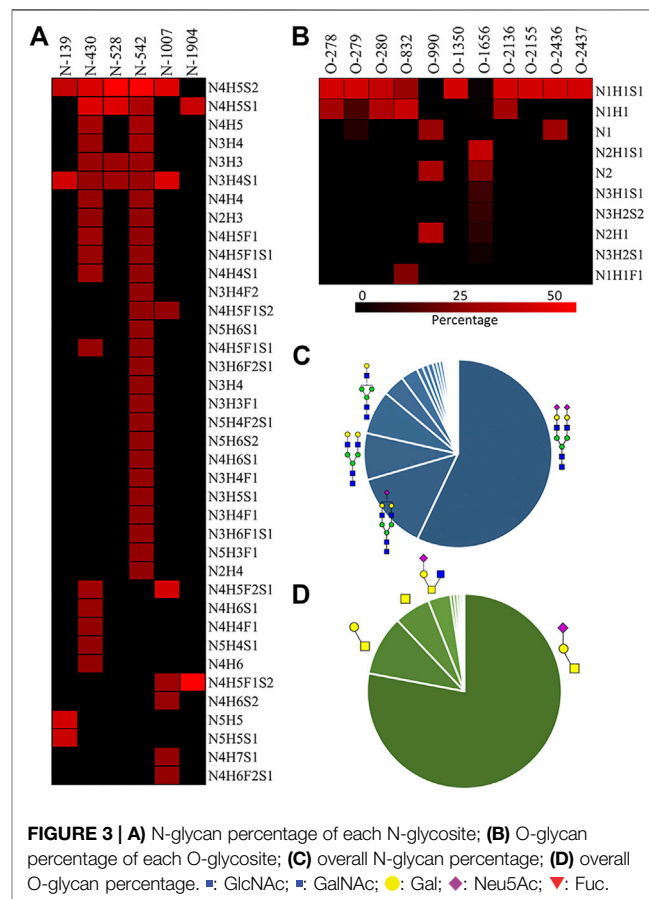
We observed that fibronectin is highly sialylated, with 77.31% of identified N-glycans containing one (19.68%) or two (57.63%) sialic acid residues. Sialic acid can affect conformation and oligomerization and the interaction function of proteins (Bhide and Colley, 2017). The existence of sialic acid on proteins affects their absorption, half-life, and clearance from the serum, as well as the physical, chemical, and immunogenic properties (Bork et al., 2009). Besides, high degrees of sialylation were found to play functional roles in heamostasis glycoproteins. For example, sialylation of the α5β1 integrin can decrease its binding affinity to fibronectin, thereby affecting cell adhesion in myeloid cells (Pan and Song, 2010). In addition, another hemostasis protein von Willebrand factor (VWF) was reported to be highly sialylated, which can interact with the asialoglycoprotein receptor (ASGPR) in the liver (O'Sullivan et al., 2016). It is possible that high sialylation of fibronectin might affect its interaction with certain extracellular matrix proteins, including collagen (Teoh et al., 2018), fibrin (Makogonenko et al., 2002), or fibulin-1 (Gu et al., 2000).

Compared with previously reported N-glycosylations of fibronectin (Tajiri et al., 2005), two new sites were discovered, including Asn139 and Asn1904. Surprisingly, Asn139 was found to be located in the Asn-Xxx-Cys sequence instead of the consensus site Asn-Xxx-Ser/Thr. The N-glycosite in Asn-Xxx-Cys sequences was not noticed until recently. This atypical N-glycosite was discovered in human protein C and human GPR109A, which significantly affected the protein functions (Gil et al., 2009; Yasuda et al., 2015). It was noticed that three previously identified sites were not observed in intact glycopeptide analysis, including Asn877, Asn1244, and Asn2199. These sites are located in theoretical tryptic-digested long peptide chains, which were hardly observed using the current intact glycopeptide analysis method. Thereby, Glu-C was used to further shorten the peptide chain and determine the sites through ^{18}O-labeling. The peptide fragmentation results of the three sites with ^{18}O-labeled glycosylation sites are shown in **Supplementary Figures S11–S13**. Unfortunately, we were not able to identify intact glycopeptides after Glu-C and trypsin combined digestion due to software limitations. The peptide fragmentation results with other ^{18}O-labeled glycosylation sites are shown in **Supplementary Figures S5–S10**.

O-Glycosylation Profiling and Relative Quantification

Fibronectin was reported as one of the most highly O-glycosylated proteins in the blood, with 71 possible sites (lacking glycoform information) identified from human serum samples. However, only 28 of those sites were precisely localized via a lectin enrichment method (King et al., 2017). Here, we localized the exact O-glycosites by using the EXoO method (Ma et al., 2020). O-glycoforms at each site were also identified and relatively quantified by using the modified EXoO method. In total, 16 different O-glycoforms on 53 O-glycosites were identified on fibronectin. Among the 53 O-glycosites, 11 were previously localized and 14 were identified before without exact localization information (King et al., 2017). The other

FIGURE 4 | Identified O-glycosites and glycoforms on each site in fibronectin. Red highlighted glycosites are new sites localized in this report. ■: GlcNAc; ■: GalNAc; ●: Gal; ◆: Neu5Ac; ▼: Fuc.
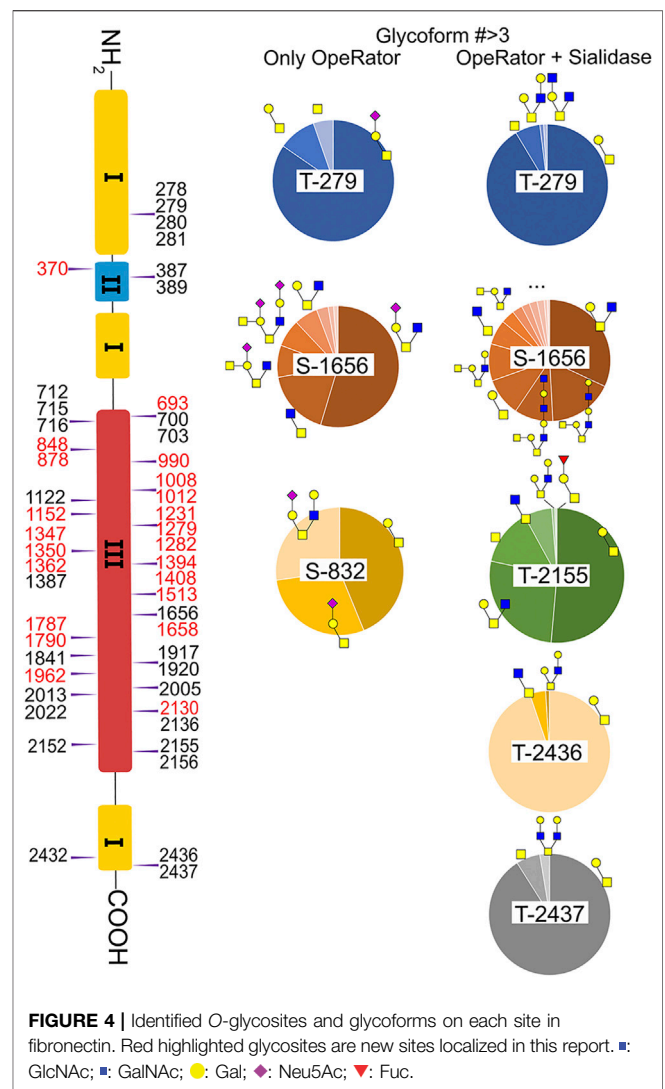
28 O-glycosites were never identified before. Specifically, our method enabled the identification of a new O-glycosite on type I domains (Ser281), a new site on type II domains (Thr370), and 26 sites on type III domains (**Figure 4**). Collectively, 43 identified O-glycosites are located within type III domains and 10 O-glycosites were distributed on other domains. We did not detect the other 17 possible O-glycosites, which may be due to different glycopeptide enrichment methods (Ma et al., 2020). The core 1 and Tn structures (**Supplementary Table S2**, structures **39–42**) were the major O-glycoforms observed on most sites, but core 2, core 3, and core 4 O-glycans were also identified (**Supplementary Tables S2, S3**, **43–59**).

In this study, the EXoO method with or without sialidase was used to study the O-glycosylation of fibronectin (Yang et al., 2020). Sialic acid information of the O-glycoforms was obtained without sialidase treatment, but treatment with sialidase enabled the identification of more O-glycosites and glycoforms. Anyway, we successfully identified 10 O-glycoforms (5 are sialylated) on 11 O-glycosites using the EXoO methods without sialidase treatment (**Supplementary Table S2**). This is the first example

of determining complete O-glycan structures with sialic acid information on glycoproteins, as the previously reported EXoO method requires de-sialylation (Ma et al., 2020). Additional sialidase treatment enabled identifying 6 more glycoforms and 43 more O-glycosites (**Supplementary Table S3**). For example, as shown in **Figure 4**, Thr279 was found mainly occupied by the sialyl-T antigen with also T and Tn antigens identified when using the non-sialidase–treated method, while some low abundant core 2 structures were also observed at the site when using the sialidase-treated method. Most strikingly, Thr2155 was only detected to have sialyl-T antigen without sialidase treatment but was found to be the second most complex O-glycosite with six glycoforms using sialidase-treated EXoO (**Figure 4**). Interestingly, the poly-LacNAc structure that could enhance the binding affinity between gelatin and fibronectin (Sauerzapfe et al., 2009) was possibly identified in core 2 O-glycans on Ser1656. However, the ExoO method cannot exclude multiple O-glycosites for peptides with repetitive sequences (Yang et al., 2020).

The relative abundance of glycoforms at each O-glycosite was also determined through peak area integration using the same methods for determining N-glycoform abundance. The relative abundance of O-glycans on some sites (>3 O-glycoforms) is shown in pie charts in **Figure 4**. Among O-glycoforms identified at Thr279, sialyl-T antigen N1H1S1 (**41**), T antigen N1H1 (**40**), and Tn antigen N1 (**39**) occupy 84.6, 10.1, and 5.3% of total glycoforms, respectively. The most heterogeneous O-glycosite Ser1656 contains sialylated core 2 structure N2H1S1 (**43**, 54.6%), asialayted core 3 structure N2 (**45**, 17.6%), sialylated core 4 structure N3H1S1 (**48**, 8.5%), disialylated core 4 structure N3H2S2 (**46**, 7.3%), core 2 structure N2H1 (**44**, 6.3%), and 13 other glycoforms. **Figure 3D** shows the overall distribution of fibronectin O-glycans identified by the non-sialidase–treated EXoO method. The top two abundant glycoforms are both core 1 structures. The most abundant structure is sialyl-T antigen (78.0%), which was identified on 10 O-glycosites. The second abundant glycoform is the T antigen (10.0%) that distributed on six *O*-glycosites. The Tn antigen (46.0%) is the third abundant O-glycoform, which was found on three O-glycosites.

O-glycosylation of fibronectin was proven to determine its binding to mAb FDC-6 (Nichols et al., 1986). Moreover, a recent study found that the key O-glycosylation is located in the type III homology connective segment (IIICS) domain (Freire-De-Lima et al., 2011). Furthermore, O-glycanase–treated fibronectin showed a reduced affinity with FDC-6, suggesting that this binding might associate with less complex asialylated O-glycans (Feinberg and Wang, 1994). However, our data showed that the previously identified site T2155 is primarily occupied by the sialyl-T antigen (**Figure 3B**) and thus may not associate with the binding to FDC-6. Here, we identified three other O-glycosites in the type IIICS domain, including T2130, S2136, and T2152, all occupied with asialylated glycoforms, which may thus associate with the binding to FDC-6.

Sialylation on O-glycans was found to play key roles in molecular interactions and protein functions. For example,

platelets clearance is necessary for normal hemostasis in humans, and reduced sialylation of O-glycans in platelets causes increased clearance in the liver (Li et al., 2017). The sialylation of O-glycans on cell surface CD8 negatively affects the binding affinity between histocompatibility complex class I molecules (MHCI) and CD8 (Moody et al., 2001). Similar to N-glycans, O-glycans of fibronectin were also highly sialylated. As shown in **Figure 3D**, 82.88% of total O-glycans were sialylated. Such a high O-glycan sialylation could possibly influence the function of fibronectin or interactions with other biomolecules.

CONCLUSION

In summary, we performed site-specific N- and O-glycosylation analyses of human plasma fibronectin through an integrated strategy. Multi-enzyme digestion, ZIC-HILIC enrichment, and sNCE-HCD fragmentation enabled complete annotation of fibronectin glycopeptides. In total, 82 unique N-glycopeptides and 226 O-glycopeptides were detected from enriched fibronectin samples. From the glycopeptides, we identified 6 N-glycosites carrying 38 N-glycoforms and 53 O-glycosites carrying 16 O-glycoforms. The glycosite includes 2 new N-glycosites (one is an atypical Asn-Xxx-Cys site) and 28 new O-glycosites. Furthermore, complete O-glycan structures with sialic acid information were identified for the first time. The comprehensive N- and O-glycosylation mapping fills a knowledge gap of fibronectin and could facilitate its functional studies as well as fibronectin-related therapeutics development.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, and further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

DL and LL designed the project and wrote the manuscript. DL performed the experiment and interpreted data. XW assisted with results analysis. All authors revised and approved the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.691217/full#supplementary-material

REFERENCES

Alagesan, K., Khilji, S. K., and Kolarich, D. (2017). It Is All about the Solvent: on the Importance of the mobile Phase for ZIC-HILIC Glycopeptide Enrichment. *Anal. Bioanal. Chem.* 409, 529–538. doi:10.1007/s00216-016-0051-6

Bhide, G. P., and Colley, K. J. (2017). Sialylation of N-Glycans: Mechanism, Cellular Compartmentalization and Function. *Histochem. Cel Biol* 147, 149–174. doi:10.1007/s00418-016-1520-x

Bork, K., Horstkorte, R., and Weidemann, W. (2009). Increasing the Sialylation of Therapeutic Glycoproteins: the Potential of the Sialic Acid Biosynthetic Pathway. *J. Pharm. Sci.* 98, 3499–3508. doi:10.1002/jps.21684

Chi, H., Liu, C., Yang, H., Zeng, W.-F., Wu, L., Zhou, W.-J., et al. (2018). Comprehensive Identification of Peptides in Tandem Mass Spectra Using an Efficient Open Search Engine. *Nat. Biotechnol.* 36, 1059–1061. doi:10.1038/nbt. 4236

Erickson, H. P. (2002). Stretching Fibronectin. *J. Muscle Res. Cell Motil.* 23, 575–580. doi:10.1023/a:1023427026818

Feinberg, R. F., and Wang, C.-L. (1994). Monoclonal Antibody FDC-6 Exhibits Binding to Human Plasma Fibronectin: A Caveat for Cervicovaginal Oncofetal Fibronectin Testing?. *Am. J. Obstet. Gynecol.* 171, 1302–1308. doi:10.1016/0002-9378(94)90152-x

Freire-De-Lima, L., Gelfenbeyn, K., Ding, Y., Mandel, U., Clausen, H., Handa, K., et al. (2011). Involvement of O-Glycosylation Defining Oncofetal Fibronectin in Epithelial-Mesenchymal Transition Process. *Proc. Natl. Acad. Sci.* 108, 17690–17695. doi:10.1073/pnas.1115191108

Furie, B., and Furie, B. C. (2005). Thrombus Formation *In Vivo*. *J. Clin. Invest.* 115, 3355–3362. doi:10.1172/jci26987

Gil, G.-C., Velander, W. H., and Van Cott, K. E. (2009). N-glycosylation Microheterogeneity and Site Occupancy of an Asn-X-Cys Sequon in Plasma-Derived and Recombinant Protein C. *Proteomics* 9, 2555–2567. doi:10.1002/pmic.200800775

Gu, Y.-C., Nilsson, K., Eng, H., and Ekblom, M. (2000). Association of Extracellular Matrix Proteins Fibulin-1 and Fibulin-2 with Fibronectin in Bone Marrow Stroma. *Br. J. Haematol.* 109, 305–313. doi:10.1046/j.1365-2141.2000.02011.x

Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., et al. (2014). Assignment of Saccharide Identities through Analysis of Oxonium Ion Fragmentation Profiles in LC-MS/MS of Glycopeptides. *J. Proteome Res.* 13, 6024–6032. doi:10.1021/pr500898r

Hsiao, C.-T., Cheng, H.-W., Huang, C.-M., Li, H.-R., Ou, M.-H., Huang, J.-R., et al. (2017). Fibronectin in Cell Adhesion and Migration *via* N-Glycosylation. *Oncotarget* 8, 70653–70668. doi:10.18632/oncotarget.19969

Huang, J., Li, X., Shi, X., Zhu, M., Wang, J., Huang, S., et al. (2019). Platelet Integrin αIIbβ3: Signal Transduction, Regulation, and its Therapeutic Targeting. *J. Hematol. Oncol.* 12, 1–22. doi:10.1186/s13045-019-0709-6

Jara, C. P., Wang, O., Paulino do Prado, T., Ismail, A., Fabian, F. M., Li, H., et al. (2020). Novel Fibrin-Fibronectin Matrix Accelerates Mice Skin Wound Healing. *Bioactive Mater.* 5, 949–962. doi:10.1016/j.bioactmat.2020.06.015

King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., et al. (2017). Characterizing the O-Glycosylation Landscape of Human Plasma, Platelets, and Endothelial Cells. *Blood Adv.* 1, 429–442. doi:10.1182/bloodadvances.2016002121

Kuriakose, A., Chirmule, N., and Nair, P. (2016). Immunogenicity of Biotherapeutics: Causes and Association with Posttranslational Modifications. *J. Immunol. Res.* 2016, 1298473. doi:10.1155/2016/1298473

Li, Y., Fu, J., Ling, Y., Yago, T., Mcdaniel, J. M., Song, J., et al. (2017). Sialylation on O-Glycans Protects Platelets from Clearance by Liver Kupffer Cells. *Proc. Natl. Acad. Sci. USA* 114, 8360–8365. doi:10.1073/pnas.1707662114

Liu, M.-Q., Zeng, W.-F., Fang, P., Cao, W.-Q., Liu, C., Yan, G.-Q., et al. (2017). pGlyco 2.0 Enables Precision N-Glycoproteomics with Comprehensive Quality Control and One-step Mass Spectrometry for Intact Glycopeptide Identification. *Nat. Commun.* 8, 1–14. doi:10.1038/s41467-017-00535-2

Ma, C., Liu, D., Li, D., Zhang, J., Xu, X. Q., Zhu, H., et al. (2020). Comprehensive N- and O-glycosylation Mapping of Human Coagulation Factor V. *J. Thromb. Haemost.* 18 (8), 1884–1892. doi:10.1111/jth.14861

Makogonenko, E., Tsurupa, G., Ingham, K., and Medved, L. (2002). Interaction of Fibrin(ogen) with Fibronectin: Further Characterization and Localization of the Fibronectin-Binding Site†. *Biochemistry* 41, 7907–7913. doi:10.1021/bi025770x

Michalski, A., Neuhauser, N., Cox, J., and Mann, M. (2012). A Systematic Investigation into the Nature of Tryptic HCD Spectra. *J. Proteome Res.* 11, 5479–5491. doi:10.1021/pr3007045

Mitra, N., Sharon, N., and Surolia, A. (2003). Role of N-Linked Glycan in the Unfolding Pathway ofErythrina corallodendronLectin†. *Biochemistry* 42, 12208–12216. doi:10.1021/bi035169e

Moody, A. M., Chui, D., Reche, P. A., Priatel, J. J., Marth, J. D., and Reinherz, E. L. (2001). Developmentally Regulated Glycosylation of the CD8αβ Coreceptor Stalk Modulates Ligand Binding. *Cell* 107, 501–512. doi:10.1016/s0092-8674(01)00577-3

Nichols, E. J., Fenderson, B. A., Carter, W. G., and Hakomori, S. (1986). Domain-specific Distribution of Carbohydrates in Human Fibronectins and the Transformation-dependent Translocation of Branched Type 2 Chain Defined by Monoclonal Antibody C6. *J. Biol. Chem.* 261, 11295–11301. doi:10.1016/s0021-9258(18)67382-x

O'sullivan, J. M., Aguila, S., Mcrae, E., Ward, S. E., Rawley, O., Fallon, P. G., et al. (2016). N-linked glycan truncation causes enhanced clearance of plasma-derived von Willebrand factor. *J. Thromb. Haemost.* 14, 2446–2457. doi:10.1111/jth.13537

Pan, D., and Song, Y. (2010). Role of Altered Sialylation of the I-like Domain of β1 Integrin in the Binding of Fibronectin to β1 Integrin: Thermodynamics and Conformational Analyses. *Biophys. J.* 99, 208–217. doi:10.1016/j.bpj.2010.03.063

Pankov, R., and Yamada, K. M. (2002). Fibronectin at a Glance. *J. Cell Sci.* 115, 3861–3863. doi:10.1242/jcs.00059

Park, J.-H., Katagiri, T., Chung, S., Kijima, K., and Nakamura, Y. (2011). Polypeptide N-Acetylgalactosaminyltransferase 6 Disrupts Mammary Acinar Morphogenesis through O-Glycosylation of Fibronectin. *Neoplasia* 13, 320–IN10. doi:10.1593/neo.101440

Patten, J., and Wang, K. (2020). Fibronectin in Development and Wound Healing. *Adv. Drug Deliv. Rev.* 170, 353–368. doi:10.1016/j.addr.2020.09.005

Qu, J., Ma, C., Xu, X. Q., Xiao, M., Zhang, J., Li, D., et al. (2020). Comparative Glycosylation Mapping of Plasma-Derived and Recombinant Human Factor VIII. *PLoS One* 15, e0233576. doi:10.1371/journal.pone.0233576

Sano, K., Asahi, M., Yanagibashi, M., Hashii, N., Itoh, S., Kawasaki, N., et al. (2008). Glycosylation and Ligand-Binding Activities of Rat Plasma Fibronectin during Liver Regeneration after Partial Hepatectomy. *Carbohydr. Res.* 343, 2329–2335. doi:10.1016/j.carres.2008.03.027

Sauerzapfe, B., Křenek, K., Schmiedel, J., Wakarchuk, W. W., Pelantová, H., Křen, V., et al. (2009). Chemo-enzymatic Synthesis of Poly-N-Acetyllactosamine (Poly-LacNAc) Structures and Their Characterization for CGL2-Galectin-Mediated Binding of ECM Glycoproteins to Biomaterial Surfaces. *Glycoconj J.* 26, 141–159. doi:10.1007/s10719-008-9172-2

Shental-Bechor, D., and Levy, Y. (2008). Effect of Glycosylation on Protein Folding: a Close Look at Thermodynamic Stabilization. *Proc. Natl. Acad. Sci.* 105, 8256–8261. doi:10.1073/pnas.0801340105

Tajiri, M., Yoshida, S., and Wada, Y. (2005). Differential Analysis of Site-specific Glycans on Plasma and Cellular Fibronectins: Application of a Hydrophilic Affinity Method for Glycopeptide Enrichment. *Glycobiology* 15, 1332–1340. doi:10.1093/glycob/cwj019

Takahashi, M., Kuroki, Y., Ohtsubo, K., and Taniguchi, N. (2009). Core Fucose and Bisecting GlcNAc, the Direct Modifiers of the N-Glycan Core: Their Functions and Target Proteins. *Carbohydr. Res.* 344, 1387–1390. doi:10.1016/j.carres.2009.04.031

Teoh, S. T., Ogrodzinski, M. P., Ross, C., Hunter, K. W., and Lunt, S. Y. (2018). Sialic Acid Metabolism: a Key Player in Breast Cancer Metastasis Revealed by Metabolomics. *Front. Oncol.* 8, 174. doi:10.3389/fonc.2018.00174

Toghi Eshghi, S., Shah, P., Yang, W., Li, X., and Zhang, H. (2015). GPQuest: a Spectral Library Matching Algorithm for Site-specific Assignment of Tandem Mass Spectra to Intact N-Glycopeptides. *Anal. Chem.* 87, 5181–5188. doi:10.1021/acs.analchem.5b00024

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., et al. (2015). *Essentials of Glycobiology [internet]*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., et al. (2014). ProteomeXchange Provides Globally Coordinated Proteomics Data Submission and Dissemination. *Nat. Biotechnol.* 32, 223–226. doi:10.1038/nbt. 2839

Wohlgemuth, J., Karas, M., Eichhorn, T., Hendriks, R., and Andrecht, S. (2009). Quantitative Site-specific Analysis of Protein Glycosylation by LC-MS Using Different Glycopeptide-Enrichment Strategies. *Anal. Biochem.* 395, 178–188. doi:10.1016/j.ab.2009.08.023

Yang, W., Song, A., Ao, M., Xu, Y., and Zhang, H. (2020). Large-scale Site-specific Mapping of the O-GalNAc Glycoproteome. *Nat. Protoc.* 15, 2589–2610. doi:10. 1038/s41596-020-0345-1

Yasuda, D., Imura, Y., Ishii, S., Shimizu, T., and Nakamura, M. (2015). The Atypical N-glycosylation Motif, Asn-Cys-Cys, in Human GPR109A Is Required for normal Cell Surface Expression and Intracellular Signaling. *FASEB j.* 29, 2412–2422. doi:10.1096/fj.14-267096

Zhou, J., Yang, W., Hu, Y., Höti, N., Liu, Y., Shah, P., et al. (2017). Site-specific Fucosylation Analysis Identifying Glycoproteins Associated with Aggressive Prostate Cancer Cell Lines Using Tandem Affinity Enrichments of Intact Glycopeptides Followed by Mass Spectrometry. *Anal. Chem.* 89, 7623–7630. doi:10.1021/acs.analchem.7b01493

Zhu, H., Wang, S., Liu, D., Ding, L., Chen, C., Liu, Y., et al. (2020). Identifying Sialylation Linkages at the Glycopeptide Level by Glycosyltransferase Labeling Assisted Mass Spectrometry (GLAMS). *Anal. Chem.* 92, 6297–6303. doi:10. 1021/acs.analchem.9b05068

Chemistry-Assisted Proteomic Profiling of O-GlcNAcylation

*Qiang Zhu and Wen Yi**

Department of Hepatobiliary and Pancreatic Surgery, Zhejiang Provincial Key Laboratory of Pancreatic Disease, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

The modification on proteins with O-linked N-acetyl-β-D-glucosamine (O-GlcNAcylation) is essential for normal cell physiology. Dysregulation of O-GlcNAcylation leads to many human diseases, such as cancer, diabetes and neurodegenerative diseases. Recently, the functional role of O-GlcNAcylation in different physiological states has been elucidated due to the booming detection technologies. Chemical approaches for the enrichment of O-GlcNAcylated proteins combined with mass spectrometry-based proteomics enable the profiling of protein O-GlcNAcylation in a system-wide level. In this review, we summarize recent progresses on the enrichment and proteomic profiling of protein O-GlcNAcylation.

Keywords: O-GlcNAcylation, enrichment strategies, mass spectrometry, quantitative proteomics, chemical tools

INTRODUCTION

O-GlcNAcylation is a prevalent form of posttranslational modifications on the hydroxyl group of serine and/or threonine residues (Torres and Hart, 1984). Starting from fructose-6-phosphate, a glycolytic intermediate, a series of enzymatic reactions (collectively termed the hexosamine biosynthetic pathway) generate Uridine-Diphosphate N-acetylglucosamine (UDP-GlcNAc), the sugar donor for protein O-GlcNAcylation (**Figure 1**). UDP-GlcNAc can also be generated from the exogenous GlcNAc through the salvage pathway (Bond and Hanover, 2015), and by the enzymatic conversion of UDP-GalNAc by UDP-galactose-4'-epimerase (GALE) (**Figure 1**) (Boyce et al., 2011). Despite the occurrence of O-GlcNAcylation on numerous proteins, only two enzymes are responsible for the recycling of this modification in cells. O-GlcNAc transferase (OGT) catalyzes the addition of O-GlcNAc onto diverse protein substrates, while O-GlcNAc hydrolase (OGA) catalyzes the removal of this modification (Haltiwanger et al., 1990; Lubas et al., 1997). Notably, O-GlcNAcylation is reversible and highly dynamic in response to different cellular stimuli to regulate the structure and function of various intracellular proteins (Gao et al., 2001; Jang et al., 2012; Li and Yi, 2014; Ong et al., 2018). Besides, O-GlcNAcylation can interact with other posttranslational modifications including phosphorylation, acetylation and ubiquitination (Vercoutter-Edouart et al., 2015). These features make O-GlcNAcylation a regulator of various important and basic biological processes such as transcription, stem cell differentiation, signal transduction, cell cycle progression, and metabolic reprogramming (Hart et al., 2011; Bond and Hanover, 2015). For example, recent studies revealed that O-GlcNAcylation of Notch1 elevated its stability by abolishing the binding of E3 ubiquitin ligase Itch, thus maintaining the self-renewal of adult neural stem cells (Chen et al., 2021). Tan et al. found that O-GlcNAcylation of serine/arginine-rich protein kinase 2 (SRPK2) promoted *de novo* lipogenesis by regulating pre-mRNA splicing (Tan et al., 2021). Duan et al. revealed O-GlcNAcylation of RACK1 on serine 122 promoted its protein stability, ribosome binding and interaction with PKCβII to modulate hepatocellular carcinoma (HCC) tumorigenesis (Duan et al., 2018). Consequently, dysregulation of O-GlcNAc cycling has

been implicated in the pathology of various diseases, including but not limited to, diabetes, cancer, cardiovascular diseases, and neuronal disorders (Slawson and Hart, 2011; Vaidyanathan and Wells, 2014; Yuzwa and Vocadlo, 2014). However, the specific molecular mechanisms by which O-GlcNAcylation contributes to the development and progression of these diseases remain to be elucidated.

Elucidating the function of O-GlcNAcylation in both physiological and pathological processes requires reliable and powerful detection tools to visualize and quantify the dynamics of O-GlcNAcylation. However, it is challenging to detect O-GlcNAcylated proteins by conventional techniques due to the regulatory nature of the modification (e.g., low abundance and highly dynamic) and the unique chemical characteristics (e.g., low immunogenicity and chemically/enzymatically labile) (Thompson et al., 2018). These features have called for the development of effective approaches to enrich and quantify this modification. In this review, we aim to provide a concise summary of recent advances to use chemistry-assisted proteomic methods to profile protein O-GlcNAcylation in a system-wise level.

ENRICHMENT STRATEGIES FOR PROTEIN O-GLCNACYLATION

Antibodies and Lectins

Unlike phosphorylation and other PTMs for which site-specific antibodies are available, effective and specific antibodies for O-GlcNAc are difficult to develop due to the low immunogenicity of the neutral O-GlcNAc sugar (Monsigny et al., 1979). The commonly used O-GlcNAc antibodies are two pan-antibodies (CTD110.6 and RL2), raised against glycopeptides derived from the C-terminal domain of RNA polymerase II, and rat liver nuclear envelopes, respectively (Snow et al., 1987; Comer et al., 2001). In addition, a few mouse monoclonal antibodies were developed, including HGAC85 (Turner et al., 1990), 10D8 (Yoshida et al., 1989), 18B10.7C (#3), 9D1. E4 (#10), and 1F5. D6 (#14) (Teo et al., 2010). These pan-antibodies were produced to yield the broad possible coverage of the modification. Although these antibodies can be employed for the detection of O-GlcNAcylated proteins, they exhibit different substrate recognition specificity. For example, CTD110.6, 18B10. 7C (#3), and 9D1. E4 (#10) are more inclined to recognize O-GlcNAc on the cell surface glycoproteins, and CTD110.6 shows cross-reactivity toward GlcNAc-containing N-glycans. RL2 also has a preference toward specific peptide sequences (Tashima and Stanley, 2014). In addition to the antibodies, specific lectins were also used in studies to detect O-GlcNAc. The lectin WGA (Wheat Germ Agglutinin) was first applied to detect and enrich O-GlcNAcylated proteins. But this plant lectin can recognize all terminal GlcNAc sugars as well as sialic acids (Monsigny et al., 1979; Snow et al., 1987). To increase the specificity, the succinyl WGA (sWGA) was developed, in which the recognition of sialic acid was inhibited via succinylation of WGA into the sialic acid

FIGURE 1 | The source of O-GlcNAcylation donor UDP-GlcNAc in the cells. Glucose is converted to UDP-GlcNAc through the hexosamine biosynthetic pathway (HBP). Enzyme names shown in blue. GFAT: Glutamine fructose-6-phosphate amino-transferase, GNPNAT1: Glucosamine-Phosphate N-Acetyltransferase 1, PGM3: Phosphoglucomutase 3, UAP1: UDP-N-Acetylglucosamine pyrophosphorylase 1. UDP-GlcNAc also can be produced by the GlcNAc and GalNAc salvage pathways. Enzyme names shown in blue. NAGK: GlcNAc kinase, AGM1: GlcNAc-6-phosphate mutase, AGX1 or AGX2: UDP-GlcNAc pyrophosphorylase, GALK2: Galactokinase 2, GALE: UDP-galactose-4′-epimerase. Bold arrow represents that reaction is easier to perform. O-GlcNAc transferase (OGT) and O-GlcNAc hydrolase (OGA) regulate the addition and the removal of O-GlcNAc, respectively.

FIGURE 2 | Metabolic labeling strategy for capture and detection of O-GlcNAcylated proteins. (A) Schematic of metabolic labeling. Unnatural monosaccharides enter cells and are metabolically converted to UDP-GlcNAc analogs which serve as donors for O-GlcNAcylation by OGT. The labeled O-GlcNAcylated proteins are enriched through bioorthogonal reactions and analyzed by LC-MS/MS or imaged in living cells. (B) Metabolic chemical reporters used for the labeling of O-GlcNAcylated proteins in cells.

recognition domain (Nakanuma et al., 1993). Another two fungal lectins PVL and AAL2 can bind to terminal non-reducing GlcNAc moieties (Kochibe and Matta, 1989; Ren et al., 2013). Recently, a recombinant lectin PVL (rPVL) produced from *Escherichia coli* was reported to have a higher specificity and affinity for proteins with multiple GlcNAc than WGA, AAL2 and PVLA (Machon et al., 2017). All of these lectins do not distinguish between terminal N-linked GlcNAc and O-GlcNAc residues, thus, the addition of PNGase F and/or sialidase is needed to remove the complex N-glycans and sialic acids during the detection and enrichment of O-GlcNAcylated proteins (Cieniewski-Bernard et al., 2004; Lefebvre et al., 2004; Zachara, 2009).

Metabolic Labeling of O-GlcNAcylated Proteins

Metabolic chemical reporters (MCRs) of glycosylation are unnatural monosaccharide analogs that contain bioorthogonal functionalities such as an alkyne or azide (Grammel and Hang, 2013; Chuh and Pratt, 2015; Chuh et al., 2016). Metabolic incorporation of these MCRs followed by bioorthogonal reactions such as the copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC) has been extensively used to label complex glycans containing sialic acids, fucose, GalNAc, or GlcNAc (**Figure 2A**) (Prescher and Bertozzi, 2005; Best, 2009).

In the detection of O-GlcNAcylation, a few specific MCRs were developed (**Figure 2B**). The first O-GlcNAc-targeted MCR 1,3,5,6-tetra-O-acetyl-N-azidoacetyl-glucosamine (Ac4GlcNAz, 1) has been employed for the visualization and proteomic profiling of O-GlcNAcylated proteins (Vocadlo et al., 2003). The acetyl groups act as protecting groups to enhance the permeability of GlcNAz into the cell. After deacetylation by cellular esterases, Ac4GlcNAz could be metabolically converted to UDP-GlcNAz which was then transferred to proteins by OGT (Zaro et al., 2011). Using this MCR, Hahne et al. identified about 1,500 O-GlcNAc proteins in cells. Coupled with β-elimination reaction, they mapped 185 O-GlcNAc modification sites on 80 proteins (Hahne et al., 2013). However, GlcNAz showed low selectivity since it could be incorporated into N-glycans (Cieniewski-Bernard et al., 2014). Subsequently, 1,3,5,6-tetra-O-acetyl-N-azidoacetyl-galactosamine (Ac$_4$GalNAz, 2) was used to label O-GlcNAcylated proteins (Boyce et al., 2011). However, further studies revealed that GlcNAz and GalNAz could interconvert to each other in cells, causing the labeling reaction unable to distinguish O-GlcNAc from mucin-type O-linked glycans (Chuh et al., 2014; Qin et al., 2020). To solve this problem, Zaro et al. (2011) employed alkyneacetyl-GlcNAc analogue (GlcNAlk, 3) and alkyneacetyl-GalNAc analogue (GalNAlk, 4) to label O-GlcNAcylated proteins. They found that GlcNAlk could not metabolically convert to GalNAlk, thus it would not label mucin-type O-linked glycans. GalNAlk showed a lower labeling efficiency

FIGURE 3 | Schematic of chemoenzymatic labeling strategy for enrichment and detection of O-GlcNAc proteins. O-GlcNAcylated proteins in cell lysates are labeled with azide group using β-1,4-galactosyltransferase GalT (Y289L). The biotin with the alkyne group is added to O-GlcNAcylated proteins by copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) chemistry. Following labeling with biotin, O-GlcNAcylated proteins can be captured using streptavidin resin and then simultaneously proteins are digested and analyzed by LC-MS/MS.

since it was hard to metabolically convert to GlcNAlk in cells. Unfortunately, GlcNAlk was also incorporated into N-linked glycans, compromising the labeling specificity (Zaro et al., 2011). Chuh et al. (2014) circumvented this limitation by using 1,3,5-tri-O-acetyl-6-azido-6-deoxy-N-acetyl-glucosamine (Ac₃6AzGlcNAc, **5**) (Chuh et al., 2014). Unlike the other MCRs mentioned above, 6AzGlcNAc could be directly phosphorylated at the 1-hydroxyl to bypass the canonical GlcNAc salvage pathway, which endowed it with a higher degree of selectivity for O-GlcNAcylated proteins (Chuh et al., 2014). However, 6AzGlcNAc showed a lower conversion efficiency to UDP-GlcNAc as compared to GlcNAz, emphasizing the potential balance between labeling efficiency and selectivity. Notably, O-GlcNAcylation is a dynamic process, in which the endogenous OGA can rapidly remove metabolic labels on proteins to decrease the labeling signal. Recently, hydrolysis-resistant MCRs such as 1, 3, 6-tri-O-acetyl-4-deoxy- N-azidoacetyl-glucosamine (Ac₃4dGlcNAz, **6**) and 1,3,5,6-tetra-O-acetyl-2-azido-2-deoxy-glucose (Ac₄2AzGlc, **7**) have exhibited higher labeling efficiency and specificity for O-GlcNAc-modified proteins (Li et al., 2016; Zaro et al., 2017). Metabolic incorporation of Ac₄2AzGlc is resistant to hydrolysis due to the lack of anchimeric assistance of the N-acetyl group (Macauley et al., 2005). Differently, Ac₃4dGlcNAz reduces nonspecific incorporation into extracellular glycans and increases resistance to OGA hydrolysis due to the absence of the 4′-OH group. (Cecioni and Vocadlo, 2013). Despite the enhancement of MCR permeability into cells by acetyl protecting groups, a recent study showed that per-O-acetylated azido and alkynyl sugars may spontaneously react with the cysteine side chains to generate S-glycosylation through a nonenzymatic mechanism (Qin et al., 2017). Therefore, per-O-acetylated sugar MCRs likely cause some false positives during the profiling of O-GlcNAcylation (Hao et al., 2019). On the other hand, the non-O-acetylated sugars are hard to cross the cell membrane for efficient labeling. To address this issue, 1,3-di-O-propionyl-N-azidoacetylgalactosamine (1,3-Pr2GalNAz, **8**) was developed as a novel metabolic probe for O-GlcNAc labeling which could be

readily incorporated into O-GlcNAcylated proteins without introducing artificial S-glycosylation (Hao et al., 2019). Collectively, MCRs are robust and powerful tools to label and profile O-GlcNAcylated proteins in living cells.

Chemoenzymatic Labeling of O-GlcNAc Proteins

As a complementary approach to the metabolic labeling, a chemoenzymatic labeling strategy has also been widely used in the capture and profiling of O-GlcNAcylated proteins (**Figure 3**). The Hart lab first developed a radioassay using a radiolabeled UDP-Gal and the enzyme β-1, 4-galactosyltransferase 1 (GalT1), which specifically transfers Gal to terminal GlcNAc moieties (Hayes et al., 1995; Torres and Hart, 1984). To extend the application of this method, a GalT1 mutant (Y289L) was generated to expand the substrate binding pocket, which allowed for the transfer of UDP-Gal analogues appended with chemical tags including azide, followed by bioorthogonal reactions, O-GlcNAc-modified peptides were biotinylated and subsequently were captured with avidin beads, eluted with free biotin, and sequenced by ETD mass spectrometry. (**Figure 3**) (Khidekel et al., 2003; Clark et al., 2008). Using this chemoenzymatic labeling strategy, the Hsieh-Wilson group carried out the first glycoproteomic study of O-GlcNAcylated proteins in the rat brain, in which some O-GlcNAc sites on 25 O-GlcNAcylated proteins were mapped (Khidekel et al., 2004). The bulky biotin group compromised the glycopeptide recovery efficiency. In the follow-up studies, a few cleavable enrichment probes were employed to improve the recovery of enriched glycopeptides and increase the rate of true assignment. The Hart group used a photocleavable biotin-alkyne probe to capture GalNAz-tagged O-GlcNAcylated peptides. When exposed to UV light (365 nm), O-GlcNAcylated peptides were released from the avidin chromatography column, followed by protein identification and site mapping by mass spectrometry

Detection methods	Specificity	Sensitivity	Applications	References
Lectin	Low	Low	Western blot, proteomics	Vosseller et al. (2006), Chalkley et al. (2009)
Antibody	Moderate	Moderate	Western blot, proteomics	Snow et al. (1987), Teo et al. (2010), Comer et al. (2001)
Metabolic labeling	Moderate	High	Western blot, proteomics, live cell imaging, flow cytometry, in-gel fluorescence	Tan et al. (2018), Zaro et al. (2017), Qin W. et al. (2018), Li et al. (2016), Clark et al. (2008)
Chemoenzymatic labeling	High	High	Western blot, proteomics, in-gel fluorescence, histology, *in vivo* imaging	Rouhanifard et al. (2014), Aguilar et al. (2017), Li et al. (2019), Clark et al. (2008)

(Wang et al., 2010a). Alfaro et al. used a photocleavable biotin probe to enrich and identify 274 O-GlcNAcylated proteins in mouse cerebral tissues (Alfaro et al., 2012). Tsumoto et al. used a novel alkyne probe containing thiol-alkyne to capture O-GlcNAcylated peptides. Glycopeptides were released by the reversible disulfide formation with a thiol-reactive resin (Tsumoto et al., 2015). Griffin et al. developed a probe that would be positively charged when it was cleaved to facilitate ETD-MS detection. An alkyne- 1-(4, 4-dimethyl-2, 6-dioxocyclohex-1-ylidene)ethyl (Dde)-biotin linker was used to label O-GlcNAcylated proteins. The Dde moiety can be quantitatively removed by hydrazine and showed higher labeling efficiency than PC-biotin-alkyne (Griffin et al., 2016). On the other hand, 3-ethynylbenzaldehyde probe was used to react with GalNAz via the copper-catalyzed Huisgen 1, 3-cycloaddition to form aromatic aldehyde-derivatized glycopeptides which were enriched by reversible hydrazone formation with hydrazide resins. Subsequently, glycopeptides could be eluted using hydroxylamine (Nishikaze et al., 2013).

In addition to capturing O-GlcNAcylated proteins, this strategy can be applied in live cell imaging, histological detection and modification stoichiometry quantification. Clark *el at*. labeled O-GlcNAcylated proteins selectively with a fluorescent reporter group to detect and image cellular O-GlcNAcylated proteins in living cells. (Clark et al., 2008). The Wu group applied this strategy to label histological specimens and demonstrated the change of O-GlcNAc levels during tumor development (Aguilar et al., 2017). To quantify O-GlcNAc stoichiometries on specific proteins, the Hsieh-Wilson group conjugated O-GlcNAcylated proteins with PEG mass tags. Compared to the nonglycosylated proteins, O-GlcNAcylated proteins showed the mass-shifted bands detected by immunoblotting with indicated antibodies. The occupancy levels of O-GlcNAcylation were determined by the intensity ratio of the glycosylated and nonglycosylated bands (Rexach et al., 2010). The Pratt group employed a semisynthetic O-GlcNAcylated protein standard combined with Strain-Promoted Cycloaddition (SPAAC) chemoenzymatic mass tagging protocol to improve the accuracy of O-GlcNAc stoichiometries analysis (Darabedian et al., 2018).

In summary, the biochemical tools and methods such as antibodies, lectins, metabolic, or chemoenzymatic labeling have different specificity and sensitivity in terms of enrichment and detection of O-GlcNAcylated proteins (**Table 1**). These strategies exhibit a broad range of applications including immunoblotting, proteomics, cellular and histological imaging (**Table 1**).

QUANTITATIVE PROTEOMICS FOR O-GLCNACYLATION

O-GlcNAcylation is highly dynamic in response to various environmental stimuli. Quantifying its dynamics is key to elucidating the roles of O-GlcNAcylation in biological processes. Mass spectrometry-based quantitative proteomics, in combination with the aforementioned O-GlcNAcylated peptide enrichment methods, have recently emerged as a powerful tool to quantify protein O-GlcNAcylation in various biological settings.

Stable Isotope Labeling With Amino Acids in Cell Culture-Based Quantitative Proteomics

SILAC (Stable Isotope Labeling with Amino Acids in Cell Culture) is one of the most widely used quantitative proteomic techniques (Chen et al., 2015). Cells treated under different conditions are grown in the presence of normal (light) or isotopically enriched (heavy) versions of a specific label (amino acid, carbon, nitrogen) to produce unlabeled and fully labeled proteins. The glycosylated proteins are enriched, combined, followed by proteolysis and quantification by MS/MS. The mass shift due to the addition of isotope labeling in mass spectrometry can be used to quantify the difference in protein glycosylation abundance (**Figure 4**). Using SILAC-based quantitative proteomics, Zachara et al. identified 15 proteins that were dynamically modified by O-GlcNAc in response to heat stress (Zachara et al., 2011). Myers et al. found that occupancies of O-GlcNAc on different sites within the same protein were affected by polycombrepressive complex 2 (PRC2) in mouse embryonic stem cells, emphasizing the site-specific regulation of O-GlcNAcylation (Myers et al., 2011). The Hart group found that 10 proteins had an apparent increase of O-GlcNAcylation and 19 proteins showed a reduction of O-GlcNAcylation upon GSK-3 inhibition, indicating a complex interaction between phosphorylation and O-GlcNAcylation (Wang et al., 2007). Using SILAC combined with the chemoenzymatic labeling with a PC-biotin-alkyne tag, Wang et al. monitored the changes in the abundance of proteins and their O-GlcNAcylation during cytokinesis (Wang et al., 2010a). Recently, Qin et al. combined SILAC-based quantitative chemoproteomics with metabolic labeling using Ac36AzGlcNAc to analyze the turnover dynamics of O-GlcNAcylated proteins. Eventually, they identified 896 O-GlcNAcylated proteins, 86% of which showed a dynamic turnover in 12 h in the experiments (Qin et al., 2017).

FIGURE 4 | Schematic of SILAC-based Quantitative proteomics for profiling O-GlcNAcylation.
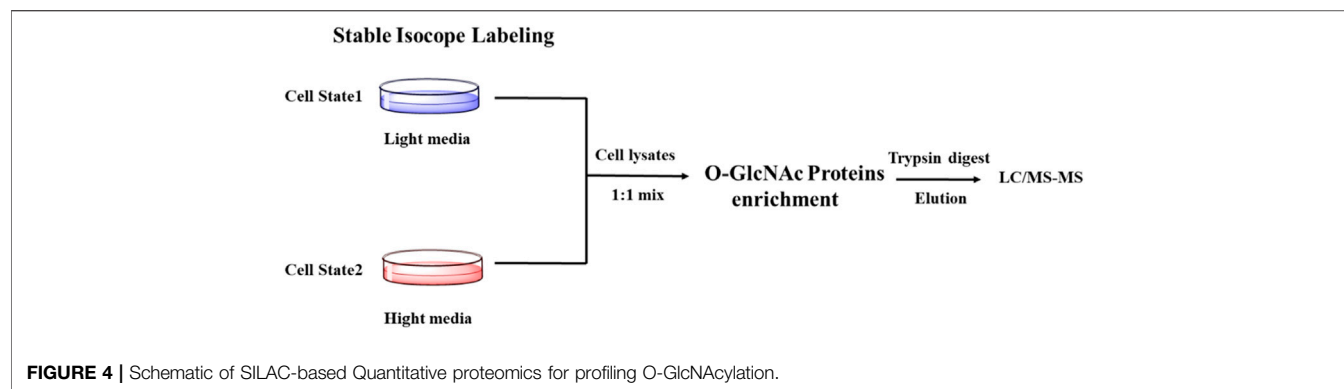
β-Elimination Followed by Michael Addition With Dithiothreitol-Based Quantitative Strategy

Mapping O-GlcNAcylation sites is vital for elucidating the functional role of O-GlcNAcylation in specific biological environment. However, this can be very challenging because there is no known consensus sequence of O-GlcNAcylation on proteins, and that O-GlcNAc occurs in substoichiometry on many proteins. In addition, the O-glycosidic bond is labile, easily lost in collision-induced dissociation (CID) mass spectrometry (Greis et al., 1996; Whelan and Hart, 2006). To address these challenges, Hart and his colleagues developed the BEMAD method (β-elimination followed by Michael addition with dithiothreitol), which chemically converts O-GlcNAcylated serine and threonine residues into stable thiol derivatives (**Figure 5**) (Wells et al., 2002). Using the isotope-labeled dithiothreitol, BEMAD had been applied to quantify and map O-GlcNAcylation sites after chemoenzymatic labeling (Zachara et al., 2011; Tsumoto et al., 2017). The replacement of labile glycosylation with a more stable dithiothreitol modification significantly improved the efficiency of site-identification (Wells et al., 2002). The Hart group further employed isobaric tags for relative and absolute quantification (iTRAQ) and BEMAD coupled with the chemoenzymatic labeling to compare the site-specific O-GlcNAc occupancy on proteins obtained from normal and diabetic erythrocytes, highlighting the differentially regulated O-GlcNAcylation in diabetic erythrocytes (Wang et al., 2009). Sakabe et al. used the chemoenzymatic labeling and BEMAD

method to identify various O-GlcNAc sites on histones H2A, H2B, and H4, elucidating that dynamic O-GlcNAcylation is a critical part of the histone code (Sakabe et al., 2009). Morover, Lund et al. employed a similar strategy to detect O-GlcNAc dynamics in response to T cell activation. More than 200 O-GlcNAcylated proteins were identified, among which are a number of proteins functionally related to RNA metabolism in human T cells, implying the functional importance of O-GlcNAcylation in T cell biology (Lund et al., 2016).

QUIC and TMT Tag Strategies for Profiling O-GlcNAcylation

Another approach couples O-GlcNAcylated peptide labeling/ enrichment methods with tandem mass tagging for quantitative profiling of O-GlcNAcylation. The Hsieh-Wilson group developed a quantitative isotopic and chemoenzymatic tag (QUIC-Tag) strategy to identify and quantify O-GlcNAcylation in mouse brains in response to cellular stimulation. O-GlcNAcylated proteins were labeled selectively with a ketone-containing galactose analog via the chemoenzymatic strategy. The ketone functionality was reacted with an aminooxy biotin derivative, which can be captured by avidin chromatography. Subsequently, the proteins were digested and labeled with formaldehyde/NaCNBH3 or deuterated formaldehyde/NaCNBD3 by a modified dimethyl labeling strategy (**Figure 6A**). Coupled with tandem mass spectrometry, they demonstrated the dynamic O-GlcNAcylation

FIGURE 5 | Schematic of BEMAD-based Quantitative strategy. O-GlcNAcylated serine and threonine residues can be converted to dehydroalanine by β-elimination reaction. Subsequently, isotope-labeled light DTT (DTTd0) or heavy DTT (DTT-d6) is added to O-GlcNAcylated peptides by Michael addition. The peptides are mixed and analyzed by LC-MS/MS. The red balls and blue balls represent heavy or light labeling sites, respectively.

FIGURE 6 | Schematic of QUIC and TMT Quantitative strategy for O-GlcNAcylation. **(A)** QUIC-Tag strategy. O-GlcNAcylated proteins first are chemoenzymatic labeled with a tag containing cleavable group and biotin, followed by trypsin digestion. The peptides are labeled with NaCNBH3 or NaCNBD3 by a modified dimethyl labeling strategy. Subsequently, O-GlcNAcylated peptides are enriched by streptavidin agarose and then released, mixed and analyzed by LC-MS/MS. **(B)** TMT strategy. The proteins are digested, labeled with TMT. O-GlcNAcylated peptides are then enriched, released and subjected to LC-MS/MS analysis.

in mediating neuronal communication (Khidekel et al., 2007). The tandem mass tag (TMT) probes containing an amine-reactive NHS-ester group, a spacer arm and an MS/MS reporter, are commonly used to label two to six peptide samples and measure relative protein expression levels with MS/MS (Thompson et al., 2003). Wang et al. integrated isobaric TMT labeling with chemoenzymatic enrichment to quantify O-GlcNAcylation between Alzheimer's diseased brain and normal brain tissues. They identified 530 O-GlcNAcylated proteins covering 1,094 O-GlcNAcylation sites in the brain. The O-GlcNAcylation levels of 81 proteins in the Alzheimer's patients brain were changed, indicating that dysregulation of O-GlcNAcylation may play an important role in the development of Alzheimer's disease (Wang et al., 2017).

IsoTaG-Based Quantitative Proteomics

Recently, an approach termed Isotope Targeted Glycoproteomics (IsoTaG) was developed by the Bertozzi group to enrich labeled glycopeptides and confidently profile the intact glycoproteome by MS (Woo et al., 2015). Specifically, glycoproteins were labeled metabolically with the azido functionality. Then, the soTaG silane Probe **1**, composed of an acid-cleavable biotin reagent containing an isotopic label and a terminal alkyne, was conjugated with the labeled glycoproteins. After capturing by streptavidin-agarose beads, and on-bead proteolytic digestion, the bound glycopeptides were released from the biotin tag and further sequenced by MS (**Figure 7B**). The quantification of the glycopeptides was achieved by using the pattern-searching algorithm mediated MS analysis to isotopically recoded species. IsoTaG shows a high sensitivity and repeatability when applied to low-abundance glycopeptides. Another strength of this method is that it promotes mass-independent

targeted database searching for high-confidence distribution. With this strategy, Woo et al. metabolically labeled O-GlcNAcylated proteins with $Ac_4GalNAz$ to explore O-GlcNAcylation alterations in response to T-cell activation. They found that more than 500 glycopeptides underwent significant changes during T cell activation, facilitating the functional understanding of O-GlcNAcylation in resting and activated primary human T cells (Woo et al., 2018). Inspired by the IsoTaG strategy, Qin et al. developed an acid-cleavable dialkoxydiphenylsilane (DADPS) linker (Probe **2**) to quantify O-GlcNAcylation. The isotopically labeled DADPS probe could be used to capture glycopeptides, which were then released after cleavage with mild acid and quantified by comparing the isotopic ratios using ETD-based tandem mass spectroscopy (Qin K. et al., 2018). Similarly, Li et al. designed an isotope-coded photocleavable probe for quantifying O-GlcNAcylation. O-GlcNAcylated proteins from two different cell states were chemoenzymatically captured by the Probe **3**. The linker was cleaved when exposed to ultraviolet light (365 nm). The released glycopeptides were further analyzed for sites mapping and relative quantification (Li et al., 2019). They found that compared with sorafenib-sensitive liver carcinoma cells, 55 glycopeptides in the sorafenib-resistant cells showed an increase in O-GlcNAcylation stoichiometry, suggesting a role of O-GlcNAcylation in regulating tumor chemoresistance. Taken together, IsoTaG-based quantitative O-GlcNAcylation proteomics strategy greatly facilitates the quantification of glycoproteins by installing isotopic tags directly onto the O-GlcNAc moiety. The isotopic labeling can be used as the dual function to improve the reliability of glycopeptide assignment.
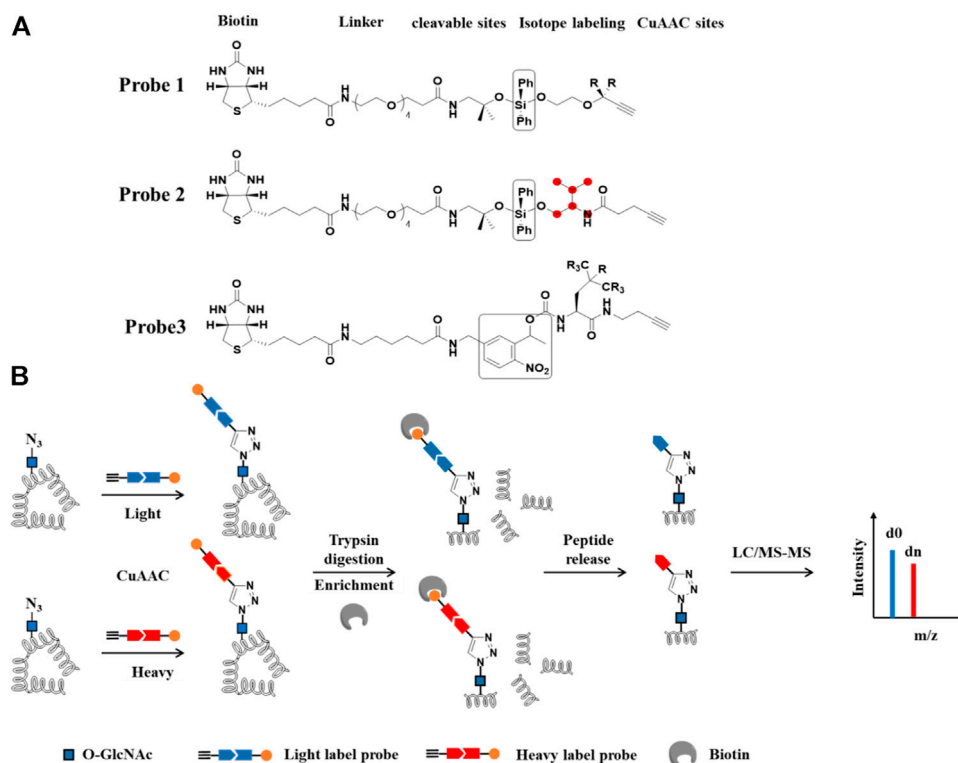
FIGURE 7 | IsoTaG-based quantitative proteomics for profiling of protein O-GlcNAc modification. **(A)** Structures of probes. R and red balls represent heavy or light labeling sites. Black box represents cleavable groups **(B)** Workflow of differential labeling and quantitative analysis of O-GlcNAcylated proteins. O-GlcNAcylated proteins first are labeled with azide group by Metabolic labeling or chemoenzymatic strategies. Isotope probes are added to O-GlcNAcylated proteins via CuAAC reaction, followed by incubation with streptavidin beads. The proteins on the beads are digested with trypsin and O-GlcNAcylated peptides are released by cleavable sites and analyzed by LC-MS/MS.

SUMMARY AND OUTLOOK

O-GlcNAcylation of intracellular proteins plays a fundamental role in health and disease. Effective ways to characterize the existence and dynamics of this modification will greatly promote the study of its functional significance. However, traditional methods, such as tritium labeling and the use of pan-O-GlcNAc antibodies, lack sensitivity and specificity. In addition, it is difficult to apply these methods to detect the changes and stoichiometry of O-GlcNAcylation in a complex system. The recent development of various chemical tools has provided exciting solutions to these problems. As stated above, a number of metabolic probes have been applied to detect and enrich O-GlcNAcylated proteins in living cells. A complementary chemoenzymatic labeling approach is also widely used to detect and profile O-GlcNAcylated proteins from cell lysates and tissues. Beyond the enrichment strategies, improvements in mass spectrometry technology have enabled quantifying and mapping O-GlcNAc sites with unprecedented accuracy.

O-GlcNAcylation is highly dynamic in response to nutrient availability and various environmental cues. With the variety of tools available to researchers, it is logical to profile and quantify O-GlcNAcylation under specific physiological contexts to reveal context-dependent functions of O-GlcNAc. In addition, current ways to modulate cellular O-GlcNAcylation rely on the use of small-molecule inhibitors or genetic knockdown/knockout, which lack the spatiotemporal resolution. Strategies that confer spatiotemporal control of O-GlcNAcylation are much needed. Moreover, O-GlcNAcylation on specific proteins has been shown to govern the protein function. Although there are some advances in developing strategies to manipulate protein-specific O-GlcNAcylation in cells (Gorelik and van Aalten, 2020; Ramirez et al., 2020; Ge et al., 2021), such studies are still in the infancy. In addition, the quantitative proteomics technique can't necessarily distinguish between changes in O-GlcNAc stoichiometry vs. changes in protein expression, which remains to be addressed. Nevertheless, we anticipate that further development of chemical tools will provide an important foundation for uncovering the functional importance of O-GlcNAcylation in the frontiers of biology and human health.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Aguilar, A. L., Hou, X., Wen, L., Wang, P. G., and Wu, P. (2017). A Chemoenzymatic Histology Method for O-GlcNAc Detection. *Chembiochem* 18, 2416–2421. doi:10.1002/cbic.201700515

Alfaro, J. F., Gong, C.-X., Monroe, M. E., Aldrich, J. T., Clauss, T. R. W., Purvine, S. O., et al. (2012). Tandem Mass Spectrometry Identifies many Mouse Brain O-GlcNAcylated Proteins Including EGF Domain-specific O-GlcNAc Transferase Targets. *Proc. Natl. Acad. Sci.* 109, 7280–7285. doi:10.1073/pnas.1200425109

Best, M. D. (2009). Click Chemistry and Bioorthogonal Reactions: Unprecedented Selectivity in the Labeling of Biological Molecules. *Biochemistry* 48, 6571–6584. doi:10.1021/bi9007726

Bond, M. R., and Hanover, J. A. (2015). A Little Sugar Goes a Long Way: The Cell Biology of O-GlcNAc. *J. Cell Biol* 208, 869–880. doi:10.1083/jcb.201501101

Boyce, M., Carrico, I. S., Ganguli, A. S., Yu, S.-H., Hangauer, M. J., Hubbard, S. C., et al. (2011). Metabolic Cross-Talk Allows Labeling of O-Linked -N-Acetylglucosamine-Modified Proteins via the N-Acetylgalactosamine Salvage Pathway. *Proc. Natl. Acad. Sci.* 108, 3141–3146. doi:10.1073/pnas.1010045108

Cecioni, S., and Vocadlo, D. J. (2013). Tools for Probing and Perturbing O-GlcNAc in Cells and *In Vivo*. *Curr. Opin. Chem. Biol.* 17, 719–728. doi:10.1016/j.cbpa.2013.06.030

Chalkley, R. J., Thalhammer, A., Schoepfer, R., and Burlingame, A. L. (2009). Identification of Protein O-GlcNAcylation Sites Using Electron Transfer Dissociation Mass Spectrometry on Native Peptides. *Proc. Natl. Acad. Sci. USA* 106, 8894–8899. doi:10.1073/pnas.0900288106

Chen, J., Dong, X., Cheng, X., Zhu, Q., Zhang, J., Li, Q., et al. (2021). Ogt Controls Neural Stem/progenitor Cell Pool and Adult Neurogenesis through Modulating Notch Signaling. *Cell Rep.* 34, 108905. doi:10.1016/j.celrep.2021.108905

Chen, X., Wei, S., Ji, Y., Guo, X., and Yang, F. (2015). Quantitative Proteomics Using SILAC: Principles, Applications, and Developments. *Proteomics* 15, 3175–3192. doi:10.1002/pmic.201500108

Chuh, K. N., Batt, A. R., and Pratt, M. R. (2016). Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chem. Biol.* 23, 86–107. doi:10.1016/j.chembiol.2015.11.006

Chuh, K. N., and Pratt, M. R. (2015). Chemical Methods for the Proteome-wide Identification of Posttranslationally Modified Proteins. *Curr. Opin. Chem. Biol.* 24, 27–37. doi:10.1016/j.cbpa.2014.10.020

Chuh, K. N., Zaro, B. W., Piller, F., Piller, V., and Pratt, M. R. (2014). Changes in Metabolic Chemical Reporter Structure Yield a Selective Probe of O-GlcNAc Modification. *J. Am. Chem. Soc.* 136, 12283–12295. doi:10.1021/ja504063c

Cieniewski-Bernard, C., Bastide, B., Lefebvre, T., Lemoine, J., Mounier, Y., and Michalski, J.-C. (2004). Identification of O-Linked N-Acetylglucosamine Proteins in Rat Skeletal Muscle Using Two-Dimensional Gel Electrophoresis and Mass Spectrometry. *Mol. Cell Proteomics* 3, 577–585. doi:10.1074/mcp.M400024-MCP200

Cieniewski-Bernard, C., Dupont, E., Deracinois, B., Lambert, M., and Bastide, B. (2014). Multiplexed Detection of O-GlcNAcome, Phosphoproteome, and Whole Proteome within the Same Gel. *Front. Endocrinol.* 5, 184. doi:10.3389/fendo.2014.00184

Clark, P. M., Dweck, J. F., Mason, D. E., Hart, C. R., Buck, S. B., Peters, E. C., et al. (2008). Direct In-Gel Fluorescence Detection and Cellular Imaging of O-GlcNAc-Modified Proteins. *J. Am. Chem. Soc.* 130, 11576–11577. doi:10.1021/ja8030467

Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A., and Hart, G. W. (2001). Characterization of a Mouse Monoclonal Antibody Specific for O-Linked N-Acetylglucosamine. *Anal. Biochem.* 293, 169–177. doi:10.1006/abio.2001.5132

Darabedian, N., Thompson, J. W., Chuh, K. N., Hsieh-Wilson, L. C., and Pratt, M. R. (2018). Optimization of Chemoenzymatic Mass Tagging by Strain-Promoted Cycloaddition (SPAAC) for the Determination of O-GlcNAc Stoichiometry by Western Blotting. *Biochemistry* 57, 5769–5774. doi:10.1021/acs.biochem.8b00648

Duan, F., Wu, H., Jia, D., Wu, W., Ren, S., Wang, L., et al. (2018). O -GlcNAcylation of RACK1 Promotes Hepatocellular Carcinogenesis. *J. Hepatol.* 68, 1191–1202. doi:10.1016/j.jhep.2018.02.003

Gao, Y., Wells, L., Comer, F. I., Parker, G. J., and Hart, G. W. (2001). Dynamic O-Glycosylation of Nuclear and Cytosolic Proteins. *J. Biol. Chem.* 276, 9838–9845. doi:10.1074/jbc.M010420200

Ge, Y., Ramirez, D. H., Yang, B., D'Souza, A. K., Aonbangkhen, C., Wong, S., et al. (2021). Target Protein Deglycosylation in Living Cells by a Nanobody-Fused Split O-GlcNAcase. *Nat. Chem. Biol.* 17, 593–600. doi:10.1038/s41589-021-00757-y

Gorelik, A., and van Aalten, D. M. F. (2020). Tools for Functional Dissection of Site-specific O-GlcNAcylation. *RSC Chem. Biol.* 1, 98–109. doi:10.1039/D0CB00052C

Grammel, M., and Hang, H. C. (2013). Chemical Reporters for Biological Discovery. *Nat. Chem. Biol.* 9, 475–484. doi:10.1038/nchembio.1296

Greis, K. D., Hayes, B. K., Comer, F. I., Kirk, M., Barnes, S., Lowary, T. L., et al. (1996). Selective Detection and Site-Analysis of O-GlcNAc-Modified Glycopeptides by β-Elimination and Tandem Electrospray Mass Spectrometry. *Anal. Biochem.* 234, 38–49. doi:10.1006/abio.1996.0047

Griffin, M. E., Jensen, E. H., Mason, D. E., Jenkins, C. L., Stone, S. E., Peters, E. C., et al. (2016). Comprehensive Mapping of O-GlcNAc Modification Sites Using a Chemically Cleavable Tag. *Mol. Biosyst.* 12, 1756–1759. doi:10.1039/c6mb00138f

Hahne, H., Sobotzki, N., Nyberg, T., Helm, D., Borodkin, V. S., van Aalten, D. M. F., et al. (2013). Proteome Wide Purification and Identification of O-GlcNAc-Modified Proteins Using Click Chemistry and Mass Spectrometry. *J. Proteome Res.* 12, 927–936. doi:10.1021/pr300967y

Haltiwanger, R. S., Holt, G. D., and Hart, G. W. (1990). Enzymatic Addition of O-GlcNAc to Nuclear and Cytoplasmic Proteins. Identification of a Uridine Diphospho-N-Acetylglucosamine:peptide Beta-N-Acetylglucosaminyltransferase. *J. Biol. Chem.* 265, 2563–2568. doi:10.1016/s0021-9258(19)39838-2

Hao, Y., Fan, X., Shi, Y., Zhang, C., Sun, D.-e., Qin, K., et al. (2019). Next-generation Unnatural Monosaccharides Reveal that ESRRB O-GlcNAcylation Regulates Pluripotency of Mouse Embryonic Stem Cells. *Nat. Commun.* 10, 4065. doi:10.1038/s41467-019-11942-y

Hart, G. W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. (2011). Cross Talk between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* 80, 825–858. doi:10.1146/annurev-biochem-060608-102511

Hayes, B. K., Greis, K. D., and Hart, G. W. (1995). Specific Isolation of O-Linked N-Acetylglucosamine Glycopeptides from Complex Mixtures. *Anal. Biochem.* 228, 115–122. doi:10.1006/abio.1995.1322

Jang, H., Kim, T. W., Yoon, S., Choi, S.-Y., Kang, T.-W., Kim, S.-Y., et al. (2012). O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. *Cell stem cell* 11, 62–74. doi:10.1016/j.stem.2012.03.001

Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., et al. (2003). A Chemoenzymatic Approach toward the Rapid and Sensitive Detection of O-GlcNAc Posttranslational Modifications. *J. Am. Chem. Soc.* 125, 16162–16163. doi:10.1021/ja038545r

Khidekel, N., Ficarro, S. B., Clark, P. M., Bryan, M. C., Swaney, D. L., Rexach, J. E., et al. (2007). Probing the Dynamics of O-GlcNAc Glycosylation in the Brain Using Quantitative Proteomics. *Nat. Chem. Biol.* 3, 339–348. doi:10.1038/nchembio881

Khidekel, N., Ficarro, S. B., Peters, E. C., and Hsieh-Wilson, L. C. (2004). Exploring the O-GlcNAc Proteome: Direct Identification of O-GlcNAc-Modified Proteins from the Brain. *Proc. Natl. Acad. Sci.* 101, 13132–13137. doi:10.1073/pnas.0403471101

Kochibe, N., and Matta, K. L. (1989). Purification and Properties of an N-acetylglucosamine-specific Lectin from Psathyrella Velutina Mushroom. *J. Biol. Chem.* 264, 173–177. doi:10.1016/s0021-9258(17)31239-5

Lefebvre, T., Baert, f. d. r., Bodart, J.-F. o., Flament, S. p., Michalski, J.-C., and Vilain, J.-P. (2004). Modulation of O-GlcNAc Glycosylation duringXenopus Oocyte Maturation. *J. Cell. Biochem.* 93, 999–1010. doi:10.1002/jcb.20242

Li, J., Li, Z., Duan, X., Qin, K., Dang, L., Sun, S., et al. (2019). An Isotope-Coded Photocleavable Probe for Quantitative Profiling of Protein O-GlcNAcylation. *ACS Chem. Biol.* 14, 4–10. doi:10.1021/acschembio.8b01052

Li, J., Wang, J., Wen, L., Zhu, H., Li, S., Huang, K., et al. (2016). An OGA-Resistant Probe Allows Specific Visualization and Accurate Identification of O-GlcNAc-Modified Proteins in Cells. *ACS Chem. Biol.* 11, 3002–3006. doi:10.1021/acschembio.6b00678

Li, Z., and Yi, W. (2014). Regulation of Cancer Metabolism by O-GlcNAcylation. *Glycoconj J.* 31, 185–191. doi:10.1007/s10719-013-9515-5

Lubas, W. A., Frank, D. W., Krause, M., and Hanover, J. A. (1997). O-linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats. *J. Biol. Chem.* 272, 9316–9324. doi:10.1074/jbc.272.14.9316

Lund, P. J., Elias, J. E., and Davis, M. M. (2016). Global Analysis of O-GlcNAc Glycoproteins in Activated Human T Cells. *J. Immunol.* 197, 3086–3098. doi:10.4049/jimmunol.1502031

Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D., and Vocadlo, D. J. (2005). O-GlcNAcase Uses Substrate-Assisted Catalysis. *J. Biol. Chem.* 280, 25313–25322. doi:10.1074/jbc.M413819200

Machon, O., Baldini, S. F., Ribeiro, J. P., Steenackers, A., Varrot, A., Lefebvre, T., et al. (2017). Recombinant Fungal Lectin as a New Tool to investigateO-GlcNAcylation Processes. *Glycobiology* 27, 123–128. doi:10.1093/glycob/cww105

Monsigny, M., Sene, C., Obrenovitch, A., Roche, A.-C., Delmotte, F., and Boschetti, E. (1979). Properties of Succinylated Wheat-Germ Agglutinin. *Eur. J. Biochem.* 98, 39–45. doi:10.1111/j.1432-1033.1979.tb13157.x

Myers, S. A., Panning, B., and Burlingame, A. L. (2011). Polycomb Repressive Complex 2 Is Necessary for the normal Site-specific O-GlcNAc Distribution in Mouse Embryonic Stem Cells. *Proc. Natl. Acad. Sci.* 108, 9490–9495. doi:10.1073/pnas.1019289108

Nakanuma, Y., Sasaki, M., and Kono, N. (1993). Succinylated Wheat Germ Agglutinin Lectin Binding in Intrahepatic Vessels. A New Histochemical Tool. *Arch. Pathol. Lab. Med.* 117, 809–811. doi:10.1016/0264-410X(93)90064-5

Nishikaze, T., Kawabata, S.-I., Iwamoto, S., and Tanaka, K. (2013). Reversible Hydrazide Chemistry-Based Enrichment for O-GlcNAc-Modified Peptides and Glycopeptides Having Non-reducing GlcNAc Residues. *Analyst* 138, 7224–7232. doi:10.1039/c3an00880k

Ong, Q., Han, W., and Yang, X. (2018). O-GlcNAc as an Integrator of Signaling Pathways. *Front. Endocrinol.* 9, 599. doi:10.3389/fendo.2018.00599

Prescher, J. A., and Bertozzi, C. R. (2005). Chemistry in Living Systems. *Nat. Chem. Biol.* 1, 13–21. doi:10.1038/nchembio0605-13

Qin, K., Zhang, H., Zhao, Z., and Chen, X. (2020). Protein S-Glyco-Modification through an Elimination-Addition Mechanism. *J. Am. Chem. Soc.* 142, 9382–9388. doi:10.1021/jacs.0c02110

Qin, K., Zhu, Y., Qin, W., Gao, J., Shao, X., Wang, Y.-l., et al. (2018). Quantitative Profiling of Protein O-GlcNAcylation Sites by an Isotope-Tagged Cleavable Linker. *ACS Chem. Biol.* 13, 1983–1989. doi:10.1021/acschembio.8b00414

Qin, W., Lv, P., Fan, X., Quan, B., Zhu, Y., Qin, K., et al. (2017). Quantitative Time-Resolved Chemoproteomics Reveals that Stable O-GlcNAc Regulates Box C/D snoRNP Biogenesis. *Proc. Natl. Acad. Sci. USA* 114, E6749–E6758. doi:10.1073/pnas.1702688114

Qin, W., Qin, K., Fan, X., Peng, L., Hong, W., Zhu, Y., et al. (2018). Artificial Cysteine S-Glycosylation Induced by Per-O-Acetylated Unnatural Monosaccharides during Metabolic Glycan Labeling. *Angew. Chem. Int. Ed.* 57, 1817–1820. doi:10.1002/anie.201711710

Ramirez, D. H., Aonbangkhen, C., Wu, H.-Y., Naftaly, J. A., Tang, S., O'Meara, T. R., et al. (2020). Engineering a Proximity-Directed O-GlcNAc Transferase for Selective Protein O-GlcNAcylation in Cells. *ACS Chem. Biol.* 15, 1059–1066. doi:10.1021/acschembio.0c00074

Ren, X., Jiang, S., Li, D., Sun, H., and Wang, D. (2013). Crystallization and Preliminary Crystallographic Studies of AAL-2, a Novel Lectin fromAgrocybe Aegeritathat Binds Nonreducing terminalN-Acetylglucosamine. *Acta Cryst. Sect F* 69, 650–652. doi:10.1107/S1744309113011639

Rexach, J. E., Rogers, C. J., Yu, S.-H., Tao, J., Sun, Y. E., and Hsieh-Wilson, L. C. (2010). Quantification of O-Glycosylation Stoichiometry and Dynamics Using Resolvable Mass Tags. *Nat. Chem. Biol.* 6, 645–651. doi:10.1038/nchembio.412

Rouhanifard, S. H., López-Aguilar, A., and Wu, P. (2014). CHoMP: A Chemoenzymatic Histology Method Using Clickable Probes. *ChemBioChem.* 15, 2667–2673. doi:10.1002/cbic.201402433

Sakabe, K., Wang, Z., and Hart, G. W. (2010). -N-acetylglucosamine (O-GlcNAc) Is Part of the Histone Code. *Proc. Natl. Acad. Sci.* 107, 19915–19920. doi:10.1073/pnas.1009023107

Slawson, C., and Hart, G. W. (2011). O-GlcNAc Signalling: Implications for Cancer Cell Biology. *Nat. Rev. Cancer* 11, 678–684. doi:10.1038/nrc3114

Snow, C. M., Senior, A., and Gerace, L. (1987). Monoclonal Antibodies Identify a Group of Nuclear Pore Complex Glycoproteins. *J. Cell Biol.* 104, 1143–1156. doi:10.1083/jcb.104.5.1143

Tan, H. Y., Eskandari, R., Shen, D., Zhu, Y., Liu, T W., Willems, L. I., et al. (2018). Direct One-Step Fluorescent Labeling of O-GlcNAc-Modified Proteins in Live Cells Using Metabolic Intermediates. *J. Am. Chem. Soc.* 140, 15300–15308. doi:10.1021/jacs.8b08260

Tan, W., Jiang, P., Zhang, W., Hu, Z., Lin, S., Chen, L., et al. (2021). Posttranscriptional Regulation of De Novo Lipogenesis by Glucose-Induced O-GlcNAcylation. *Mol. Cell* 81, 1890–1904.e7. doi:10.1016/j.molcel.2021.02.009

Tashima, Y., and Stanley, P. (2014). Antibodies that Detect O-Linked β-d-N-Acetylglucosamine on the Extracellular Domain of Cell Surface Glycoproteins. *J. Biol. Chem.* 289, 11132–11142. doi:10.1074/jbc.M113.492512

Teo, C. F., Ingale, S., Wolfert, M. A., Elsayed, G. A., Nöt, L. G., Chatham, J. C., et al. (2010). Glycopeptide-specific Monoclonal Antibodies Suggest New Roles for O-GlcNAc. *Nat. Chem. Biol.* 6, 338–343. doi:10.1038/nchembio.338

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003). Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904. doi:10.1021/ac0262560

Thompson, J. W., Griffin, M. E., and Hsieh-Wilson, L. C. (2018). Methods for the Detection, Study, and Dynamic Profiling of O-GlcNAc Glycosylation. *Method Enzymol.* 598, 101–135. doi:10.1016/bs.mie.2017.06.009

Torres, C. R., and Hart, G. W. (1984). Topography and Polypeptide Distribution of Terminal N-Acetylglucosamine Residues on the Surfaces of Intact Lymphocytes. Evidence for O-Linked GlcNAc. *J. Biol. Chem.* 259, 3308–3317. doi:10.1016/s0021-9258(17)43295-9

Tsumoto, H., Akimoto, Y., Endo, T., and Miura, Y. (2017). Quantitative Analysis of O -GlcNAcylation in Combination with Isobaric Tag Labeling and Chemoenzymatic Enrichment. *Bioorg. Med. Chem. Lett.* 27, 5022–5026. doi:10.1016/j.bmcl.2017.10.005

Tsumoto, H., Ogasawara, D., Hashii, N., Suzuki, T., Akimoto, Y., Endo, T., et al. (2015). Enrichment of O-GlcNAc-Modified Peptides Using Novel Thiol-Alkyne and Thiol-Disulfide Exchange. *Bioorg. Med. Chem. Lett.* 25, 2645–2649. doi:10.1016/j.bmcl.2015.04.082

Turner, J. R., Tartakoff, A. M., and Greenspan, N. S. (1990). Cytologic Assessment of Nuclear and Cytoplasmic O-Linked N-Acetylglucosamine Distribution by Using Anti-streptococcal Monoclonal Antibodies. *Proc. Natl. Acad. Sci.* 87, 5608–5612. doi:10.1073/pnas.87.15.5608

Vaidyanathan, K., and Wells, L. (2014). Multiple Tissue-specific Roles for the O-GlcNAc Post-translational Modification in the Induction of and Complications Arising from Type II Diabetes. *J. Biol. Chem.* 289, 34466–34471. doi:10.1074/jbc.R114.591560

Vercoutter-Edouart, A.-S., El Yazidi-Belkoura, I., Guinez, C., Baldini, S., Leturcq, M., Mortuaire, M., et al. (2015). Detection and Identification ofO-GlcNAcylated Proteins by Proteomic Approaches. *Proteomics* 15, 1039–1050. doi:10.1002/pmic.201400326

Vocadlo, D. J., Hang, H. C., Kim, E.-J., Hanover, J. A., and Bertozzi, C. R. (2003). A Chemical Approach for Identifying O-GlcNAc-Modified Proteins in Cells. *Proc. Natl. Acad. Sci.* 100, 9116–9121. doi:10.1073/pnas.1632821100

Vosseller, K., Trinidad, J. C., Chalkley, R. J., Specht, C. G., Thalhammer, A., Lynn, A. J., et al. (2006). O-Linked N-Acetylglucosamine Proteomics of Postsynaptic Density Preparations Using Lectin Weak Affinity Chromatography and Mass Spectrometry. *Mol. Cell Proteomics* 5, 923–934. doi:10.1074/mcp.T500040-MCP200

Wang, S., Yang, F., Petyuk, V. A., Shukla, A. K., Monroe, M. E., Gritsenko, M. A., et al. (2017). Quantitative Proteomics Identifies Altered O-GlcNAcylation of Structural, Synaptic and Memory-associated Proteins in Alzheimer's Disease. *J. Pathol.* 243, 78–88. doi:10.1002/path.4929

Wang, Z., Pandey, A., and Hart, G. W. (2007). Dynamic Interplay between O-Linked N-Acetylglucosaminylation and Glycogen Synthase Kinase-3-dependent Phosphorylation. *Mol. Cell Proteomics* 6, 1365–1379. doi:10.1074/mcp.M600453-MCP200

Wang, Z., Park, K., Comer, F., Hsieh-Wilson, L. C., Saudek, C. D., and Hart, G. W. (2009). Site-Specific GlcNAcylation of Human Erythrocyte Proteins: Potential Biomarker(s) for Diabetes. *Diabetes* 58, 309–317. doi:10.2337/db08-0994

Wang, Z., Udeshi, N. D., O'Malley, M., Shabanowitz, J., Hunt, D. F., and Hart, G. W. (2010a). Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell Proteomics* 9, 153–160. doi:10.1074/mcp.M900268-MCP200

Wang, Z., Udeshi, N. D., Slawson, C., Compton, P. D., Sakabe, K., Cheung, W. D., et al. (2010b). Extensive Crosstalk between O-GlcNAcylation and Phosphorylation Regulates Cytokinesis. *Sci. Signaling* 3, ra2. doi:10.1126/scisignal.2000526

Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002). Mapping Sites of O-GlcNAc Modification Using Affinity Tags for Serine and Threonine post-translational Modifications. *Mol. Cell Proteomics* 1, 791–804. doi:10.1074/mcp.m200048-mcp200

Whelan, S. A., and Hart, G. W. (2006). Identification of O-GlcNAc Sites on Proteins. *Methods Enzymol.* 415, 113–133. doi:10.1016/S0076-6879(06)15008-9

Woo, C. M., Iavarone, A. T., Spiciarich, D. R., Palaniappan, K. K., and Bertozzi, C. R. (2015). Isotope-targeted Glycoproteomics (IsoTaG): a Mass-independent Platform for Intact N- and O-Glycopeptide Discovery and Analysis. *Nat. Methods* 12, 561–567. doi:10.1038/nmeth.3366

Woo, C. M., Lund, P. J., Huang, A. C., Davis, M. M., Bertozzi, C. R., and Pitteri, S. J. (2018). Mapping and Quantification of over 2000 O-Linked Glycopeptides in Activated Human T Cells with Isotope-Targeted Glycoproteomics (Isotag). *Mol. Cell Proteomics* 17, 764–775. doi:10.1074/mcp.RA117.000261

Yoshida, N., Mortara, R. A., Araguth, M. F., Gonzalez, J. C., and Russo, M. (1989). Metacyclic Neutralizing Effect of Monoclonal Antibody 10D8 Directed to the 35- and 50-kilodalton Surface Glycoconjugates of Trypanosoma Cruzi. *Infect. Immun.* 57, 1663–1667. doi:10.1128/IAI.57.6.1663-1667.1989

Yuzwa, S. A., and Vocadlo, D. J. (2014). O-GlcNAc and Neurodegeneration: Biochemical Mechanisms and Potential Roles in Alzheimer's Disease and beyond. *Chem. Soc. Rev.* 43, 6839–6858. doi:10.1039/c4cs00038b

Zachara, N. E. (2009). Detecting the "O-GlcNAcome"; Detection, Purification, and Analysis of O-GlcNAc Modified Proteins. *Methods Mol. Biol.* 534, 250–279. doi:10.1007/978-1-59745-022-5_19

Zachara, N. E., Molina, H., Wong, K. Y., Pandey, A., and Hart, G. W. (2011). The Dynamic Stress-Induced "O-GlcNAc-Ome" Highlights Functions for O-GlcNAc in Regulating DNA Damage/repair and Other Cellular Pathways. *Amino acids* 40, 793–808. doi:10.1007/s00726-010-0695-z

Zaro, B. W., Batt, A. R., Chuh, K. N., Navarro, M. X., and Pratt, M. R. (2017). The Small Molecule 2-Azido-2-Deoxy-Glucose Is a Metabolic Chemical Reporter of O-GlcNAc Modifications in Mammalian Cells, Revealing an Unexpected Promiscuity of O-GlcNAc Transferase. *ACS Chem. Biol.* 12, 787–794. doi:10.1021/acschembio.6b00877

Zaro, B. W., Yang, Y.-Y., Hang, H. C., and Pratt, M. R. (2011). Chemical Reporters for Fluorescent Detection and Identification of O-GlcNAc-Modified Proteins Reveal Glycosylation of the Ubiquitin Ligase NEDD4-1. *Proc. Natl. Acad. Sci.* 108, 8146–8151. doi:10.1073/pnas.1102458108

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Quantification of Intact O-Glycopeptides on Haptoglobin in Sera of Patients With Hepatocellular Carcinoma and Liver Cirrhosis

Hong Shu[1,2†], Lei Zhang[3†], Yiwei Chen[4], Yijie Guo[5], Limin Li[6], Fanghua Chen[1], Zhao Cao[6], Guoquan Yan[3], Chunlai Lu[4], Chao Liu[5]* and Shu Zhang[1]*

[1]Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion (Ministry of Education), Fudan University, Shanghai, China, [2]Department of Clinical Laboratory, Cancer Hospital of Guangxi Medical University, Nanning, China, [3]Institutes of Biomedical Sciences, Fudan University, Shanghai, China, [4]Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China, [5]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing, China, [6]Department of Clinical Laboratory, First Affiliated Hospital of Guangxi Medical University, Nanning, China

Haptoglobin (Hp) is one of the acute-phase response proteins secreted by the liver, and its aberrant N-glycosylation was previously reported in hepatocellular carcinoma (HCC). Limited studies on Hp O-glycosylation have been previously reported. In this study, we aimed to discover and confirm its O-glycosylation in HCC based on lectin binding and mass spectrometry (MS) detection. First, serum Hp was purified from patients with liver cirrhosis (LC) and HCC, respectively. Then, five lectins with Gal or GalNAc monosaccharide specificity were chosen to perform lectin blot, and the results showed that Hp in HCC bound to these lectins in a much stronger manner than that in LC. Furthermore, label-free quantification based on MS was performed. A total of 26 intact O-glycopeptides were identified on Hp, and most of them were elevated in HCC as compared to LC. Among them, the intensity of HYEG$\underline{S}^{316}$TVPEK (H1N1S1) on Hp was the highest in HCC patients. Increased HYEG$\underline{S}^{316}$TVPEK (H1N1S1) in HCC was quantified and confirmed using the MS method based on $^{18}O/^{16}O$ C-terminal labeling and multiple reaction monitoring. This study provided a comprehensive understanding of the glycosylation of Hp in liver diseases.

Keywords: haptoglobin, O-glycosylation, mass spectrometry, lectin, hepatocellular carcinoma

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related deaths, especially in China. The risk of HCC development depends on many factors such as hepatitis B virus (HBV) infection, and it causes liver complications and liver cirrhosis (Yang et al., 2019). Alpha-fetoprotein (AFP) is the most widely used diagnostic serum biomarker for HCC, however, it has low sensitivity. Glycosylation was reported to be related to most serum tumor biomarkers (Zhu et al., 2019a; Yang et al., 2019). In our previous study, the fucosylated N-glycans of haptoglobin (Hp) were found to be increased in HCC patients (Zhang et al., 2011; Zhang et al., 2013).

As an acute-phase protein mainly synthesized in the liver, Hp comprises heavy chains (β,~ 40 kDa) and light chains (α_1, ~ 9.1 kDa)/(α_2, ~ 16 kDa). Its function is to bind and transport

TABLE 1 | Clinical characteristics of liver cirrhosis and HCC patients.

Group	Set 1 (lectin; ^{16}O/^{18}O labeling)		Set 2 (label free)		Set 3 (MRM)	
	LC	HCC	LC	HCC	LC	HCC
Number	60	60	7	7	12	12
Age	47.40 ± 11.73	49.65 ± 11.55	48.57 ± 8.02	47.4 ± 10.27	50.13 ± 8.55	51.0 ± 10.29
Gender (M/F)	50/10	55/5	6/1	6/1	8/4	11/1
ALT, U/L	30 (5–216)	42 (12–428)	32 (17–121)	36 (15–199)	27 (17–121)	53 (15–199)
AST, U/L	35 (8–182)	38 (15–355)	33 (19–145)	50 (24–220)	34 (19–145)	43 (24–220)
HBsAg^{+}(%)[a]	100	100	100	100	100	100
AFP, ng/mL	6.00 (1.2–65.1)	153.3 (1.6–60500.0)	3.86 (1.35–170.0)	94.1 (1.41–60500.0)	3.26 (1.35–170)	73.76 (1.41–60500)

ALT, alanine aminotransferase; AST, aspartate transaminase; HbsAg, hepatitis B surface antigen; AFP, alpha fetoprotein.
[a]Normal results of HBsAg+ (%) are negative or nonreactive, meaning that no HBsAg was found; if the test is positive or reactive, it may mean active infection with HBV or chronic hepatitis B infection.

free hemoglobin to degrade and recycle iron in the liver (Zhang et al., 2016). Hp contains four N-glycosylation sites, and most of the studies focused on its N-glycosylation, for example, N-glycosylation site occupancy and site-specific N-glycoforms in liver diseases (Zhang et al., 2012). However, limited studies on O-glycans of Hp have been reported. The presence of mono- and disialyl core type 1 O-glycans was reported in prostate cancer Hp (Fujimura et al., 2008).

Development in mass spectrometry (MS) and software has enabled the characterization and quantification of intact glycopeptides in complex biological matrixes. Zhu et al. have identified and quantified N-glycopeptides of Hp based on electron-transfer higher-energy collision dissociation (ET-hcD), MS/MS fragmentation, and Byonic software (Zhu et al., 2019b). We have previously used a glycopeptide method based on ^{18}O/^{16}O C-terminal labeling and multiple reaction monitoring (MRM) to quantify N-glycopeptides of the 40-kDa band in liver diseases (Zhang et al., 2019).

For this study, serum Hp was purified from patients with HCC and liver cirrhosis (LC), respectively. First, five lectins with the specificity to recognize Gal or GalNAc monosaccharides were used to determine whether Hp could bind. Furthermore, label-free quantification for intact glycopeptide was performed to discover O-glycosylation on Hp. The potential O-glycosylation features were confirmed using the MS quantification method based on ^{18}O/^{16}O C-terminal labeling and MRM.

MATERIALS AND METHODS

Preparation of Specimens and Purification of Serum Haptoglobin

The serum samples were obtained from the Cancer Hospital of Guangxi Medical University and approved by the Institution Ethics Committee of the Cancer Hospital of Guangxi Medical University (LW2019043). All participants had signed an informed consent form. A summary of pathological patients' data is given in **Table 1**. Individuals with autoimmune diseases were excluded from this study. All serum samples were collected using a standard protocol and stored at −80°C until use. For lectin blot analysis and MS quantification based on ^{18}O/^{16}O C-terminal

labeling, pooled serum samples were used. For label-free MS detection and MRM, individual serum samples were chosen. The purification of Hp was performed according to previous reports (Okuyama et al., 2006).

Lectin Blot Analysis

Purified Hp of 60 μl of sera (pooled from 20 patients, 3 μl per patient) from LC and HCC, respectively, was analyzed by lectin blot according to the previous description (Shu et al., 2011). Five biotinylated lectins (Vector Laboratories, Burlingame, CA) including *Amaranthus caudatus* lectin (ACA, 1 μg/ml), *Griffonia simplicifolia* lectin I (GSI-L, 1 μg/ml), Jacalin (JAC, 1 μg/ml), *Vicia villosa* lectin (VVA, 1 μg/ml), and *Wisteria floribunda* lectin (WFA, 2 μg/ml) were incubated for 45 min, respectively. After washing and incubating with horseradish peroxidase Avidin D (Vector Laboratories, Burlingame, CA), the bands were developed using chemiluminescence detection reagents (GE Healthcare, Piscataway, NJ).

In-Gel Digestion and O-Glycopeptide Enrichment

Purified Hp was obtained from an equal volume of each sample and subjected to SDS-PAGE. The bands of the $α_1$, $α_2$, and $β$ chains were excised, destained, reduced, and alkylated. PNGase F treatment (enzyme:substrate = 1:20, v/w) was performed to release N-glycans from the gel at 37°C overnight. Then, the gel particles were incubated in sequencing grade trypsin (Promega, Madison, WI) at 37°C overnight for in-gel digestion (enzyme: substrate = 1:50, w/w). Subsequently, the peptides were extracted three times with 50% ACN and 0.1% trifluoroacetic acid and lyophilized. The O-glycopeptides were enriched in accordance with the manufacturer's protocol for the glycopeptide enrichment kit (Novagen, Darmstadt, Germany).

Nano-Liquid Chromatography Tandem Mass Spectrometry

The experiments were performed using an Orbitrap Fusion MS (Thermo Fisher Scientific, Bremen, Germany). Purified Hp from 30 μl sera of LC and HCC, respectively, was used as starting material. Freeze-dried peptides were redissolved with 10 μl

solvent A [solvent A: 0.1% formic acid (FA) and water solution], and were directly injected into the analytical column (Acclaim PepMap C18, 75 µm × 25 cm). To achieve sufficient separation, the samples were eluted for 90 min with the following linear gradient: 1% solvent B (solvent B: 0.1% FA in ACN) to 25% mobile phase B in 60 min, from 25 to 45% mobile phase B in 20 min, followed by an increase to 90% mobile phase B in 1 min, which is maintained for 3 min and finally re-equilibrated for 6 min at 1% B. The flow rate of the column was 300 nL/min, and the column temperature was maintained at RT. Ions with charge states between 2 + and 6+ were sequentially fragmented by HCD with a stepped collision energy of 20, 30, and 40%.

O-Glycopeptide Identification Using Byonic

The raw files were searched using Byonic (Protein Metrics, San Carlos, CA) against the fasta-containing sp|P00738| HPT_HUMAN (the UniProt human haptoglobin database) due to the searching capacity. A human O-glycan database that included 70 human O-glycans was used for searching. Specific parameters were as follows: the tolerances were 20 ppm for precursors and 0.05 Da for fragment ions, respectively. Up to two missed cleavages were allowed. The Carbamidomethyl/ +57.022 Da of C was set as a fixed modification, while Oxidation/+15.995 Da of M and Deamidated/+0.98 Da of N were set as the variable modifications. Quality control was performed by FDR estimation, and those below the criterion of a 1% FDR were considered. The mass range for the search was set between 350 and 5,000 Da. The predicted O-glycosylation sites were also considered. The NetOGlyc 4.0 server was used to predict O-glycosylation sites on Hp in humans (http://www.cbs.dtu.dk/ services/NetOGlyc). Identification of MS/MS spectra was presented using the gLabel software tool (Liu et al., 2017), which contributes toward manual verification.

Label-Free Quantification for O-Glycopeptide

For the MS^1 level-based label-free method, the normalization was performed using pQuant (Liu et al., 2014). It first obtains the original intensity matrix X (**Eq. 1**) and then calculates the normalized intensity according to each sample. For each intensity $X_{i,j}$, pQuant divides it by the mean intensity of the corresponding sample and calculates the logarithm to get X_mean_process (**Eq. 2**). Then, the median intensity of each sample processed by the mean value is subtracted from the intensity just calculated, and the intensity of the whole sample is shifted so that the median is 0 (**Eq. 3**). Finally, the normalized intensity is obtained as follows:

$$X = (x_1, x_2, \ldots, x_n), n = sample\ number, \quad (1)$$

$$x\_mean\_process_{i,j} = \log\left(x_{i,j} \Big/ \frac{1}{m} \times \sum_{l=1}^{m} x_{i,l}\right), \quad (2)$$

m = peptide number,

$$Normalized\ Intensity\ x_{i,j} = x\_mean\_process_{i,j}$$
$$- median\left(x\_mean\_process_{,i}\right). \quad (3)$$

After normalization, the intensities of intact glycopeptide were calculated using the formula; for example, intensity of O-glycopeptide 1 in LC1 = (normalized intensity of O-glycopeptide 1 in LC1) × adjust ratio = (normalized intensity of O-glycopeptide 1 in LC1) × [(normalized intensity of O-glycopeptide 1 in LC1)/(all normalized intensities of O-glycopeptides in LC1)].

O-Glycopeptide Quantification Using ^{18}O/^{16}O Labeling

Purified Hp of 300 µl sera (pooled from 20 patients, 15 µl per patient) from LC and HCC was obtained, respectively. After in-gel digestion and O-glycopeptide enrichment, immobilized trypsin (Thermo Scientific, Rockford, IL) was added (enzyme: substrate = 1:5 v:w) and freeze-dried. The two freeze-dried powders were resolubilized with 20% ACN in $H_2^{16}O/H_2^{18}O$ (97%, Cambridge Isotope Laboratories, Andover, MA), respectively, at 37°C for 24 h. The reaction was quenched by adding 1 µl FA, and the supernatant was collected using centrifuge columns (Pierce, Rockford, IL). The ^{16}O- and ^{18}O-fractions were redissolved with 5 µl solvent A, respectively (solvent A: 0.1% FA in water), and combined, isolated by nano-LC, and detected by online electrospray tandem MS. ^{18}O/^{16}O labeled O-glycopeptides were identified using Byonic and quantified using pQuant. The ^{18}O/^{16}O glycopeptide ratio is based on a pair of isotope chromatograms with the least interference to exclude the overlap of the ^{16}O and ^{18}O isotopic peaks. pQuant usually chooses the monoisotopic peak from light labeling because it is usually the least interfered peak for light labeling. Then, due to different interference conditions, a different peak from heavy labeling was considered. If the monoisotopic peak of heavy labeling was chosen, the algorithm in our previous study (Zhang et al., 2019) was used. The denominator of the formula, such as

$$I_2 + I_4 - \left(\frac{M_2}{M_0}\right)I_2 - \left[\left(\frac{M_2}{M_0}\right) + \left(\frac{M_4}{M_0}\right) - \left(\frac{M_2}{M_0}\right)^2\right]I_0,$$ subtracts the

light labeling effect from heavy labeling. If the other peak from heavy labeling was chosen (not the monoisotopic or the first peak), all the isotopic chromatograms were normalized using pQuant. The normalized chromatograms of the light peptide are $L_1 = p_1^l \ (\Sigma t_i^l/t_1^l)$, $L_2 = p_2^l \ (\Sigma t_i^l/t_2^l)$, and $L_N = pl_N \ (\Sigma t_i^l/t_N^l)$. The normalized chromatograms of the heavy peptide are $H_1 = p_1^h \ (\Sigma t_i^h/t_1^h)$, $H_2 = p_2^h \ (\Sigma t_i^h/t_2^h)$, and $H_M = p_M^h(\Sigma t_i^h/t_M^h)$. Once normalized, the monoisotopic peak from the light peptide and the selected peak from the heavy labeling can be used to calculate the H/L ratio of the glycopeptide.

Multiple Reaction Monitoring Validation

Purified Hp was obtained from 100 µl of serum of each sample, and a total of 24 patients were enrolled (12 HCC patients and 12 LC patients). After digestion overnight with trypsin, the samples were diluted in 2% ACN and 0.1% FA and ionized using an Ultimate 3,000 HPLC (150 × 2.1 mm, 3 µm, 100 Å column, Thermo, Framingham, United States) coupled with a 6,500 QTRAP MS (ABSciex, Waltham, United States). Peptide
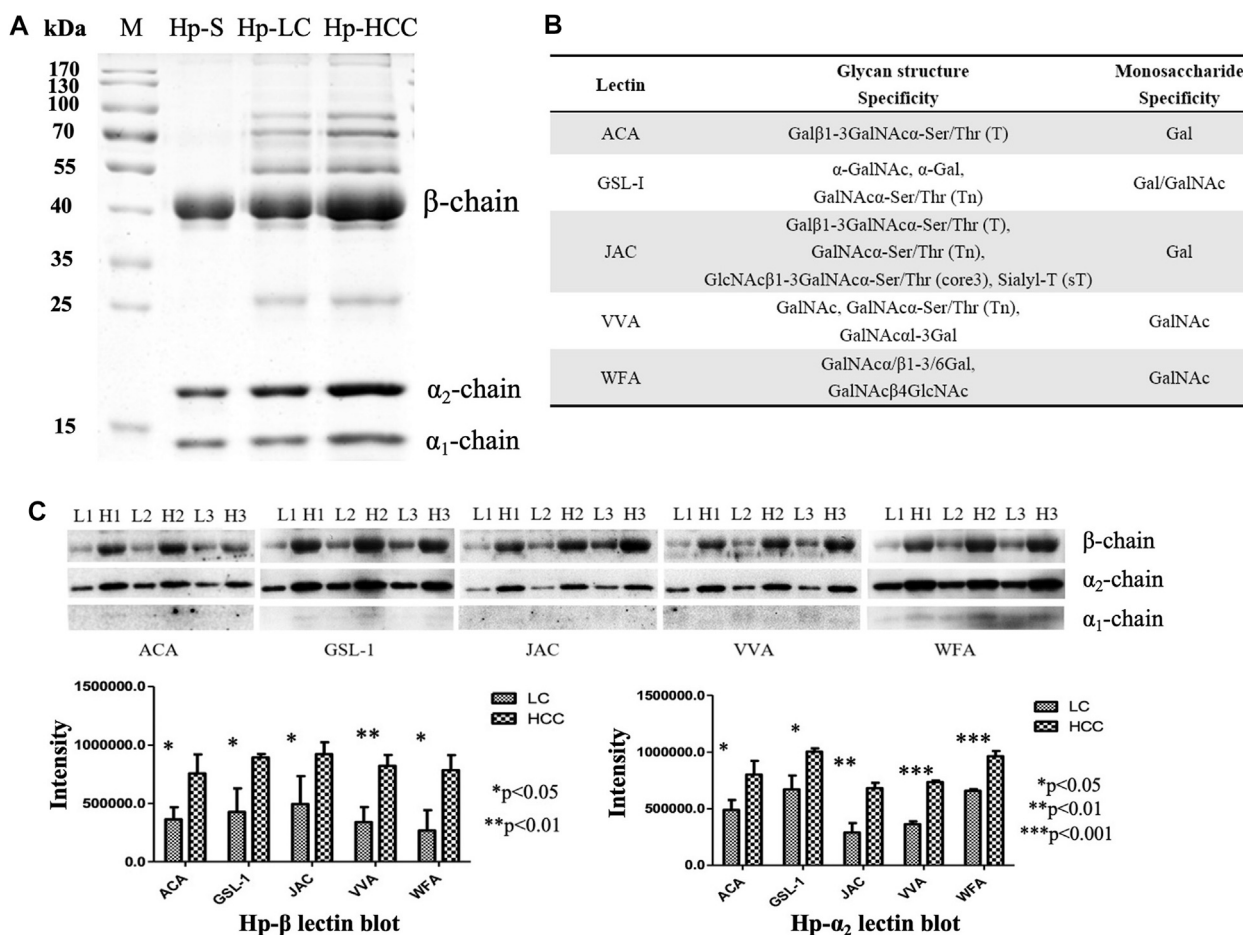
FIGURE 1 | Purification of Hp and its O-glycosylation status by lectin blot analysis. **(A)** Purified Hp from the sera of LC and HCC patients. M, marker; Hp-S, Hp standard; Hp-LC, purification of Hp from LC patients; Hp-HCC, purification of Hp from HCC patients. **(B)** Carbohydrate-binding specificities of five lectins (ACA, GSL-1, JAC, VVA, and WFA). **(C)** Lectin blot was applied to reveal the O-glycan levels of Hp. Hp from an equal volume of LC and HCC sera was used, and three biological repeats were performed. *p*-value of less than 0.05 showing statistical significance using nonparametric Mann–Whitney *U* tests. L, LC; H, HCC.

separation of the individual samples was achieved using the following 30-min gradient for flow rates of 200 μL/min, with solvent A (0.1% FA in water) and solvent B (0.1% FA in ACN): 5% B for 1 min, 5%–50% B for 13 min, 80% B for 0.5 min, and 50% B for 7.5 min. The optimum transitions and one unique peptide of Hp were selected for each glycopeptide/peptide for MRM monitoring. Analyzer parameters were optimized for each peptide/transition pair to ensure maximum selectivity. Both Q1 and Q3 resolutions were chosen as "Unit" (±0.7 Da). The acquired MRM wiff files were analyzed using Skyline software. The precursor ion of the O-glycopeptide was chosen, and the product ions were used to determine the peak area for the O-glycopeptide.

Statistical Analysis

All data and graphs were generated using GraphPad Prism 7.0 software (GraphPad Software, Inc.). Statistical comparisons were carried out using the *t*-test, and *p* values < 0.05 presented as statistically significant. Lectin blot and MRM data were evaluated using nonparametric Mann–Whitney *U* tests.

Data Availability Statement

The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral. proteomexchange.org) *via* the iProX partner repository (Ma et al., 2019) with the dataset identifier PXD023447.

RESULTS AND DISCUSSION

O-Glycosylation Status of Haptoglobin by Lectin Blot Analysis

Hp is a glycoprotein produced in the liver that is secreted into the blood. Lectins are defined as carbohydrate-binding proteins and have biotechnological implications. Lectin-based methods were developed for the detection of fucosylated Hp (Miyoshi and

FIGURE 2 | Workflow for label-free and labeling quantification of O-glycopeptides on Hp.
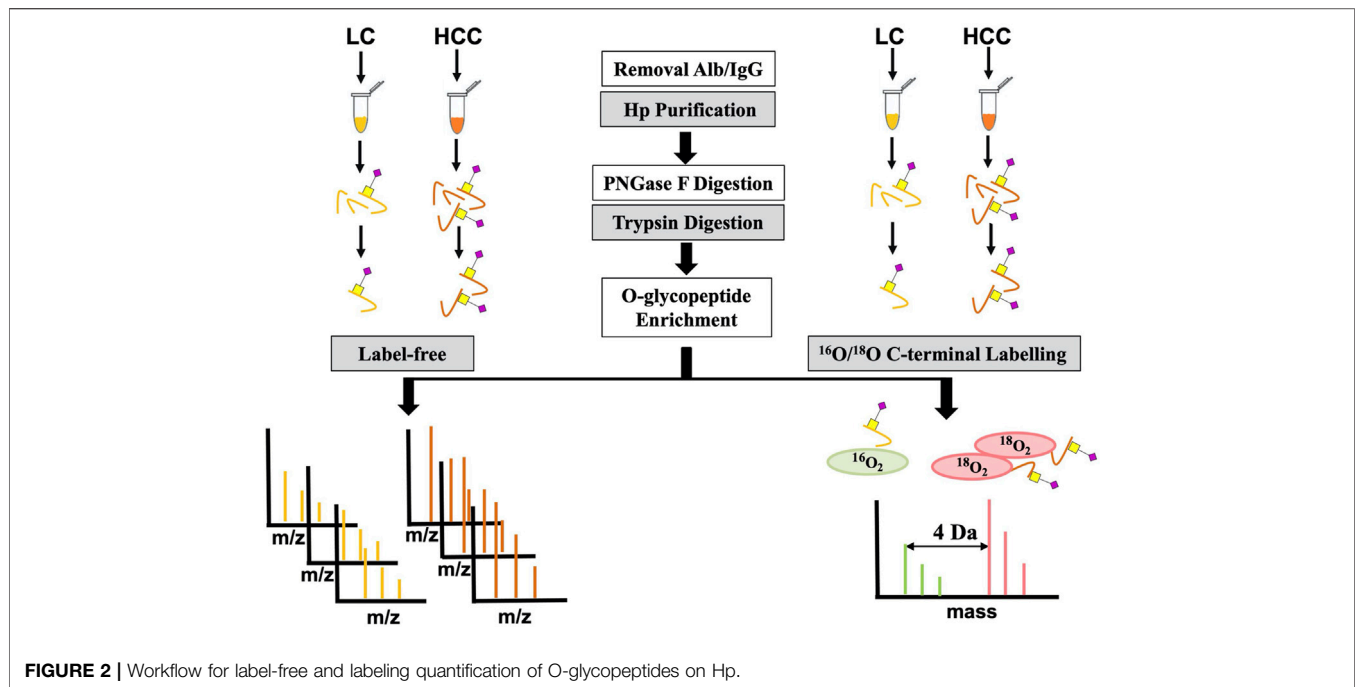
Kamada. 2016; Shang et al., 2017), and they may serve as a cancer biomarker for clinical application. Fucosylation is one of the most important types of glycosylation in cancer and inflammation, for example, the core-fucosylated T-cell receptor was necessary for T-cell signaling and production of inflammatory cytokines (Fujii et al., 2016). However, limited studies on O-glycans of Hp have been reported. O-glycosylation of Hp was found in pigs with PCV2-SD infection, and the presence of mono- and disialyl core type 1 O-glycans of Hp was found in prostate cancer (Fujimura et al., 2008; Marco-Ramell et al., 2014). Here, lectin blotting was used to determine the O-glycosylation of Hp in liver diseases.

First, Hp was purified from an equal volume of LC and HCC sera, respectively (**Figure 1A**), and the gel bands were cut, digested, and determined by LC-MS/MS. The corresponding MS results confirmed that these proteins in the gels were Hp (**Supplementary Figure S1**). Then, five lectins including ACA, GSL-1, JAC, VVA, and WFA were chosen to reveal the potential O-glycosylation status of Hp in LC and HCC. The monosaccharide-binding specificities of these lectins are presented in **Figure 1B**. Lectin blot analysis showed that the β- and α2-chains of Hp in HCC had a higher binding ability with regard to these lectins than those in cirrhosis (**Figure 1C** and **Supplementary Material**). Both predicted and experimentally determined O-glycosylation sites were nearly in accordance with those already known O-glycosylation sites, such as fibrinogen, α-2-HS-glycoprotein, and so on (Hoffmann et al., 2016). In this study, according to the Hp sequence, O-glycosylation at four Thr residues (Thr67, 126, 317, and 323) and one Ser residue (Ser316) has been predicted using the NetOGlyc 4.0 server (Steentoft et al., 2013). The results indicated the possible presence of O-glycans on Hp, and it also displayed differences in liver diseases.

Intact O-Glycopeptides of Haptoglobin in Liver Cirrhosis and Hepatocellular Carcinoma

To further confirm its O-glycosylation in HCC, both the label-free and labeling methods based on MS were applied (**Figure 2**). A total of seven HCC and seven LC serum patients were first used for label-free quantification. For this method, serum Hp was purified from an equal volume of patients with LC and HCC. Then, purified Hp was treated with PNGase F and trypsin. After glycopeptide enrichment, they were subjected to LC-MS/MS analysis, respectively. All spectra raw files were automatically identified using Byonic.

As shown in **Figure 3A**, 26 intact O-glycopeptides on four O-glycosylation sites (Thr126, 317, 323, and Ser316) were identified on the Hp protein and corresponded to 18 types of glycan compositions. Most of them were elevated in HCC as compared to LC. As shown in **Supplementary Table S1**, 57.69% of the changed glycopeptides were observed in seven patients and 96.15% in at least four patients. In addition, the majority of the glycoforms were located on Ser316 and Thr317, while one glycoform was located on Thr126 and Thr323. Moreover, one glycoform (H1N4) on Thr317 could be detected in HCC; however, this was absent in cirrhosis. Among these O-glycopeptides on Hp, the intensity of HYEGS[316]TVPEK (H1N1S1) was the highest, and it was significantly increased in HCC patients (**Figure 3B**, $p < 0.05$). The glycan composition of this O-glycopeptide was core 1 type O-glycans with one NeuAc residue.

Aberrant fucosylation and sialylation have been reported in various cancers (Verhelst et al., 2020). As one of the acute-phase response proteins, Hp contains four N-glycosylation sites (Asn184, 207, 211, and 241), and its fucosylated N-glycans were
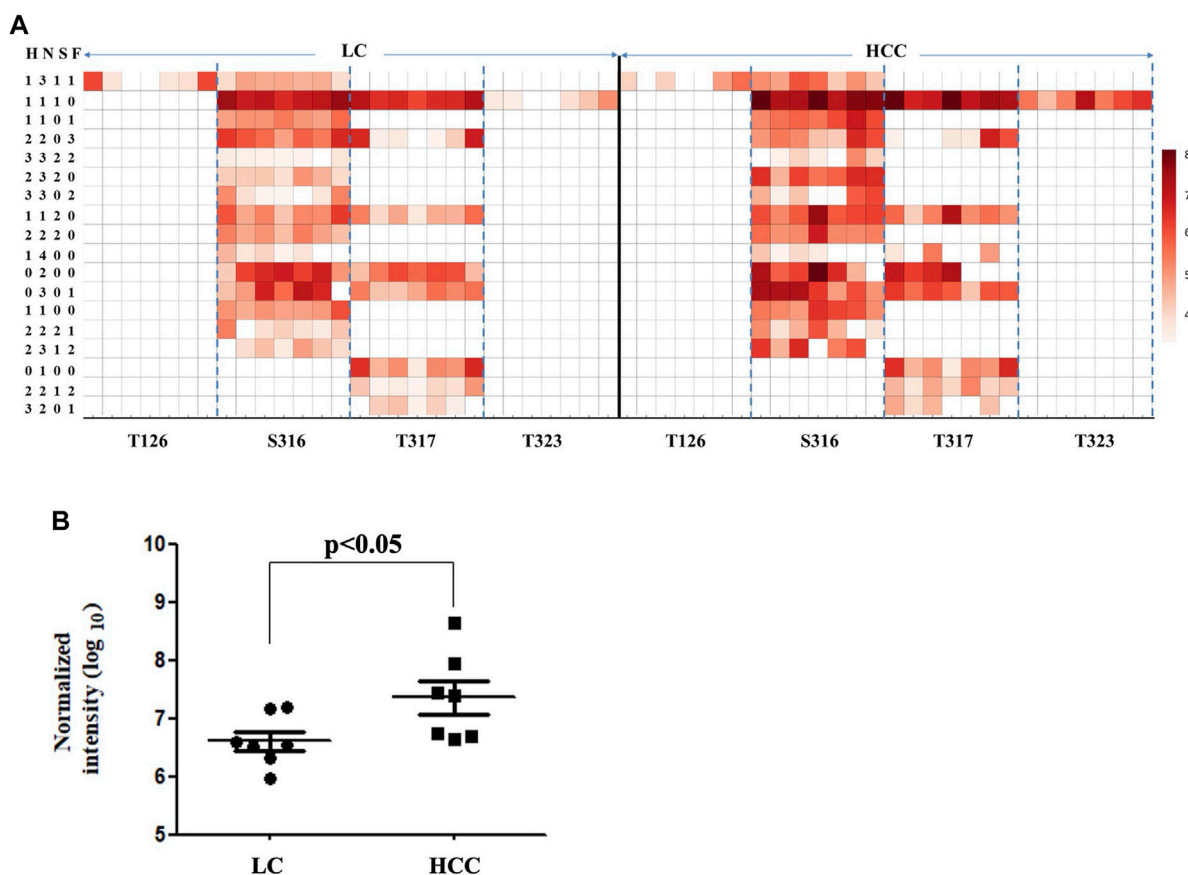
FIGURE 3 | Abundance of Hp O-glycopeptides in LC and HCC. **(A)** Intensity distributions of Hp O-glycopeptides in seven HCC and seven LC patients. The *X*-axis represents identified O-glycosylation sites of Hp. The *Y*-axis represents the intensity (Log10 conversion) of different glycoforms in Hp. **(B)** Label-free quantification of HYEGS316TVPEK (H1N1S1) of Hp and the difference was statistically significant.

reported in multiple cancer types (Takahashi et al., 2016). In this study, the sialylation alteration of O-glycans on Hp provided clues to the comprehensive understanding of its glycosylation in cancers.

Intact O-Glycopeptide HYEGS[316]TVPEK (H1N1S1) on Haptoglobin

We focused on HYEGS[316]TVPEK (H1N1S1), which had the highest intensity and was significantly increased in HCC patients. The labeling method was applied to confirm this change. For the labeling method, O-glycopeptides from LC were labeled by $H_2^{16}O$ and those from HCC were labeled by $H_2^{18}O$. There were 4-Da mass shifts between ^{16}O- and ^{18}O-labeled samples. Labeled samples were pooled and detected by LC-MS/MS. The reproducibility and quantitative accuracy of this labeling strategy were first evaluated using standard Hp. A series of theoretical ratios of standard (1:2, 1:1, 2:1, 5:1, and 10:1) was applied. The O-glycopeptide HYEGS316TVPEK (H1N1S1) was identified, and experimental ratios of this O-glycopeptide showed good linearity with a correlation coefficient (R^2) approximate of 0.99 (**Figure 4A**). For LC and HCC samples, this O-glycopeptide of Hp was elevated significantly in HCC patients (**Figure 4B**). Based on three

biological replicates, the average fold change (HCC/LC) of this O-glycopeptide was about 6.2 (**Figure 4C**). **Figure 4D** shows the representative MS^2 spectrum of this O-glycopeptide, annotated using the gLabel software tool.

Intact glycopeptide analysis including glycosylation sites and site-specific glycans is crucial for the understanding of glycosylation (Medzihradszky et al., 2015). The degree of O-glycan sialylation was observed to be associated with the pathogenesis of many diseases (Campbell et al., 2001). However, the minor level of O-glycosylation posed a challenge for further analyses of it (Yang et al., 2018; Zhang et al., 2018). Using label-free and labeling quantification, elevated HYEGS[316]TVPEK (H1N1S1) of Hp in HCC was found in this study.

Multiple Reaction Monitoring Analyses of HYEGS[316]TVEPK (H1N1S1)

To further confirm elevated HYEGS[316]TVEPK (H1N1S1) of Hp in HCC, MRM analysis was also applied (Sanda et al., 2013). The target O-glycopeptide HYEGS[316]TVPEK (H1N1S1) and the unique peptide of Hp were chosen for MRM transitions (**Figure 5A**). In this study, peak areas of the O-glycopeptide were employed for absolute quantification, and a unique peptide was used to confirm the

FIGURE 4 | $^{16}O/^{18}O$ C-terminal labeling for quantification of Hp O-glycopeptides. **(A)** Series of theoretical ratios of Hp O-glycopeptide standard (1:2, 1:1, 2:1, 5:1, and 10:1) was applied, and experimental ratios showed good linearity with a correlation coefficient (R^2) approximate of 0.99. **(B)** Representative MS^1 spectrum of HYEGS^{316}TVPEKK (H1N1S1). **(C)** Three biological repeats were performed, and the average fold change (HCC/LC) of this O-glycopeptide was about 6.2. **(D)** Representative MS^2 spectrum was shown using the gLabel software tool. The O-glycosylation site is "S." The glycan symbols are as follows: purple diamond for sialic acid (S), yellow circle for hexose (H), and yellow square for N-acetylhexosamine (N). The glycan composition and peptide sequence are in the upper box of each spectrum. The peak annotations are displayed in the middle box containing fragment ions of the glycan part or diagnostic glycan ions (green, blue, and purple), y-ions from glycan fragmentation (red), and the b/y ions from peptide backbone fragmentation (yellow/cyan). Mass deviations in the annotated peaks are displayed in the box below.

purification of Hp. **Figure 5B** shows that this O-glycopeptide was also significantly elevated in HCC patients as compared to LC ($p < 0.05$). Thus, MRM analyses validated this intact O-glycopeptide on Hp which exhibited differences in HCC. We also used the unique peptide of Hp (VGYVSGWGR) as an internal standard, and the O-glycopeptide abundance was divided by the unique peptide abundance to separate out the contribution of protein concentration. The result showed that the increased HYEGS^{316}TVEPK (H1N1S1) was caused by elevated protein expression in HCC (**Figure 5C**).

Mass spectrometry–based strategies for glycopeptide quantification include isotopic labeling and the label-free method (Delafield and Li, 2021). For the labeling strategy, the quantitative result can be obtained simultaneously by comparing the abundance of the isotopologues. The label-free strategy requires the stability of the production, and it has benefited from the implementation of MRM. For MRM analysis, effective ionization of glycopeptides and

the reproducible fragments are important. In our study, MS^1 level–based label-free quantification was performed first to screen O-glycopeptides. The $^{16}O/^{18}O$-labeled method and MS^2 level–based MRM were used to quantify the significantly changed O-glycopeptide and to accurately confirm this alteration. All the above results indicated the existence of sialylated O-glycans on Hp. Compared to the LC patients, significantly elevated HYEGS^{316}TVPEK (H1N1S1) of Hp was identified in HCC. More studies are still needed to increase the sensitivity for O-glycosylation analyses and uncover its biological function in liver diseases.

CONCLUSION

In the present study, we demonstrated that the intact O-glycopeptide HYEGS^{316}TVPEK (H1N1S1) increased significantly in HCC patients. Changes in protein glycosylation
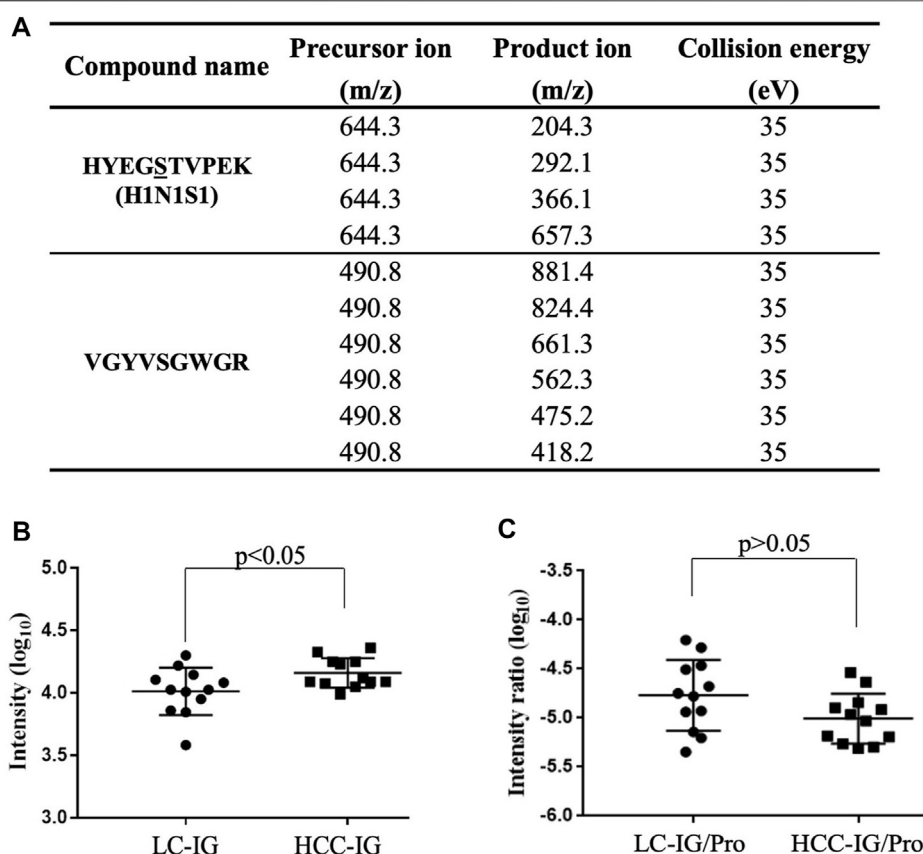
FIGURE 5 | MRM for validation. **(A)** MRM transitions for monitoring the O-glycopeptide and the unique peptide of Hp. **(B)** HYEGS[316]TVPEK (H1N1S1) of Hp from 12 LC patients and 12 HCC patients was detected, and the difference was statistically significant. **(C)** O-glycopeptide abundance was divided by the unique peptide abundance to separate out the contribution of protein concentration. The result showed that the increased HYEGS[316]TVEPK (H1N1S1) was caused by elevated protein expression in HCC. LC-IG, the intact O-glycopeptide in LC; HCC-IG, the intact O-glycopeptide in HCC; LC-IG/Pro, the intact O-glycopeptide abundance divided by the protein abundance in LC; HCC-IG/Pro, the intact O-glycopeptide abundance divided by the protein abundance in HCC.

might be used as diagnostic and prognostic markers, as well as targets of therapy for cancer. The aberrant O-glycopeptide on Hp was explored in connection with liver disease, and this study provided clues to a comprehensive understanding of its glycosylation in cancers.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: the ProteomeXchange Consortium (http://proteomecentral. proteomexchange.org) *via* the iProX partner repository with the dataset identifier PXD023447.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institution Ethics Committee of the Cancer

Hospital of Guangxi Medical University. The patients/ participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HS, SZ, and LZ designed the research study; HS, YC, YG, LL, ZC, GY, and CL performed the experiments; HS, LZ, FC, and CL analyzed the data; and HS, SZ, and CL wrote the manuscript.

FUNDING

Science Fund Project of Guangxi Province of China (2018GXNSFAA281053), the Self-Raised Scientific Research Fund of the Ministry of Health of Guangxi Province (Z2015587), and the National Natural Science Foundation of China (31700727).

REFERENCES

Campbell, B. J., Yu, L. G., and Rhodes, J. M. (2001). Altered Glycosylation in Inflammatory Bowel Disease: A Possible Role in Cancer Development. *Glycoconj J* 18 (11-12), 851–858. doi:10.1023/a:1022240107040

Delafield, D. G., and Li, L. (2021). Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. *Mol Cell Proteomics* 20, 100054. doi:10.1074/mcp.R120.002095

Fujii, H., Shinzaki, S., Iijima, H., Wakamatsu, K., Iwamoto, C., Sobajima, T., et al. (2016). Core Fucosylation on T Cells, Required for Activation of T-Cell Receptor Signaling and Induction of Colitis in Mice, Is Increased in Patients with Inflammatory Bowel Disease. *Gastroenterology* 150, 1620–1632. doi:10.1053/j.gastro.2016.03.002

Fujimura, T., Shinohara, Y., Tissot, B., Pang, P. C., Kurogochi, M., Saito, S., et al. (2008). Glycosylation Status of Haptoglobin in Sera of Patients with Prostate Cancer vs. Benign Prostate Disease or normal Subjects. *Int. J. Cancer* 122, 39–49. doi:10.1002/ijc.22958

Hoffmann, M., Marx, K., Reichl, U., Wuhrer, M., and Rapp, E. (2016). Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol. Cel Proteomics* 15, 624–641. doi:10.1074/mcp.M115.053546

Liu, C., Song, C. Q., Yuan, Z. F., Fu, Y., Chi, H., Wang, L. H., et al. (2014). pQuant Improves Quantitation by Keeping Out Interfering Signals and Evaluating the Accuracy of Calculated Ratios. *Anal. Chem.* 86, 5286–5294. doi:10.1021/ac404246w

Liu, M. Q., Zeng, W. F., Fang, P., Cao, W. Q., Liu, C., Yan, G. Q., et al. (2017). pGlyco 2.0 Enables Precision N-Glycoproteomics with Comprehensive Quality Control and One-step Mass Spectrometry for Intact Glycopeptide Identification. *Nat. Commun.* 8, 438. doi:10.1038/s41467-017-00535-2

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., et al. (2019). iProX: an Integrated Proteome Resource. *Nucleic Acids Res.* 47, D1211–D1217. doi:10.1093/nar/gky869

Marco-Ramell, A., Miller, I., Nobauer, K., Moginger, U., Segales, J., Razzazi-Fazeli, E., et al. (2014). Proteomics on Porcine Haptoglobin and IgG/IgA Show Protein Species Distribution and Glycosylation Pattern to Remain Similar in PCV2-SD Infection. *J. Proteomics* 101, 205–216. doi:10.1016/j.jprot.2014.02.018

Medzihradszky, K. F., Kaasik, K., and Chalkley, R. J. (2015). Tissue-Specific Glycosylation at the Glycopeptide Level. *Mol Cell Proteomics* 14, 2103–2110. doi:10.1074/mcp.M115.050393

Miyoshi, E., and Kamada, Y. (2016). Application of Glycoscience to the Early Detection of Pancreatic Cancer. *Cancer Sci.* 107, 1357–1362. doi:10.1111/cas.13011

Okuyama, N., Ide, Y., Nakano, M., Nakagawa, T., Yamanaka, K., Moriwaki, K., et al. (2006). Fucosylated Haptoglobin Is a Novel Marker for Pancreatic Cancer: a Detailed Analysis of the Oligosaccharide Structure and a Possible Mechanism for Fucosylation. *Int. J. Cancer* 118, 2803–2808. doi:10.1002/ijc.21728

Sanda, M., Pompach, P., Brnakova, Z., Wu, J., Makambi, K., and Goldman, R. (2013). Quantitative Liquid Chromatography-Mass Spectrometry-Multiple Reaction Monitoring (LC-MS-MRM) Analysis of Site-specific Glycoforms of Haptoglobin in Liver Disease. *Mol Cell Proteomics* 12, 1294–1305. doi:10.1074/mcp.M112.023325

Shang, S., Li, W., Qin, X., Zhang, S., and Liu, Y. (2017). Aided Diagnosis of Hepatocellular Carcinoma Using Serum Fucosylated Haptoglobin Ratios. *J. Cancer* 8, 887–893. doi:10.7150/jca.17747

Shu, H., Zhang, S., Kang, X., Li, S., Qin, X., and Sun, C. (2011). Protein Expression and Fucosylated Glycans of the Serum Haptoglobin-{beta} Subunit in Hepatitis B Virus-Based Liver Diseases. *Acta Biochim. Biophys. Sinica* 43, 528–534. doi:10.1093/abbs/gmr038

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., et al. (2013). Precision Mapping of the Human O-GalNAc

Glycoproteome through SimpleCell Technology. *EMBO J.* 32, 1478–1488. doi:10.1038/emboj.2013.79

Takahashi, S., Sugiyama, T., Shimomura, M., Kamada, Y., Fujita, K., Nonomura, N., et al. (2016). Site-specific and Linkage Analyses of Fucosylated N-Glycans on Haptoglobin in Sera of Patients with Various Types of Cancer: Possible Implication for the Differential Diagnosis of Cancer. *Glycoconj J.* 33, 471–482. doi:10.1007/s10719-016-9653-7

Verhelst, X., Dias, A. M., Colombel, J. F., Vermeire, S., Van Vlierberghe, H., Callewaert, N., et al. (2020). Protein Glycosylation as a Diagnostic and Prognostic Marker of Chronic Inflammatory Gastrointestinal and Liver Diseases. *Gastroenterology* 158, 95–110. doi:10.1053/j.gastro.2019.08.060

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. (2019). A Global View of Hepatocellular Carcinoma: Trends, Risk, Prevention and Management. *Nat. Rev. Gastroenterol. Hepatol.* 16, 589–604. doi:10.1038/s41575-019-0186-y

Yang, W., Ao, M., Hu, Y., Li, Q. K., and Zhang, H. (2018). Mapping the O-Glycoproteome Using Site-specific Extraction of O-Linked Glycopeptides (EXoO). *Mol. Syst. Biol.* 14, e8486. doi:10.15252/msb.20188486

Zhang, S., Cao, X., Liu, C., Li, W., Zeng, W., Li, B., et al. (2019). N-glycopeptide Signatures of IgA in Serum from Patients with Hepatitis B Virus-Related Liver Diseases. *Mol Cell Proteomics* 18, 2262–2272. doi:10.1074/mcp.RA119.001722

Zhang, S., Jiang, K., Sun, C., Lu, H., and Liu, Y. (2013). Quantitative Analysis of Site-specific N-Glycans on Sera Haptoglobin Beta Chain in Liver Diseases. *Acta Biochim. Biophys. Sin (Shanghai)* 45, 1021–1029. doi:10.1093/abbs/gmt110

Zhang, S., Liu, X., Kang, X., Sun, C., Lu, H., Yang, P., et al. (2012). iTRAQ Plus ^{18}O: a New Technique for Target Glycoprotein Analysis. *Talanta* 91, 122–127. doi:10.1016/j.talanta.2012.01.033

Zhang, S., Shang, S., Li, W., Qin, X., and Liu, Y. (2016). Insights on N-Glycosylation of Human Haptoglobin and its Association with Cancers. *Glycobiology* 26, 684–692. doi:10.1093/glycob/cww016

Zhang, S., Shu, H., Luo, K., Kang, X., Zhang, Y., Lu, H., et al. (2011). N-linked Glycan Changes of Serum Haptoglobin Beta Chain in Liver Disease Patients. *Mol. Biosyst.* 7, 1621–1628. doi:10.1039/c1mb05020f

Zhang, Y., Xie, X., Zhao, X., Tian, F., Lv, J., Ying, W., et al. (2018). Systems Analysis of Singly and Multiply O-Glycosylated Peptides in the Human Serum Glycoproteome via EThcD and HCD Mass Spectrometry. *J. Proteomics* 170, 14–27. doi:10.1016/j.jprot.2017.09.014

Zhu, J., Chen, Z., Zhang, J., An, M., Wu, J., Yu, Q., et al. (2019b). Differential Quantitative Determination of Site-specific Intact N-Glycopeptides in Serum Haptoglobin between Hepatocellular Carcinoma and Cirrhosis Using LC-EThcD-MS/MS. *J. Proteome Res.* 18, 359–371. doi:10.1021/acs.jproteome.8b00654

Zhu, J., Warner, E., Parikh, N. D., and Lubman, D. M. (2019a). Glycoproteomic Markers of Hepatocellular Carcinoma-Mass Spectrometry Based Approaches. *Mass. Spectrom. Rev.* 38, 265–290. doi:10.1002/mas.21583

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.705341/full#supplementary-material

Recent Advances in Mass Spectrometry-Based Glycomic and Glycoproteomic Studies of Pancreatic Diseases

Dylan Nicholas Tabang[1], Megan Ford[2] and Lingjun Li[1,3]*

[1]Department of Chemistry, University of Wisconsin-Madison, Madison, WI, United States, [2]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, United States, [3]School of Pharmacy, University of Wisconsin-Madison, Madison, WI, United States

Modification of proteins by glycans plays a crucial role in mediating biological functions in both healthy and diseased states. Mass spectrometry (MS) has emerged as the most powerful tool for glycomic and glycoproteomic analyses advancing knowledge of many diseases. Such diseases include those of the pancreas which affect millions of people each year. In this review, recent advances in pancreatic disease research facilitated by MS-based glycomic and glycoproteomic studies will be examined with a focus on diabetes and pancreatic cancer. The last decade, and especially the last five years, has witnessed developments in both discovering new glycan or glycoprotein biomarkers and analyzing the links between glycans and disease pathology through MS-based studies. The strength of MS lies in the specificity and sensitivity of liquid chromatography-electrospray ionization MS for measuring a wide range of biomolecules from limited sample amounts from many sample types, greatly enhancing and accelerating the biomarker discovery process. Furthermore, imaging MS of glycans enabled by matrix-assisted laser desorption/ionization has proven useful in complementing histology and immunohistochemistry to monitor pancreatic disease progression. Advances in biological understanding and analytical techniques, as well as challenges and future directions for the field, will be discussed.

Keywords: mass spectrometry, glycosylation, glycation, pancreatic cancer, pancreatitis, diabetes, glycoproteomics, glycomics

---

Abbreviations: PTM, post-translational modification; AGE, advanced glycation end product; CML, carboxymethyllysine; MS, mass spectrometry; MALDI, matrix-assisted laser desorption/ionization; ESI, electrospray ionization; LC, liquid chroma-tography; T1D/T2D, type 1/type 2 diabetes; PDAC, pancreatic ductal adenocarcinoma; IGP, intact glycopeptide; GlcNAc, N-acetylglucosamine; SNA, *S. nigra* lectin; MAL II, *M. amurensis* lectin II; AGP, alpha-1-acid glycoprotein; DN, diabetic nephropathy; TMA, tumor microarray; GAG, glycosaminoglycan; ECM, extracellular matrix; HSA, human serum albumin; IgG, immunoglobulin G; TOF, time-of-flight; CA 19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; FT-ICR, Fourier transform-ion cyclotron resonance.

INTRODUCTION

Glycosylation and other post-translational modifications (PTMs) have been frequently studied in the context of diseases (Chen Z. et al., 2018; Shi et al., 2019; Zhang H. et al., 2020; Chen et al., 2021). No other PTM carries as much possible structural heterogeneity as glycosylation, the enzymatic addition of glycan moieties to a protein backbone. Glycan moieties on proteins have been implicated in mediating cell signaling processes, preventing protein degradation, and regulating substrate binding. In many diseases, aberrations in normally functioning pathways lead to changes in the glycome. Thus, glycans have also become useful biomarkers for disease diagnosis or monitoring responses to disease treatment.

Much work has been done underscoring the importance of glycans in these pancreatic diseases. It is important to note the distinction between two different processes leading to addition of glycans to proteins: glycation and glycosylation.

Glycation is a non-enzymatic addition of a carbohydrate monomer, most often glucose, to a protein backbone. It is a non-enzymatic process that is concentration dependent. This process leads to protein structural heterogeneity through a series of reactions that transform the carbohydrate monomer into advanced glycation end products (AGEs), including carboxymethyllysine (CML). Glucose is the monosaccharide most often added *via* glycation. Glycation often modifies the amino groups of Lys and Arg residues and the protein N-terminus (Brownlee, 1995).

Glycosylation, on the other hand, is the enzymatic addition of a carbohydrate moiety, often with a core glycan structural template, and is not necessarily concentration dependent. These two processes are linked through the hexosamine biosynthetic pathway, in which glucose is transformed into uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), the first monomer involved in glycosylation. The most common forms of glycosylation are N-linked (Asn) and O-linked (Ser and Thr) (Akella et al., 2019). Protein structures, and consequently their functions, are affected by glycation, glycosylation, and other PTMs.

Mass spectrometry (MS) has emerged as the best analytical tool for glycan structural analysis and quantification, with the possibility of glycan isomer-specific resolution. The two most common forms of MS ionization are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), which is usually interfaced with liquid chromatography (LC). In MALDI, laser irradiation of matrix molecules co-crystallized with sample is used to generate analyte ions. In ESI, samples are subjected to high voltage, nebulized into a mist, and are desolvated into analyte ions. Both methods produce gas-phase ions with minimal fragmentation of the analyte.

Advances in sample preparation and enrichment (Riley et al., 2020), mass spectrometer instrumentation (Thomas and Scott, 2020), and data analysis (Abrahams et al., 2020; Delafield and Li, 2021) have made glycan, glycopeptide, and glycoprotein-level analyses more facile and accessible. These advances, discussed in more detail in these other recent reviews, have provided new insights into how glycans are altered in and drive human diseases.

Diseases of the pancreas affect millions of people annually. The pancreas performs both exocrine and endocrine functions by secreting digestive enzymes and hormones. Diabetes is a disease state characterized by abnormal blood glucose concentration involving dysfunction of the pancreatic hormone insulin, produced by beta cells in the islets of Langerhans. The two main mechanisms leading to irregular blood glucose are beta cell destruction, where the pancreas produces little to no insulin (type 1, T1D), and beta cell dysfunction, where the pancreas does not produce enough insulin or responds improperly to insulin (type 2, T2D). According to the Centers for Disease Control and Prevention's National Diabetes Statistics Report, from 2013 to 2016 in the United States alone, diabetes was diagnosed in over 34 million people. This corresponds to approximately 10% of the national population (Prevention, 2020).

When pancreatic enzymes begin digesting the organ itself instead of their intended substrates in the gastrointestinal tract, inflammation results. This is known as pancreatitis, which may be acute or chronic depending on the underlying cause of inflammation. The pancreas may also be affected by cancer. Pancreatic ductal adenocarcinoma (PDAC), where tumors begin to grow in the pancreatic ducts, is the most common form of pancreatic cancer. Furthermore, various pancreatic lesions and cysts may progress into malignancy, so research efforts have also been directed into studying possible precursors to pancreatic cancer (Paziewska et al., 2018; Pan et al., 2020). Pancreatitis and pancreatic cancer are often studied together in the same experimental workflows due to their similarity in symptoms such as abdominal pain, weight loss, and loss of appetite.

This review will discuss both glycation and glycosylation in the context of MS-derived advances in glycomic and glycoproteomic research of pancreatic diseases. Overall trends and analytical considerations will first be discussed, followed by a more focused discussion on individual studies on new insights into pancreatic diseases facilitated by MS analyses. To narrow the scope of this review, studies from the previous five years (2016–2021) will be surveyed. Summaries of the 38 reviewed research articles are in the supplementary material, **Supplementary Tables S1, S2**.

TRENDS IN THE REVIEWED STUDIES

Workflow and Analyte Considerations

The main steps of a glycomic or glycoproteomic MS experiment are shown graphically in **Figure 1**. Analytes of interest must first be extracted from biological samples, which can range from cells and tissues to biofluids like urine and the components of blood. Early studies, before developments in MS enabled large-scale analyses of complex mixtures, often purified proteins through incubation with an antibody or with lectin arrays, which bind specific glycan motifs (Higai et al., 2003; Patwa et al., 2006; Zhao et al., 2006; Nakano et al., 2008). Enrichment of glycosylated peptides is often needed because non-modified peptides are the most abundant species in complex mixtures and ionize more efficiently than more hydrophilic glycopeptides (Cech and Enke, 2000; Pouria et al., 2004). This complicates certain MS analyses operating with data dependent acquisition, which in one form

FIGURE 1 | A representative workflow for mass spectrometry-based glycoproteomic analyses. Purification/enrichment, digestion, and derivatization of analytes may be performed at the glycoprotein, glycopeptide, and glycan levels.
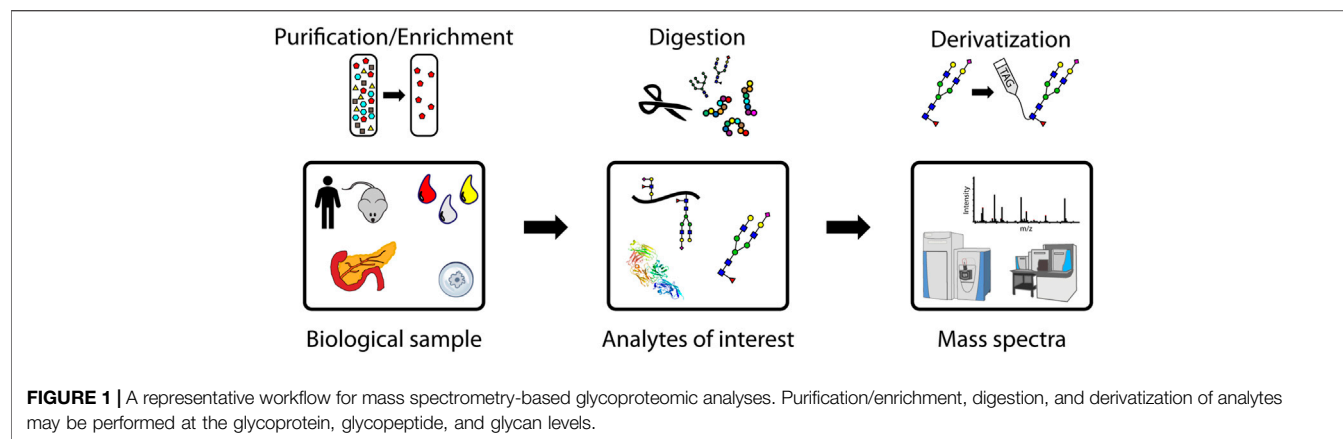
may be used to detect and fragment the most abundant ions. This is also known as a "top N" method based on how many of the most abundant ion m/z values chosen for fragmentation. Targeting specific m/z values for fragmentation can help measure analytes of comparatively lower abundance, including glycopeptides.

Another strategy used more recently relies on deglycosylation of enriched glycopeptides prior to MS analysis (Krishnan et al., 2017; Nigjeh et al., 2017; Kang et al., 2020). This is helpful by decreasing overall analyte complexity and by improving the ionization efficiency of peptides, which are blunted by the increased hydrophilicity contributed by glycan modifications. Samples are usually resuspended in acidic solvents to facilitate positive ion generation. MS-based analyses of glycoproteins are most often performed on the peptides that comprise them to facilitate instrumental detection. Trypsin is the most common enzyme used in what is known as a "bottom-up" approach. Digestion by the enzyme peptide-N-glycosidase F, hereafter referred to as PNGase F, releases of N-glycans from protein backbones. While there is no universal enzyme for the release of O-glycans, which span numerous core structures, there are enzymes available for targeting specific core structures. Alternatively, any O-glycan may be released through the chemical process of β-elimination, though undesirable side reactions may occur (Wilkinson and Saldova, 2020).

The major strength of MS-based analyses is the isolation and fragmentation of analyte peaks to obtain structural information, a process known as tandem MS or MS/MS. Protein, peptide, and glycan fragmentation can be predicted and used to enable database-searching for molecular identification. The efficiency of fragmentation is blunted, however, by both the size and structural properties of analytes. A "top-down" approach analyzing intact glycoproteins retains all PTM structure and localization information, though analyte separation and interpreting fragmentation of intact proteins remain barriers to making top-down analyses more widespread (Chen B. et al., 2018).

Intact glycopeptides (IGPs) are easier to analyze than intact glycoproteins *via* MS as they may exhibit higher ionization efficiencies and generate simpler MS/MS spectra partially due to their smaller sizes compared to their glycoprotein counterparts. Developments in separation techniques and enrichment strategies for glycopeptides have made IGP analysis more routine in recent years (Riley et al., 2020). Information about an intact proteoform's identity or its site-specific modification information could be lost due to enzymatic digestion and precursor proteins must be inferred from identified peptide sequences. This can make quantification of glycosylated proteoforms, or glycoforms, more difficult due to the need to infer precursor protein identities from peptides.

Released glycans, while still vastly heterogenous in their possible structures, are less complex compared to IGPs and glycoproteins. Information on the protein carriers of glycans and residue site localization is completely lost after glycan release. Advantages of releasing glycans include faster analyses and a wider range of instruments capable of performing their analysis. These advantages are consequences of not needing as high mass resolution due to glycan's smaller size and that a mixture of glycans can be analyzed without separation.

Various enzymes have been used to facilitate N-glycan characterization. The most widely used is the previously mentioned PNGase F, which cleaves the bond joining the innermost N-acetylglucosamine residue (GlcNAc) to a protein Asn residue to release the glycan. A related enzyme, PNGase A, can additionally cleave glycans containing α(1,3)-linked core fucosylation, which PNGase F is unable to do. Additionally, the endoglycosidase F3, or simply endo F3, facilitates studies of core fucosylation by hydrolyzing bi-and triantennary glycans, with increased reaction rates for core fucosylated glycans (Lin et al., 2014; Tan et al., 2015).

Both N-linked and O-linked glycosylation have been implicated in pancreatic diseases, and research into enzymes targeting O-glycosylation is a developing area of work. O-GlcNAcylation, the addition of GlcNAc to Ser or Thr, can be used as an indicator of cellular metabolism and has been studied by MS in earlier studies on diabetes (Majumdar et al., 2006; Ma and Hart, 2013; Lim et al., 2014). The considerations needed in studying the small O-GlcNAc modification, including enzymatic analysis methods, have been thoroughly reviewed recently (Maynard and Chalkley, 2021; Xu et al., 2021).

Due to the heterogeneity of O-glycan core structures, there is no universal enzyme for releasing all O-glycans. Chemical means
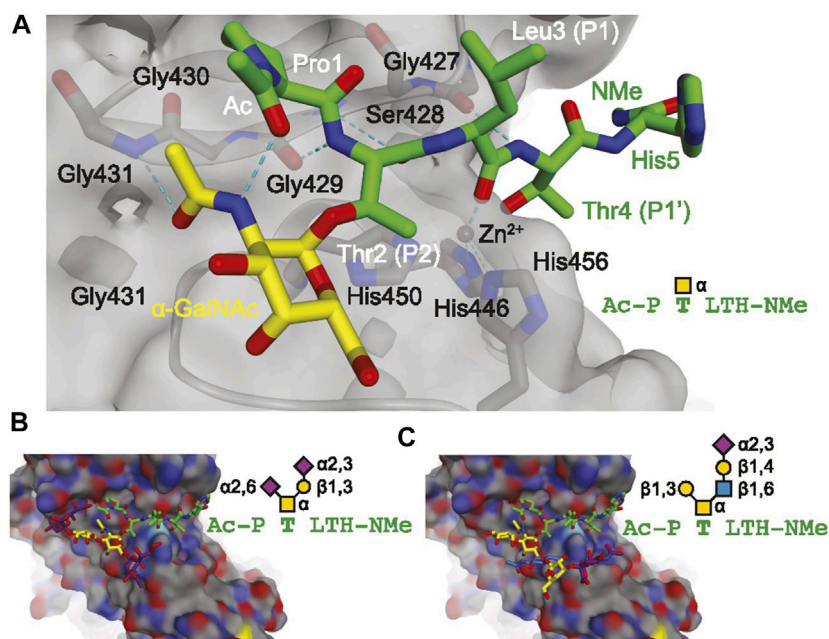
FIGURE 2 | Molecular modeling showing the docking of model glycopeptides (**A**, AcP(α-GalNAc)TLTH-NMe; (**B**), AcP(disialyl core 1)TLTH-NMe; (**C**), Ac-P(sialyl core 2)TLTH-NMe) with the mucin-selective protease StcE from *E. coli*. Reproduced from Malaker, S. A., Pedram, K., Ferracane, M. J., Bensing, B. A., Krishnan, V., Pett, C., et al. (2019). The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. U.S.A.* 116, 7278–7287. doi: 10.1073/pnas.1813020116 under the *PNAS* license to publish.

of release are possible, though they do come with disadvantages, including long reaction times and undesirable "peeling" side reactions resulting from base-catalyzed elimination reactions which can artifactually truncate glycans (Kozak et al., 2012). Endo-α-N-acetylgalactosaminidase, also known as O-glycosidase, is a commercially available enzyme for release of core 1 and core 3 O-glycans, but its use is not nearly as ubiquitous as that of PNGase F. Recent developments in O-glycan targeting enzyme research include OpeRATOR, which cleaves peptides N-terminally to O-glycosylated Ser and Thr (Yang et al., 2018), and secreted protease of C1 esterase inhibitor from *E. coli* (StcE), which releases mucin-type O-glycopeptides (Malaker et al., 2019). Mucins are proteins that are heavily O-glycosylated and are frequent targets for glycobiological analyses. Mucin domain-targeting proteases, including StcE, were recently reviewed (Shon et al., 2021). Molecular modeling of model glycopeptides docking into the active site of StcE can be seen in **Figure 2**. This figure suggests a role for α-acetylgalactosamine selectivity in the conformation and recognition of substrates by the StcE enzyme, which can also accommodate larger glycans (Panels B and C). This enzyme is thus especially helpful for analyzing mucin-type O-glycosylation since these glycans start with α-acetylgalactosamine.

In summary, cleaving of O-glycoproteins into smaller peptides greatly simplifies MS analysis and enables more confident glycan localization.

Without purification or prior protein-level enrichment, complex peptide mixtures are dominated by non-modified peptides. Enrichment is often needed to ensure that glycopeptides are able to be ionized and fragmented by the mass spectrometer. A strategy for enrichment that has seen continued use is using lectins. Lectins, many of which are derived from plants, are proteins that have binding affinities specific to different glycan motifs, such as sialyl or fucosyl moieties. Lectin affinity enrichment can be performed on intact proteins and has been performed in earlier studies to investigate diabetic nephropathy and T2D (Ahn et al., 2010; Nedić et al., 2012). Like antibodies, lectins can be used to pull down molecules of interest from complex mixtures and to stain tissues, adding a level of molecular specificity to optical microscopy (Chugh et al., 2018; Wolters-Eisfeld et al., 2018). In one example investigating sialic acid linkage isomerism in pancreatic cancer serum, two lectins with linkage isomer specificity were used. SNA (*S. nigra* lectin) was used to target terminal α(2,6) sialic acid residues while MAL II (*M. amurensis* lectin II) was used for terminal α(2,3) sialic acid residues (Kontro et al., 2014). Lectin arrays can also be used for high-throughput screening of glycans. They have been used to screen pancreatic cancer serum, tumor tissues, and exosomes (Patwa et al., 2006; Matsuda et al., 2020; Wagatsuma et al., 2020).

Derivatization reactions to enhance detection may be performed prior to MS analysis. One commonly used strategy in released glycan analysis is fluorescent labeling *via* reductive amination. Examples of typical reagents used for this purpose are 2-aminobenzoic acid and 2-aminobenzamide (Sarrats et al., 2010; Tousi et al., 2013). This strategy can add an orthogonal spectroscopic means of quantification besides MS. Other advantages for derivatization include improving ionization

efficiency (Feng et al., 2019a), enabling multiplexed analyses by introduction of heavy isotopes (Bishop et al., 2010; Zhang et al., 2013; Giménez et al., 2015; Feng et al., 2019a; Feng et al., 2019b), and differentiating glycan isomers (Tousi et al., 2013).

The resulting mass spectra can be searched against a database to identify peptide sequences and glycan moieties. Quantification can also be done using label-free approaches, such as using area-under-the-curve, or with reporter ion intensities after isobaric tagging (Shih et al., 2019; Kang et al., 2020). A more detailed discussion of glycan and glycopeptide quantification can be found in Delafield and Li (2021). Statistical analyses can then be done to identify significant changes in analytes among different conditions, including identification of disease biomarkers. For biomarker discovery, these include principal component analysis, which can show how samples cluster, and receiver operating characteristic analysis, which can show the diagnostic performance of different values (Borges et al., 2011).

Sample Type Considerations

Different analytical considerations are needed when working with different sample types. The reviewed studies primarily used samples from human patients, though mouse models for diabetes and pancreatic cancer have also proven useful for research.

Bodily fluids, such as blood and urine, are useful biological samples in that sample collection is relatively non-invasive. Indeed, the long-held gold standard for diabetes diagnosis is the evaluation of glycated hemoglobin (HbA1c) and glucose concentrations in the blood (Lapolla et al., 2000). Whole blood is composed of red and white blood cells, platelets, and plasma. The plasma fraction can be isolated by centrifuging out the solid components of blood without any clotting. To isolate serum, blood is first allowed to clot before separation of the liquid components. The soluble protein components of blood are dominated by several abundant species, including albumin. To analyze proteins of lower abundance, a depletion step is often performed. This may involve separation *via* size exclusion chromatography or use of antibodies to bind and remove the most abundant proteins (Tan et al., 2015). Serum and plasma are great sources of biomarkers. This is due to facile collection of the biological sample and that proteins circulating among different organs are found in these samples. Serum and plasma proteins can thus reflect biological processes happening elsewhere. Samples may also be collected longitudinally with ease to compare patients through time, such as in the monitoring of alpha-1-acid glycoprotein (AGP) at different timepoints detailed in Keser et al. (2021).

Urine was also used in two studies on biomarker discovery (Guo et al., 2015; Belczacka et al., 2019). Diabetic nephropathy (DN) is a complication of diabetes resulting in damaged kidney blood vessels which may lead to chronic kidney disease and kidney failure (Singh et al., 2020a). Since the kidneys are part of the urinary tract, the urinary proteome and glycome may shed light into changes in the kidney during DN. One study investigated the urinary glycoproteome to distinguish different stages of DN (Guo et al., 2015). Biomarkers found in urine could also be helpful for diagnosing pancreatic cancer. The other study

focused on endogenous urine glycopeptides as markers for different cancers, including bladder, prostate, and pancreatic cancer (Belczacka et al., 2019). While urine is plentiful and collection is non-invasive, one disadvantage for this sample type is lower concentration of glycopeptides. Centrifugation is also often needed before sample preparation to pellet insoluble material before extraction.

Tissues of internal organs, such as the pancreas, require invasive surgery for collection. Pancreatic cancer tumors can be banked after surgery and can be used to generate primary cell lines, so studies using human tissues are more common for cancer research than for diabetes. Diabetes usually does not require pancreas surgery. Standard, commercially available cell lines are available for pancreatic cancer tumors and metastases. These cell lines include PANC-1, derived from cells from the head of the pancreas, and CAPAN-1, derived from metastatic pancreatic cancer in the liver (Park et al., 2015). Cell culture provides a steady supply of samples for analyses, though there are several downsides to relying on cells. Cell-based systems do not replicate all possible *in vivo* interactions that may lead to the behavior seen in primary tissues. An example of this is the tumor microenvironment which promotes pancreatic cancer metastasis (Ligorio et al., 2019). To validate results from cultured systems, cell line authentication is also needed to ensure that contamination is not the reason for observed phenomena (Freedman et al., 2015).

Banked primary tissues have also been useful in constructing tumor microarrays (TMAs), which can bring together cores of tissues from patients with various stages of cancer for analysis on the same slide for high-throughput analyses (Pan et al., 2014). Fluids from pancreatic cysts, collected through fine needle aspiration, have also been analyzed. The glycoproteomes of these fluids have been investigated to distinguish benign cysts from malignant cysts which can progress to cancer (Mann et al., 2012; Porterfield et al., 2014).

Mouse models of diabetes and pancreatic cancer provide more flexibility in changing experimental conditions to study these diseases. Commonly used mouse models of diabetes include the streptozotocin-induced T1D mouse and the db/db obese and type 2 diabetic mouse (Sookwong et al., 2011; Liljedahl et al., 2016). A commonly used genetically engineered mouse model of PDAC is the *LSL-Kras^{G12D/+};LSL-Trp53^{R172H/+};Pdx-1-Cre* mouse, or KPC mouse, which reproduces the development of cancer from premalignant lesions called pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasm cysts (Chugh et al., 2018).

Besides models developed by genetic engineering, mice may also be implanted with patient-derived tumors. These mice are known as xenograft models (Hasehira et al., 2021). Using mice also enables functional studies by knocking down or knocking out specific genes encoding certain proteins of interest and characterizing the resulting phenotypes. This strategy was employed to truncate O-glycans by knocking down the *Cosmc* chaperone protein encoding gene in Hofmann et al. (2015) and Wolters-Eisfeld et al. (2018) or the glycotransferase *C1galt1* encoding gene in Chugh et al. (2018). Though mouse models of disease have proven useful for research, one caveat is that

findings may not necessarily translate to the human forms of disease.

Nearly all the reviewed studies focus on glycans on proteins, but it is important to note that other glycoconjugate types exist as well. These species too have been examined in various pancreatic diseases. Glycans may also be covalently attached to lipids, known as glycolipids. Glycolipids are important in maintaining cell membrane stability and have been studied as PDAC markers (Yabu et al., 2013; Zhang T. et al., 2020). Proteoglycans are another type of glycoconjugate, consisting of glycosaminoglycan (GAG) chains branching from a small core protein. Proteoglycans are an important part of the extracellular matrix (ECM) involved in cell adhesion and migration. GAGs, specifically keratan sulfate, have been studied in the context of nephrotic response to hyperglycemia in T1D (Van et al., 2020).

Clinical Utility Considerations

Analyses in clinical settings have different considerations than those in a research laboratory. Clinical assays must have standard procedures and ideally are rapid and high-throughput analyses.

Many studies in the reviewed literature used blood components as the biological samples. Plasma and serum are ideal sample sources for clinical assays due to their relative abundance and ease of collection. Recent studies have focused on the most abundant species in blood as targets for biomarker discovery, bypassing the need for enrichment of less concentrated species or depletion of the most concentrated. These studies have investigated glycation of human serum albumin (HSA) as markers for diabetes to complement HbA1c (Korwar et al., 2015). Glycosylation of immunoglobulin G (IgG), the most abundant antibody in serum, has further been investigated as a source of markers for pancreatic cancer and pancreatitis (Chen et al., 2014; Shih et al., 2019; Shiao et al., 2020). The focus on IgG is also consistent with the immune facets of both pancreatic cancer and T1D. IgG glycosylation has also been investigated in T2D and DN (Liu et al., 2019; Singh et al., 2020a) Other studies have also focused on other acute phase proteins, including haptoglobin and AGP, as sources of markers for pancreatic diseases (Higai et al., 2003; Nakano et al., 2008; Sarrats et al., 2010; Kontro et al., 2014; Balmaña et al., 2016; Mancera-Arteu et al., 2019; Keser et al., 2021). While the exact mechanisms leading to regulation of these circulating glycoproteins in pancreatic diseases are still unknown, studies have shown their potential as biomarkers, thus warranting further study.

To expand access to assays, another consideration is whether specialized instrumentation or long sample preparation times are needed to conduct such analyses. Sample preparation protocols for MS analyses may range in timescales from minutes (so-called "dilute-and-shoot" methods) to days (bottom-up proteomics with overnight enzymatic digestions and long fractionations and separations), bottlenecking throughput. Multiplexing sample preparation, like with a multichannel automated liquid handler, can increase how many samples can be prepared at once and may offset time costs. IGPs, for example, often require long, nano-flow LC separation due to sensitivity needed to resolve the diversity of glycan compositions combined with varying peptide backbones. Throughput can be increased using multiplexed

analyses enabled by an isobaric tagging strategy where samples can be pooled. Furthermore, IGP analysis often requires an instrument with a fast scan speed and a high-resolution mass analyzer for accurate mass needed to distinguish peptide glycoforms. The length of LC separations and the need for specialized instrumentation required for deep glycopeptide coverage is thus not conducive for routine clinical use.

On the other hand, released glycans can be profiled with ease and intact using MALDI-time-of-flight (TOF) analysis of spotted samples on a target plate. Though mass resolution may be lower, mass accuracy can be preserved with TOF and glycan compositions can be inferred. Another advantage of the TOF mass analyzer is its wider mass range compared to other mass analyzers like the Orbitrap. Specific linkage information, though, would require derivatization or MS/MS fragmentation. This would usually require LC separation and an instrument with a fragmentation cell. Still, rich glycan profile information can be obtained using MALDI-TOF. Automated, high-throughput workflows for released glycan analysis using MALDI-TOF have been performed with the "Sweetblot" platform (Nouso et al., 2013; Akimoto et al., 2015). Example MALDI-TOF-MS spectra of released and derivatized N-glycans from various pancreatic duct and cancer cell lines can be seen in **Figure 3**. This figure showed rich glycan profile information in a single spectrum obtainable with MALDI-TOF without the need for LC separation.

For multiple reaction monitoring assays requiring high sensitivity and quantitative accuracy, triple quadrupole instruments have become a common feature of clinical laboratories (Shiao et al., 2020). Instruments with other mass analyzers can perform the similarly named parallel reaction monitoring which can also enable fast and sensitive quantification. These analyses are often paired with short LC gradients, increasing sample throughput while maintaining high analyte sensitivity.

MS is an even more powerful tool in combination with other technologies routinely used in the clinical lab, such as spectroscopy and microscopy. Since its first reporting in 1997, MALDI-MS imaging has been routinely used to analyze tissue sections at the molecular level with high-throughput (Caprioli et al., 1997; Buchberger et al., 2018). The resulting ion images generated with MALDI provide an orthogonal visualization for optical and spectroscopic images, which have different sensitivities and molecular specificities. Pancreatic cancer, for example, is diagnosed with the help of X-ray-based computerized tomography and radio-wave based magnetic resonance imaging (Kleeff et al., 2016).

Analyses of tumor tissue sections use stains to visualize cellular components. Glycans specifically can be visualized using Alcian blue and periodic acid-Schiff staining. The most ubiquitous stains, however, are hematoxylin and eosin (H&E). These stains are used for visualizing cellular nuclei and cytoplasm. Immunohistochemistry uses an antibody for visualizing certain proteins. Detection can be accomplished optically if a color changing reaction is performed. Fluorescence can also be used for detection if antibodies are conjugated to fluorophores. A recent work examined pancreatic cancer tissues using MALDI-
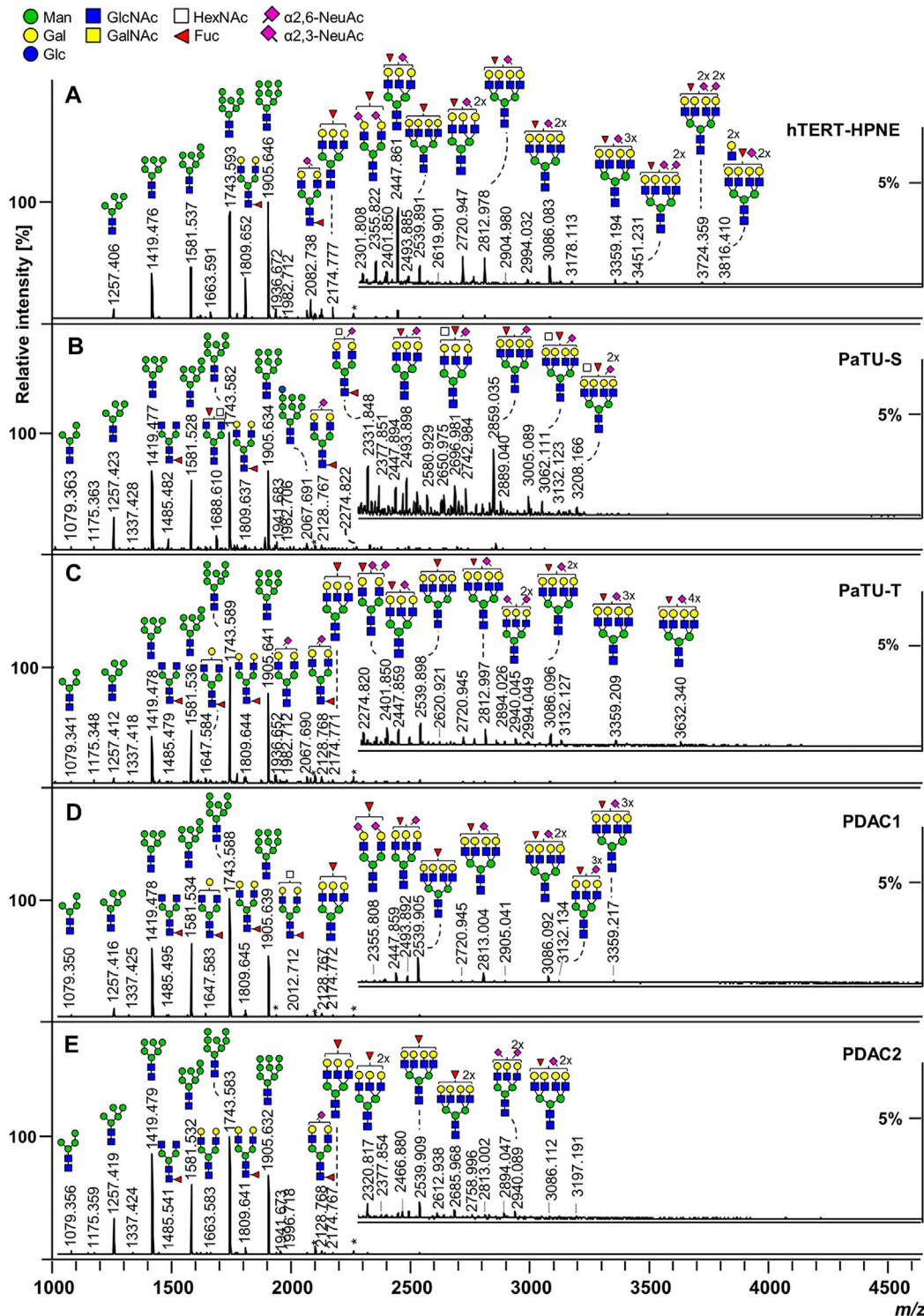
FIGURE 3 | MALDI-TOF-MS spectra of released and derivatized N-glycans from pancreatic duct **(A)** and cancer cells **(B–E)**. Reproduced from Holst, S., Belo, A.I., Giovannetti, E., Van Die, I., and Wuhrer, M. (2017). Profiling of different pancreatic cancer cells used as models for metastatic behaviour shows large variation in their N-glycosylation. *Sci. Rep.* 7, 16623. doi: 10.1038/s41598-017-16811-6 under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).
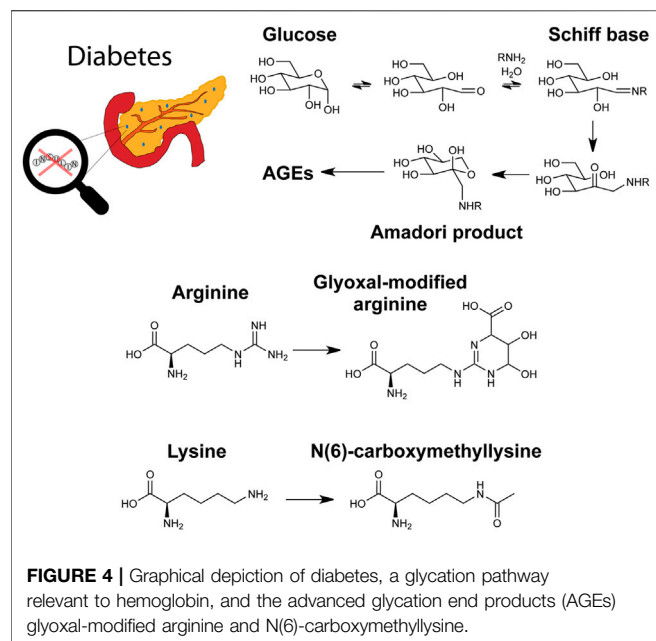
FIGURE 4 | Graphical depiction of diabetes, a glycation pathway relevant to hemoglobin, and the advanced glycation end products (AGEs) glyoxal-modified arginine and N(6)-carboxymethyllysine.

MS imaging in combination with immunohistochemistry and optical microscopy to identify potential N-glycan biomarkers and improve current diagnostic capabilities (Mcdowell et al., 2020). In short, the combination of orthogonal imaging modalities, including MALDI-MS imaging, further enhances diagnostic capabilities when combined with pathological annotation.

INSIGHTS INTO DIABETES

Most of the studies comprising the reviewed literature focused on T2D over T1D, which is consistent with T2D prevalence being much greater than T1D (Bullard et al., 2018). Gestational diabetes, a less common but nonetheless important form of hyperglycemia affecting pregnant women, has also been analyzed in earlier studies *via* MS-based glycomic and glycoproteomic methods (Lee et al., 2011; Smilowitz et al., 2013; Dupont et al., 2014). A pathway for hemoglobin glycation and recently studied molecules in diabetes are shown in **Figure 4**.

Summaries of the reviewed literature related to diabetes can be found in **Supplementary Table S1**.

Diabetes Biomarker Discovery

The long-held gold standard biomarkers used for diabetes diagnosis are concentrations of blood glucose and glycated hemoglobin (Lapolla et al., 2013; Wang et al., 2014). These measurements are typically accomplished *via* enzymatic or LC-based assays. As previously mentioned, glycation is concentration dependent. In hyperglycemia, blood glucose concentration is not controlled properly. It follows, then, that glycation of hemoglobin could be a marker for a diabetic state. There are downsides, however, to relying on HbA1c for diabetes diagnoses. HbA1c reflects average levels of blood glucose over a

long period, so it is not sensitive to short-term changes. Recent research efforts have been focused on identifying new biomarker species to complement those currently in use.

One study examined the erythrocyte, or red blood cell, glycated proteome across a range of HbA1c levels. After hemoglobin depletion, glycated proteins were enriched and analyzed *via* MS. Protein glycation was found to be dependent on HbA1c levels in extent and site-specificity, with 37 glycated proteins identified, nearly half of which were metabolic enzymes (Muralidharan et al., 2019). Another study identified glycation of the acute phase protein haptoglobin as a potential biomarker. Glycated K141 of haptoglobin was a glycation site that could be a biomarker for hyperglycemia. Since hemoglobin and haptoglobin have different half-lives, these two markers could then complement each other and improve sensitivity to short-term changes in blood glucose for diagnosing T2D (Soboleva et al., 2019).

Glycation in plasma proteins was also investigated for T1D biomarker discovery. In a study investigating glycemic control in T1D, two-dimensional LC-MS/MS was used to quantify 76 glycated peptides, with 6 of these peptides showing high correlation with HbA1c levels (Zhang and Zhang, 2020). A similar two-dimensional chromatographic approach was taken by the same group to study the AGE carboxymethyllysine (shown in **Figure 4**) in plasma proteins. This study quantified 58 modified peptides from 19 proteins with 57 sites. Five modified peptides were significantly higher in poor glycemic control compared to good control (Korwar and Zhang, 2021). In a different study on T1D, another AGE, modification with glyoxal (also shown in **Figure 4**), was studied on histones. Histones and their PTMs play a key role in DNA transcription, so glycation of histones should similarly affect processes. Biochemical characterization was performed on glyoxal-modified histones. T1D serum antibodies were also used to perform binding studies, revealing significant interactions with the histones (Ansari et al., 2018).

Though glycation and AGEs are a major focus of diabetes research, as recently reviewed in D'aronco et al. (2019), there is also a substantial body of work investigating glycosylation. N-glycosylation specifically has been recently reviewed as a source of biomarkers and drug targets (Rudman et al., 2019). Indeed, hyperglycemia has been linked to increased N- and O-GlcNAc glycosylation in T2D (Chatterjee and Thakur, 2018).

A broad analysis of plasma N-glycans was performed using T2D plasma. N-glycans were enzymatically released from plasma proteins and sialic acids were derivatized to enable linkage isomer resolution prior to spotting analysis. Seventy glycan compositions were identified. Compared to healthy controls, 18 glycosylation features were significantly associated with T2D. T2D was associated with higher α(2,6)-linked sialylation and sialylation of diantennary glycans and lower fucosylation and bisection of diantennary glycans and α(2,3)-linked sialylation of triantennary glycans (Dotz et al., 2018). These linkage isomer differences further emphasize the importance of glycan structure in these studies and the potential functional changes caused by these structural alterations.

FIGURE 5 | Chromatogram **(A)** and tandem mass spectra of glycopeptides **(B)**, MS^1 of glycoforms from glycosylation site I$_1$; **(C)**, MS^2 of the peptide part; **(D)**, MS^2 of the glycan (N-acetylglucosamine)$_5$(Hexose)$_6$(Sialic acid)$_3$ from human alpha-1-acid glycoprotein enriched from plasma. Reproduced from Keser, T., Tijardović, M., Gornik, I., Lukić, E., Lauc, G., Gornik, O., et al. (2021). High-throughput and site-specific N-glycosylation analysis of human alpha-1-acid glycoprotein offers a great potential for new biomarker discovery. *Mol. Cell. Proteomics*, 100044. doi: 10.1074/mcp.RA120.002433 under a Creative Commons CC-BY license (https://creativecommons.org/licenses/by/4.0/).

Two studies on N-glycosylation in T2D took a more focused approach by narrowing analysis to one species. IgG glycosylation was investigated in a Uyghur population in China at the IGP level. Analyses identified 27 directly measured and 4 derived glycan traits that were significantly associated with T2D, including decreased bisecting GlcNAc of IgG2 and agalactosylation of IgG4 and increased sialylation of IgG4 and digalactosylation of IgG2 (Liu et al., 2019). AGP glycosylation was investigated in a site-specific and high-throughput manner in T2D plasma at three timepoints. A one-step precipitation was performed using perchloric and phopshotungstic acids in a 96-well plate to perform the enrichment of AGP. A chromatogram and mass spectra of some IGPs from AGP can be seen in **Figure 5**. This figure shows the power of LC-MS in separating intact glycopeptides and fragmenting their constituent peptide backbones and glycans for confident identification and quantification. Moreover, this analysis identified markers of higher risk for T2D, including increasing branching at the second glycosite and decreased sialylation at the third glycosite (Keser et al., 2021).

Further recent studies focused on HSA. Glycation of albumin was previously found to affect its capability to bind non-esterified fatty acids. Glycated albumin in T2D had lower binding capacity, leading to higher plasma concentrations of the lipid species. This in turn contributes to platelet hyperactivity and increased thrombosis observed in T2D, again underscoring the importance of modified protein function resulting from modifications on observed phenotypes (Blache et al., 2015). More recently, glycation of HSA was investigated for its effect on antibodies, where it was found to enhance neo-epitope generation, leading to immunological complications (Raghav et al., 2017). Glycated HSA was also investigated in the context of modified pharmacokinetics. Out of 49 glycation sites identified in this work, the modification site K199 was found to be most changed in diabetic plasma. Both glycation sites and drug binding sites would be expected to be modulated by different accessibilities on the protein, so molecular docking simulations were performed to analyze differential pharmacokinetics. Heparin was found not to be significantly affected, but warfarin had higher binding affinity with glycated HSA. These two drugs are common anticoagulants prescribed for patients with cardiovascular disease. These findings show that therapeutic efficacy and safety of common drugs are changed due

to protein glycation. Further drug-specific studies are suggested to see how other drugs are affected by protein glycation resulting from diabetic status (Qiu et al., 2020).

Overall, the reviewed literature suggests a focus on and need for complementary diabetes markers besides blood glucose and HbA1c. An increased panel of markers would increase diagnostic sensitivity and specificity. Studies with large cohorts that are diverse in age, race, and sex, however, are needed to ensure experimental rigor for discovery of novel biomarkers (Oh et al., 2015).

Diabetes-Related Complications

Though diabetes is fundamentally a disease of the pancreas, specifically of the beta cells and of insulin dysfunction, the possible complications resulting from the disease are far-reaching. Common complications of diabetes include cardiomyopathy, nephropathy, and retinopathy. Biomarker discovery for diabetes-related complications is also a substantial body of work, as complications may develop long after the initial diabetes diagnosis.

DN has been investigated using both mouse models and human biological samples. One study used db/db and streptozotocin-induced diabetic mice with and without insulin treatment. Kidney tissue proteins were first extracted, with tryptic glycopeptides enriched using hydrazide chemistry. Glycans were oxidized to aldehydes and were immobilized on a hydrazide solid phase. Their formerly glycosylated peptides were then eluted and analyzed *via* MS, revealing that insulin treatment changed glycoprotein levels compared to control. Implicated proteins were related to cell adhesion and cell-matrix composition (Liljedahl et al., 2016). DN in T1D was also investigated regarding effects of glycemic control. Glycans were released from serum proteins, identifying 39 glycans, 24 of which were from IgG. Glycan quantification correlated with HbA1c, with higher HbA1c associating with decreased numbers of simple biantennary glycans but also higher branching, galactosylation, and sialylation. These changes are consistent with prior findings implicating the epidermal growth factor receptor and transforming growth factor-β pathways in kidney disease (Bermingham et al., 2018). Another study of plasma from T2D patients examined IgG N-glycosylation in DN. The glycosylation profile was linked to declining kidney function, with the inflammatory potential of IgG mediated by its glycosylation. This finding underscores the importance of the immune system in T2D and resulting nephropathy, and not just in T1D (Singh et al., 2020a).

Eye-related complications may also result from diabetes. Diabetic retinopathy is the primary cause of blindness today. Two studies analyzed diabetic retinopathy using serum and cells from the eyes (Sharma et al., 2020; Ramos-Martínez et al., 2021). Human serum was analyzed from diabetic patients with and without retinopathy. IGPs were enriched prior to MS analysis, which identified 15 IGPs from 11 glycoproteins that were significantly altered in retinopathy. The implicated proteins included fibronectin, hemopexin, and vitronectin, which are glycoproteins involved in peptidolytic processes and in chronic low-grade inflammation (Sharma et al., 2020). Epithelial cells

from the lenses of T2D patients were analyzed in the context of cataract development. Proteins were separated using two-dimensional gel electrophoresis, which in combination with lectin immunohistochemistry, identified glycosylated protein targets for in-gel tryptic digestion and MS analysis. The findings of this work were focused on the inflammatory component of eye disease and on the protein type 1 cytokeratin. Over-expression of this N-glycoprotein may contribute to increased permeability of the lens and cataract formation (Ramos-Martínez et al., 2021).

Diabetes Mechanisms and Effects of Diabetes Treatment

Mechanisms of diabetes progression and treatments for diabetes have also been studied recently. Using a mouse model, the impact of O-glycosylation on pancreatic function was assessed. The *Cosmc* molecular chaperone encoding gene was knocked down selectively in pancreatic acinar cells, leading to truncated O-glycosylation and a loss of core 1 glycan formation. Protein carriers of the truncated O-glycan Tn antigen were identified using MS after lectin pulldown and in-gel digestion. One of the identified proteins was carboxyl ester lipase, or Cel, which is known to be heavily O-glycosylated and expressed in the acini. Compared to the wild-type mice, the Cosmc knockdown mouse presented a diabetes-related (maturity-onset diabetes of the young type 8) phenotype, as evidenced by impaired beta cell function. These results link the truncated O-glycosylation of Cel with the diabetes-related phenotype and exocrine dysfunction. O-glycosylation can contribute to overall protein stability and regulate proteolysis, suggesting that glycan truncation plays a role in protein instability in exocrine dysfunction (Wolters-Eisfeld et al., 2018). In sum, this study highlights the power of MS in identifying glycoproteins directly affected by knocking down an important molecular chaperone to study mechanisms of diabetes.

The db/db mouse model was used to investigate the mechanism by which the traditional Chinese medicine *B. batryticatus*, specifically the active compound 1-deoxynojirmycin, relieves diabetic cardiomyopathy. Myocardium tissue from mice dosed with the drug were analyzed, and formerly N-glycosylated peptides were subjected to MS analysis. N-glycosylation was decreased overall due to treatment, but α(1,6) fucosylation was elevated. The authors suggest that the mechanism for relief of cardiomyopathy stems from inhibition of N-GlcNAc formation and the reduction of substrate concentration (Zhao et al., 2018).

A study of T2D human plasma examined the associations of N-glycosylation and treatments for diabetes and cardiovascular disease. The investigated treatments included metformin, statin, sulfonyl urea derivatives, and insulin. This work concluded that metformin and statins associate with N-glycosylation, suggesting a common biological effect (Singh et al., 2020b).

INSIGHTS INTO PANCREATIC CANCER

The importance of glycosylation in cancer, including pancreatic cancer, is well-established (Drake, 2015; Taniguchi and Kizuka,

FIGURE 6 | Graphical depiction of pancreatitis, pancreatic cancer, and several related O-glycans (with glycosidic linkages shown) investigated as biomarkers. R = peptide.
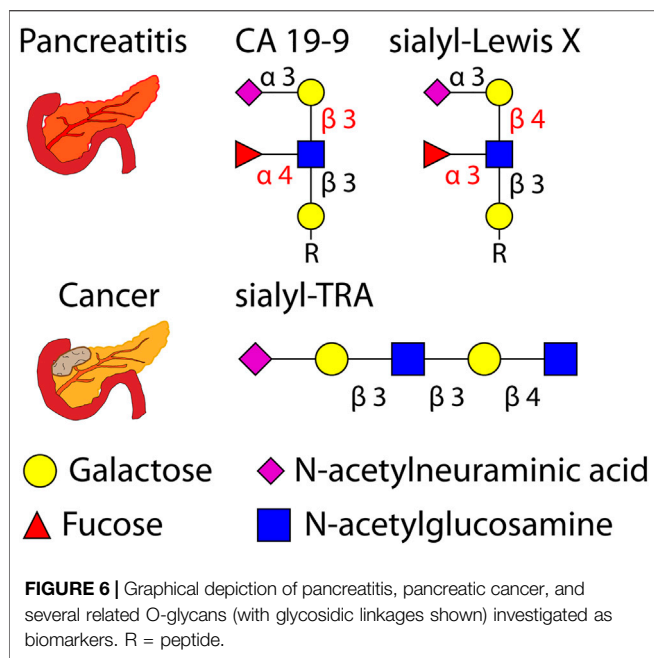
2015; Mereiter et al., 2016; Pan et al., 2016; Kailemia et al., 2017). One of the characteristic features of pancreatic cancer is the desmoplastic reaction, or desmoplasia. This phenomenon involves rapid growth of fibrous tissue around cancer cells. This fibrous tissue is enriched in various cell types but also ECM proteins, which are heavily glycosylated. This dense environment changes cellular properties, and it has been suggested that desmoplasia is a cause of the chemoresistance of pancreatic cancer (Whatcott et al., 2012). The role of the ECM in PDAC-stromal interactions was recently reviewed (Liot et al., 2021).

Chemoresistance, plus difficulty in diagnosing the pancreatic cancer until late stages when metastasis has already occurred, is a major reason for the poor prognosis of this disease. Much of the work on pancreatic cancer is thus focused on biomarker discovery for earlier diagnosis.

Protein and glycan species have been thoroughly investigated for biomarker potential in pancreatic cancer, though widely approved biomarkers remain elusive. Biomarker species have been studied, however, in monitoring response to pancreatic cancer treatment. Carbohydrate antigen 19-9 (CA 19-9), also known as sialyl-Lewis A, is an O-glycan that has been widely studied in pancreatic cancer. Its diagnostic potential, though, is blunted by its low specificity and poor predictive power. Thus, CA 19-9 is mostly used for monitoring response to treatment and is the only test for pancreatic cancer approved by the US Food and Drug Administration. The glycoprotein carcinoembryonic antigen (CEA) is also studied for diagnosis, though its use also suffers from low accuracy. Continued research efforts seek to discover novel biomarker molecules for earlier diagnosis of pancreatic cancer. These novel molecules could complement those currently studied to improve diagnostic power through biomarker panels over singular species. A recent review offers a clinical perspective on these molecules and other proteomic and glycomic biomarkers (Hanna-Sawires et al., 2021).

Though pancreatic cancer prevalence is low compared to other cancers (3% of all cancers are pancreatic), its five-year survival

rate is among the lowest of all cancers at approximately 10% (Siegel et al., 2021). Risk factors for pancreatic cancer common to other diseases include age, family history, tobacco use, and being overweight. Noteworthy, however, is that other pancreatic diseases, such as diabetes and chronic pancreatitis, also increase the risk of developing pancreatic cancer. The inflammation in pancreatitis may be caused by tobacco and alcohol use, which are both risk factors for cancer, but the exact mechanism by which diabetes and pancreatic cancer are linked is still unknown (Cho et al., 2020).

Recent efforts have often studied pancreatitis and pancreatic cancer together, as the two diseases can lead to similar symptoms. Unnecessary surgery resulting from misdiagnosis can be avoided by discovering biomarkers specific to either disease. Glycoproteins and glycans are frequent targets for such efforts, with recent studies investigating the O-glycans sialyl-Lewis X and sialyl-TRA as biomarkers (Barnett et al., 2017; Guerrero et al., 2021).

Graphical depictions of pancreatitis, pancreatic cancer, and glycans implicated in the two diseases are shown in **Figure 6**.

Summaries of the reviewed literature related to pancreatic cancer can be found in **Supplementary Table S2**.

Pancreatic Cancer Biomarker Discovery

In pursuit of earlier diagnoses, studies have also investigated pancreatic lesions and cysts that may progress into malignancy. Pancreatic intraepithelial neoplasia involves lesions that are small and are not filled with fluid. Higher grade lesions may progress into cancer. Intraductal papillary mucinous neoplasms are larger pockets of tissue and are filled with fluid. Fluid from these cysts have been investigated as biomarker sources previously (Mann et al., 2012; Cao et al., 2013). Serum glycome changes have also been investigated in patients with invasive intraductal papillary mucinous neoplasms (Akimoto et al., 2015).

As mentioned previously, currently used markers CA 19-9 and CEA are not accurate or specific enough for diagnostic use with pancreatic cancer, so efforts have been focused on discovering other molecules or molecular qualities that can discriminate pancreatic diseases. One quality of glycan profiles, sialylation, has been studied in earlier work and continues to be a target in more recent work (Almaraz et al., 2012; Amano et al., 2012). Another glycan profile quality upregulated in pancreatic cancer is increased high-mannose type glycosylation, and a high-mannose type glycan was found to associate with the efficacy of gemcitabine treatment for the disease (Miyahara et al., 2015).

Several of the reviewed studies on pancreatic cancer biomarker discovery used serum or plasma. In one study, serum from a repository of samples taken for ovarian cancer screening was analyzed. Samples were separated by time to pancreatic cancer diagnosis. After analysis of formerly glycosylated peptides, 167 N-glycoproteins were quantified, though no gross differences were identified across the five time-to-diagnosis groups. Altered proteins found were related to the inflammatory response and coagulation (Krishnan et al., 2017). Another study analyzed released N-glycans from serum. Glycans were separated using porous graphitic carbon to enable isomer resolution, with 280 isomers identified corresponding to 72 glycan compositions. Compared to control, cancer serum had 25 significantly

different isomers with sialylation and fucosylation of these glycans being of particular interest (Liu et al., 2018). Sialic acid linkage isomers were investigated using an automated liquid handler. Derivatization was performed to differentiate α(2,3) and α(2,6)-linked sialic acids prior to MALDI spotting analysis using the high mass resolution Fourier transform-ion cyclotron resonance (FT-ICR) mass analyzer. This work found PDAC serum to have higher branching, (antenna)fucosylation, and α(2,6) vs. α(2,3)-linked sialylation compared to control serum (Vreeker et al., 2020).

O-glycosylation of serum proteins has also been investigated to differentiate between pancreatic and gastric cancers from control. One study focused on sulfated O-glycans, which are non-sialic acid linked glycans as are the glycan markers CA 19-9 and sialyl-Lewis X (both of which are shown in **Figure 6**). O-glycans were first released *via* hydrazinolysis, a process which can limit the undesirable "peeling" of glycans that can be observed in base-catalyzed elimination of glycans. Sulfated glycans were enriched following enzymatic digestion with α-neuraminidase from *A. ureafaciens* to remove sialic acids regardless of linkage position. Fourteen candidate marker molecules were identified (Tanaka-Okamoto et al., 2017). Another study from the same group focused on internal sialylation of O-glycans in serum. Enzymatic digestion using α-neuraminidase from *S. typhimurium* cleaved non-internal sialic acid residues, enabling the quantification of 17 marker candidates. Precise structural analyses were later done to confirm glycan structures of the marker candidates (Tanaka-Okamoto et al., 2018).

Urine is also a plentiful source of biomarker candidates. A recent study extracted endogenous glycopeptides from urine followed by capillary electrophoresis separation coupled to MS. Urine was collected from patients with several cancers, including pancreatic cancer. Capillary electrophoresis separates molecules in an orthogonal manner to LC based on electrophoretic mobility instead of interactions with a stationary phase. This type of separation can be performed with much less sample volume, down to nanoliters as opposed to microliters as is routine in LC. Using this separation, 37 O-glycopeptides and 23 N-glycopeptides were identified. Three O-glycopeptides differed among conditions with statistical significance, while 5 N-glycopeptides also exhibited differences with statistical significance (Belczacka et al., 2019).

Besides biofluids such as serum, plasma, and urine, other studies have instead used pancreatic cancer cells and tissues as sources of biomarkers.

A study used two pancreatic cancer cell lines with opposite morphology and cell behavior to examine differences in O-glycoproteins and O-glycolipids. Indeed, differing O-glycan profiles were found between the mesenchymal-like and epithelial-like cell lines examined (Zhang T. et al., 2020). Pancreatic ductal and cancer cell lines were similarly compared in a study examining N-glycosylation. This work compared two primary PDAC cell lines and two cell lines of liver metastases of PDAC. N-glycans were released from proteins and derivatized to enable sialic acid linkage isomer resolution. Major differences in glycosylation were observed between all four cell types, suggesting large heterogeneity even among cell types of the same disease. This work suggests that conclusions derived

from one cell line may not translate to other cell lines (Holst et al., 2017). The cell line Suit2-007 was used to investigate lactosyl-sepharose binding proteins, which are involved in metastasis. MS was used to examine altered binding of these proteins to a lactosyl resin after affinity chromatography. Calcium and galactose were found to alter the binding of these proteins, and the authors concluded that galactose may have a potential use as a therapeutic for PDAC in inhibiting cell proliferation (Sagini et al., 2021).

Several of the reviewed studies used primary tumor tissue as a source of biomarkers. Previous MALDI-MS imaging analyses of tumor tissue have revealed molecular markers of tumor progressions and morphology, and a recent study examined tumors from PDAC patients. Released N-glycans were mapped across pancreatic and pancreatic tumor tissues using MALDI-quadrupole-TOF and MALDI-FT-ICR. Adding a quadrupole prior to the TOF mass analyzer helps improve mass accuracy and can enable MS/MS analyses. N-glycans from PDAC tissue sections visualized using MALDI-MS imaging can be seen in **Figure 7**. This figure shows that glycan distributions can mirror cancer and necrosis distributions in pancreatic tumors, which may complement pathological annotations. The results of these imaging experiments were combined with immunohistochemistry to improve PDAC identification using one modality alone, again emphasizing the improved predictive power that results when MS-based analyses are combined with other clinical techniques. This study additionally used the endo F3 enzyme to distinguish core and antennary fucosylation and derivatization to distinguish linkage isomers. For immunohistochemical analyses, CA 19-9 and sialyl-TRA were stained and visualized using immunofluorescence. These two glycan markers were previously found to define separate subpopulations of pancreatic cancer cells (Barnett et al., 2017). These analyses identified in PDAC tissues increased sialylation, poly-LacNAc extensions, branching, and fucosylation of high-mass glycans (Mcdowell et al., 2020).

Other studies have instead extracted the protein content from tumor tissues. In a study examining both pancreatic cancer cell lines and primary tumor tissues, extracted proteins were first separated by two-dimensional gel electrophoresis. To identify proteins containing sialyl-Lewis X, which has been associated with PDAC, Western blot was first performed to identify antigen carrier protein bands, then in-gel digestion was done to generate peptides for MS analysis. This analysis found that microfibril-associated protein 4 is a carrier of sialyl-Lewis X and was upregulated in cancer compared to control. The authors further hypothesize a link between upregulation of the protein and PDAC desmoplasia (Guerrero et al., 2021).

An untargeted quantification approach was taken in a separate study examining tumor and normal adjacent tissue. Isobaric tagging was performed using an isotopic diethyl label pooling of tissue pairs together in one analytical run. LC-MS analysis identified 20,038 intact N-glycopeptides, with 38 up-regulated and 14 down-regulated in cancer compared to control. These peptides correspond to proteins involved in several processes, including transporter and catalytic activity and binding (Lu et al., 2021). Another work studied two patient-derived xenograft mouse models to compare the effects of cancer differentiation
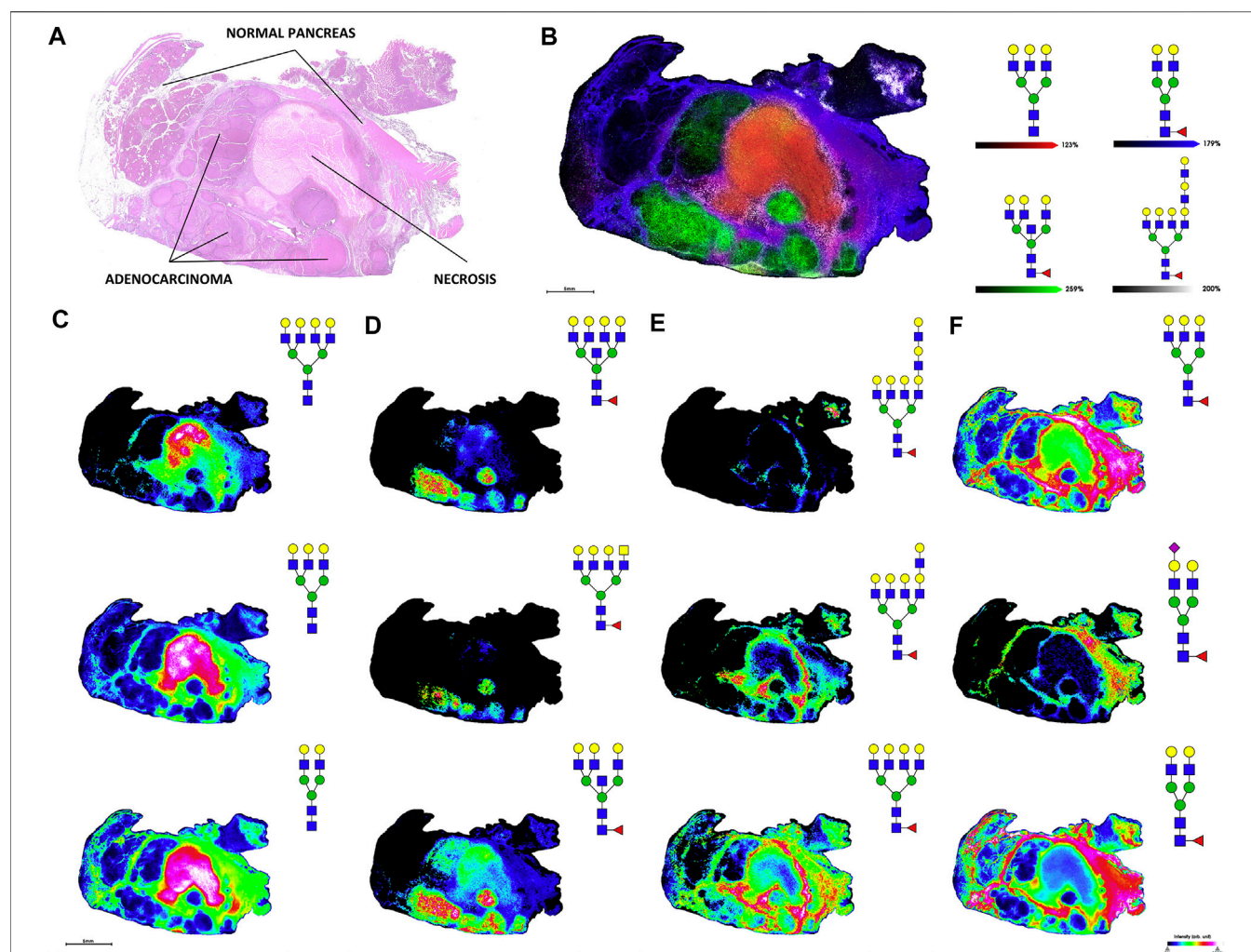
FIGURE 7 | H&E stain **(A)** and MALDI-FT-ICR MS images of N-glycans **(B)**, overlay of four glycans corresponding to necrotic tissue, adenocarcinoma, tumor margin, and adjacent non-tumor tissue; **(C)**, glycans from necrotic tissue; **(D)**, glycans from adenocarcinoma; **(E)**, glycans from tumor margin; **(F)**, glycans from adjacent non-tumor tissue) released from stage 3 pancreatic tumor tissue. Reproduced from McDowell, C. T., Klamer, Z., Hall, J., West, C. A., Wisniewski, L., Powers, T. W., et al. (2020). Imaging Mass Spectrometry and Lectin Analysis of N-Linked Glycans in Carbohydrate Antigen-Defined Pancreatic Cancer Tissues. *Mol. Cell. Proteomics* 20, 100012. doi: 10.1074/mcp.RA120.002256 under a Creative Commons CC-BY license (https://creativecommons.org/licenses/by/4.0/).

on both the N-linked and O-linked glycomes. Glycans were first released *via* hydrazinolysis before analysis with MALDI-TOF. The well-differentiated PC3 samples had higher branching and sialylation of complex N-glycans compared to the poorly-differentiated PC42 samples. PC3 had a lower percentage overall of core 1 O-glycans identified but a higher percentage of core 3 type glycans compared to PC42 (Hasehira et al., 2021).

Differentiation of Pancreatic Cancer From Pancreatitis

Biomarker disease discrimination is especially important between pancreatitis and pancreatic cancer. A recent review examined biomarkers specifically for these two diseases (Chou et al., 2020). An earlier work examined serum from patients diagnosed with various pancreatic ailments, including pancreatitis, pancreatic

cysts, and T2D for biomarker discovery. This work identified a panel of glycoproteins that could help discriminate pancreatic cancer, again emphasizing the utility of multiple molecules over single targets for diagnostic potential (Nie et al., 2014).

More recently, studies have focused on glycosylation in studies distinguishing pancreatitis from pancreatic cancer and from healthy controls. AGP N-glycans were analyzed after affinity purification from serum. Stable isotope labeling of glycans was performed using isotopic aniline to introduce light and heavy carbon labels. Pooled healthy controls were labeled and combined with pancreatic cancer and chronic pancreatitis samples for quantification using capillary electrophoresis separation with spectroscopic detection. MS analysis identified increased AGP $\alpha(1,3)$ fucosylation in cancer compared to pancreatitis and control. Furthermore, pancreatitis could be differentiated from cancer using an enzyme-linked lectin assay with *A. aurantia* lectin (Balmaña et al., 2016). A similar isotopic

labeling strategy of AGP glycans was performed in a separate study (also using serum) prior to multivariate data analysis for biomarker discovery. This analysis identified seven glycan isomers with α(2,6)-linked sialylation that could be used to differentiate chronic pancreatitis from pancreatic cancer (Mancera-Arteu et al., 2019).

Chronic pancreatitis was compared with pancreatic cancer in another study using plasma by analyzing formerly N-glycosylated peptides with data-independent acquisition, which fragments all precursor ions in a certain m/z range instead of picking individual precursors for MS/MS. This strategy can increase the robustness of quantification, which can suffer from missing values based on the stochastic nature of data-dependent acquisition. N-glycosylation changes in circulating galectin-3 binding protein, or LGALS3BP, glycoforms were detected, suggesting changes in function during cancer progression (Nigjeh et al., 2017).

IgG glycosylation has also been a target for studies comparing pancreatitis and pancreatic cancer, especially pancreatitis inflammation caused by the patient's own immune response. Autoimmune pancreatitis serum was collected from 86 patients and compared to 115 PDAC and 57 control samples, with a further validation cohort also recruited. Sixteen glycoforms were detected *via* MS after tryptic digestion, revealing higher fucosylation of IgG1 and higher sialylation of IgG subclasses 1, 2, and 4 compared to PDAC (Shih et al., 2019). A later work developed an on-bead enzymatic protein elution method for quantification of IgG glycosylation. A unique feature of this study is the incorporation of isotopically-labeled IgG as an internal standard to improve quantitative accuracy. Quantification was performed on serum using multiple reaction monitoring on a triple quadrupole mass spectrometer, identifying 7 IGPs for differentiating PDAC from autoimmune pancreatitis (Shiao et al., 2020).

Pancreatic Cancer Mechanisms

Two recent studies used mouse models of pancreatic cancer to study mechanisms related to glycosylation and disease phenotypes (Chugh et al., 2018; Engle et al., 2019).

O-glycan truncation is a marker of pancreatic cancer observed in patient samples. Processing of O-glycans that normally prevents truncation is performed by the core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 (C1GALT1) protein. To induce O-glycan truncation in a pancreatic cancer mouse model, KPC mice were crossed with mice that had the *C1galt1* gene floxed and thus not expressed. This led to the loss of core 1 glycans, which was monitored using MS and lectin pull-down assays. These crossbred mice bearing truncated O-glycans had significantly shorter survival lengths than the control KPC mice. These mice further developed precancerous lesions, PDAC, and metastases earlier and more aggressively than the KPC mice. C1GALT1 knockout was also investigated in human cells, which revealed glycan truncation on the MUC16 protein, a known cancer marker. Together, these results suggest that loss of C1galt1 led to the more aggressive phenotype than the one in KPC mice and that the MUC16 protein plays a role in pancreatic cancer (Chugh et al., 2018).

As discussed previously, CA 19-9 can be used as a marker for monitoring treatment in pancreatic cancer, but its diagnostic

power is blunted due to its low specificity as a marker in other diseases. The exact mechanisms and links between the marker and disease phenotypes have not been widely studied previously, however. To investigate the links between CA 19-9 expression and pancreatic disease phenotypes, two proteins were inducibly expressed in mice. Human fucosyltransferase 3 and β1,3-galactosyltransferase 5 are needed to generate CA 19-9, so both proteins were first transduced in mouse cells. Protein carriers of the CA 19-9 glycan in the mouse cells were analyzed using MS. Known human protein carriers of CA 19-9 were indeed identified, suggesting a successful recapitulation of the human phenotypes in mouse cells. This was repeated to generate a mouse model of CA 19-9 expression in the pancreas. These mice developed acute pancreatitis that advanced into chronic pancreatitis after 28 days, followed by development of precancerous cysts and pancreatic tumors. As mentioned previously, pancreatitis is a risk factor for PDAC. This work showed causal relationships between CA 19-9 expression and development of both pancreatitis and aggressive pancreatic cancer, illustrating a mechanism in which CA 19-9 is not just a marker of disease, but a driver of pancreatic disease progression (Engle et al., 2019).

In sum, pancreatic cancer disease mechanisms have been characterized through protein gene knockdown which can affect glycosylation. These studies illustrate the importance of protein glycosylation on cancer progression.

DISCUSSION

The past five years have provided a rich body of work using MS-based analysis of glycans and glycoproteins to derive insights into pancreatic diseases. Nevertheless, the field faces several challenges, some of which are described here. Developing research areas and future directions for the field will also be discussed later in this review.

Current Challenges

Continued research strives to provide earlier diagnoses of both diabetes and pancreatic cancer through novel biomarker discovery. For diabetes, the sensitivity of the currently used biomarkers of blood glucose and glycated hemoglobin could be further improved or complemented using a panel of markers. Research has mostly focused on T1D and T2D, though gestational diabetes is also an important target for biomarker discovery. Another important consideration in gestational diabetes is how the short and long-term health of the developing fetus is affected. Diabetes-related complications would also benefit from earlier diagnoses, as complications may take years from diagnosis to present. Organ dysfunction resulting from diabetes complications can be treated earlier to prevent organ failure with earlier diagnoses.

Previous discussion focused on studies of glycation and glycosylation in pancreatic disease, but explicit mechanisms of how these modifications contribute to disease phenotypes remain elusive. The chemoresistance of PDAC, likely due in part to desmoplasia, is a major barrier to successful treatment of the

disease. Exact mechanisms of this chemoresistance are still unknown as well. Studies using mouse models to knockdown genes involved in protein glycosylation can shed light on how aberrant glycosylation affects disease progression. These types of studies have been helpful in linking CA 19-9 and truncated O-glycosylation, known markers of cancer, to pancreatic cancer development (Chugh et al., 2018; Engle et al., 2019). Similar studies may also be possible to investigate the roles of glycosylation in diabetes-related complications.

Heterogeneity in diabetes and pancreatic cancer makes research more difficult. In diabetes, the number of possible AGEs that can result from protein glycation is quite large. Untargeted analyses for discovery of AGEs are possible with MS-based analyses, though functional analyses of various AGEs have considerably lower throughput. Several of the reviewed studies demonstrated protein functional changes imparted by different glycation modifications. Elucidating the structure-function relationship with AGEs on proteins in diabetes and its related complications will continue to be a challenge.

In pancreatic cancer, disease heterogeneity manifests in tumor proliferation and metastasis. The difficulty in finding sensitive and specific biomarkers for pancreatic cancer may be a direct result of the immense heterogeneity of pancreatic cancer tumorigenesis, proliferation, and metastasis. The importance of glycosylation, which can possess vast structural heterogeneity, in pancreatic cancer further emphasizes heterogeneity in the disease. As shown by Holst et al. (2017), even cell lines of the same disease can present different glycosylation profiles.

In studies examining pancreatic diseases, care must be taken in countering disease heterogeneity with large sample cohorts that are diverse in age, race, and sex, with appropriate binning of samples into experimental groups. Pancreatic cancer, for example, is more common in those over 65 years old, in African Americans compared to Whites, and in men compared to women. Addressing these risk factors in cohort recruitment can increase confidence in translating the conclusions drawn from the work.

Future Perspectives

Recent biological insights, analytical developments, and commercialization of new technologies may be influential in affecting the directions of pancreatic disease research.

Analyses of glycosylation on tissue sections have grown more popular as MALDI-MS imaging instrumentation has developed. As shown by Mcdowell et al. (2020), MALDI-MS imaging can complement clinical imaging modalities due to its high molecular content and throughput. One downside to MALDI analyses, however, is a lack of analyte separation that may suppress ionization of lower abundance analytes. Current developments in instrumentation have sought to improve analyte sensitivity, including by adding ion-mobility separation after ionization before mass analysis. Since glycan linkage isomers have different structures and conformations, they behave differently while moving through or interacting with a drift gas. This is the principle that enables separation of isomers in ion mobility spectrometry. MALDI-MS imaging using trapped ion-mobility time-of-flight, or timsTOF, was recently demonstrated (Spraggins et al., 2019). Other ion mobility forms, such as drift-

tube, traveling-wave, and field asymmetric ion mobility, have been coupled to mass spectrometry imaging and are explored in more detail in a recent review (Sans et al., 2018).

Another approach for improving sensitivity in MALDI-MS is pseudo-enrichment *via* improving ionization efficiency of analytes. Our lab recently used hydrazide chemistry to improve the sensitivity of N-glycans in MALDI-MS imaging. Using Girard's reagent P to react with the reducing end of glycans, a positive charge is introduced, improving ionization efficiency and thus sensitivity. Sensitivity was boosted up to 230-fold for glucose. This method was then demonstrated to improve N-glycan sensitivity in N-glycan imaging of laryngeal cancer tissues (Zhang H. et al., 2020). This method and other on-tissue chemical derivatization strategies were thoroughly reviewed recently (Harkin et al., 2021; Zhou et al., 2021).

Laser-induced postionization, or MALDI-2, is another recently commercialized improvement to mass spectrometers that can enhance glycan detection sensitivity (Soltwisch et al., 2015). Using a second laser, the analyte plume generated by the first laser pulse is further irradiated. This second irradiation improves bare ion formation to increase signal and thus improve sensitivity. Laser-induced postionization has been shown to increase N-glycan sensitivity in imaging applications (Heijs et al., 2020).

MS has not been constrained to tissue analyses on slides. *In situ* measurement of molecules in patients on the surgical table has been made possible with technologies like the iKnife (Balog et al., 2013) and MassSpecPen (Zhang et al., 2017). These handheld instruments can be used by surgeons to enable molecular analyses from tissues in real time to help map tumor margins. As glycans have been investigated as tumor markers, often against normal adjacent or healthy control tissues, it follows that real-time glycan analysis in the operating room may help discriminate tumor margins. One caveat, however, is that analyses of glycans often require numerous sample preparation steps, including glycan release and purification from other matrix components. Developing the methods to expand the surgical toolkit to involve glycosylation may prove useful in discriminating cancer tissues from healthy tissues in real time.

While sugars play a major role in protein structure and function, there are numerous other PTMs that can similarly affect proteins. Efforts to maximize the PTM information obtained with MS workflows have gained traction in recent years. Using LC-MS, glycosylation is often studied with phosphorylation, as they are among the most common PTMs (Glover et al., 2018; Cho et al., 2019; Cui et al., 2019; Zhou et al., 2020; Cui et al., 2021; Huang et al., 2021). Both PTMs are implicated in cell signaling, and O-glycans may modify Ser and Thr, also two possible residues for phosphorylation. Analyses of such cross-talk interactions, including that of phosphorylation and O-GlcNAcylation, may help unravel mechanisms of disease (Wang et al., 2008).

In MALDI-MS imaging analyses, tissues may be analyzed multiple times to detect different analyte classes. By spraying GAGases, PNGase F, and trypsin sequentially, the same tissue section can be used to map proteoglycans, N-glycans, and peptides, as surrogates of their precursor proteins, in the same experiment (Turiak et al., 2014; Clift et al., 2021). Furthermore,

FIGURE 8 | Top-down (**A,C**) and bottom-up tandem mass spectra (**B,D**) of an O-glycoform of an insulin chain from mouse pancreatic islets. Reproduced with permission from Yu, Q., Canales, A., Glover, M. S., Das, R., Shi, X., Liu, Y., et al. (2017). Targeted Mass Spectrometry Approach Enabled Discovery of O-Glycosylated Insulin and Related Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal. Chem.* 89, 9184–9191. doi: 10.1021/acs.analchem.7b01926. Copyright © 2017 American Chemical Society.

derivatization reagents and exoglycosidases can also be sprayed on tissues to obtain linkage-specific information (West et al., 2021). Overlaying of these images can then be done to reveal co-localization of analytes. Such multiomic workflows are useful to maximize the information gained when patient samples are precious and limited.

Glycosylation analysis *via* MS in pancreatic diseases has primarily focused on proteins, but glycosylation can also

modify hormones. Our lab recently investigated O-glycosylation in pancreatic hormones isolated from mouse islets. Mass spectra of an O-glycosylated isoform of insulin are shown in **Figure 8**. This figure shows confident identification of the glycan composition (panels A–B) and the peptide backbone (panels A–C) using hybrid MS/MS Electron-Transfer/Higher-Energy Collision Dissociation (EThcD) fragmentation. This

fragmentation mode also enabled confident site localization of the glycan (panel D). This work was the first report of glycosylated insulin-B chain and insulin-C peptide (Yu et al., 2017). Hormone glycosylation has not been widely studied *via* MS. Similarly, the roles of hormone glycoforms in pancreatic diseases are still unclear. It would be expected that the structural changes imparted by the glycans would change their functions. For example, an earlier report discussed the blunted effectiveness of glycated insulin compared to the non-modified counterpart (Nedić et al., 2012). Furthermore, another earlier report showed that insulin is a regulator of O-glycosylation in liver cells (Majumdar et al., 2006). Based on this evidence, hormone glycation and glycosylation may play roles in pancreatic disease, and thus warrants further study.

MS-based analyses of glycosylation rely on enzymatic digestion for fine structural analyses of glycan moieties. PNGase F, while capable of releasing most mammalian N-linked glycans, is also useful in that it can improve the sensitivity of O-glycans by removing the N-linked species. Some mammalian glycans remain resistant to PNGase F digestion, however (Stadlmann et al., 2018). Glycosidases, such as endo F3 and neuraminidase, are useful in studying specific glycan isomers, which have been shown to differ in pancreatic disease states. Isomeric characterization in glycoproteomics is a constantly evolving area of work and has been thoroughly reviewed recently (Gutierrez Reyes et al., 2021; Peng et al., 2021). Due to the heterogeneity of O-glycan core structures, a universal enzyme has not been discovered, though as shown by Malaker et al. (2019), enzyme development for use in MS-based studies is a continued area of work. Just as lectin proteins widely used for enriching glycan motifs were discovered and purified from natural, plant-based sources, O-glycan targeting enzymes may be found in bacterial sources. Further biological research may then lead to discoveries of new enzymes that may be beneficial for MS-based structural analyses of diverse glycans.

CONCLUSION

In sum, recent literature on MS-based glycomic and glycoproteomic analyses have provided insights into the pathologies of diabetes, pancreatitis, and pancreatic cancer.

Glycation and glycosylation are two important biological processes that impart vast structural heterogeneity onto proteins that inevitably change their functions, potentially causing or worsening disease states. In MS-based analyses, large, glycosylated analytes are often broken down into smaller pieces that are more amenable to currently available mass analyzers, such as time-of-flight or the Orbitrap. These same mass analyzers can be used to accurately measure mass-to-charge with high resolution to enable molecular identifications from analyte fragments. This strategy is one borne of instrumental constraints, however, and inevitably leads to loss of information due to analyte modification.

Top-down MS, as opposed to the common bottom-up approach taken by many studies in the literature reviewed here, is a field that embraces intact characterization of analytes. Top-down analyses can be performed to resolve glycoforms of large, intact glycoproteins (and

even antibodies) while minimizing loss of information. Indeed, top-down MS is a growing field that will be key in future health research, though technical challenges remain before top-down analyses outnumber those of bottom-up (Brown et al., 2020; Melby et al., 2021). Top-down MS, ultimately, is needed to maximize the accuracy of structural characterization of glycoproteins, which are molecules that can be modified with multiple glycan structures at multiple sites simultaneously. This phenomenon has been coined "meta-heterogeneity" as discussed in a recent review (Caval et al., 2020). Glycoform quantification in disease states is a burgeoning area of research. The field has also been moving towards the frontier of "structure-focused glycoproteomics" as discussed in another recent review to analyze functional changes as imparted by structural modifications (Chernykh et al., 2021).

Advances in instrumentation will further enhance analyte detection capabilities. Thus, the information lost in the transfer between molecules *in vivo* and as detected is minimized. Though current analytical capabilities lead to inevitable loss of information, new insights into pancreatic disease have been derived using MS-based glycomic and glycoproteomic analyses. Diabetes and pancreatic cancer are two diseases that have a wide reach far beyond their organ of origin. Ongoing research efforts will and must continue to discover new biomarkers for earlier diagnoses that may assuage the damages caused by the disease and to unravel the mechanisms by which these diseases progress.

AUTHOR CONTRIBUTIONS

DT and MF reviewed the literature. DT wrote the original draft; DT, MF, and LL reviewed and edited the final draft; LL acquired funding and supervised the work.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.707387/full#supplementary-material

REFERENCES

Abrahams, J. L., Taherzadeh, G., Jarvas, G., Guttman, A., Zhou, Y., and Campbell, M. P. (2020). Recent Advances in Glycoinformatic Platforms for Glycomics and Glycoproteomics. *Curr. Opin. Struct. Biol.* 62, 56–69. doi:10.1016/j.sbi.2019.11.009

Ahn, J.-M., Kim, B.-G., Yu, M.-H., Lee, I.-K., and Cho, J.-Y. (2010). Identification of Diabetic Nephropathy-Selective Proteins in Human Plasma by Multi-Lectin Affinity Chromatography and LC-MS/MS. *Prot. Clin. Appl.* 4, 644–653. doi:10.1002/prca.200900196

Akella, N. M., Ciraku, L., and Reginato, M. J. (2019). Fueling the Fire: Emerging Role of the Hexosamine Biosynthetic Pathway in Cancer. *BMC Biol.* 17, 52. doi:10.1186/s12915-019-0671-3

Akimoto, Y., Nouso, K., Kato, H., Miyahara, K., Dohi, C., Morimoto, Y., et al. (2015). Serum N-Glycan Profiles in Patients with Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Pancreatology* 15, 432–438. doi:10.1016/j.pan.2015.05.470

Almaraz, R. T., Tian, Y., Bhattacharya, R., Tan, E., Chen, S.-H., Dallas, M. R., et al. (2012). Metabolic Flux Increases Glycoprotein Sialylation: Implications for Cell Adhesion and Cancer Metastasis. *Mol. Cell Proteomics* 11, 1–12. doi:10.1074/mcp.M112.017558

Amano, M., Eriksson, H., Manning, J. C., Detjen, K. M., André, S., Nishimura, S.-I., et al. (2012). Tumour Suppressor p16INK4a- Anoikis-Favouring Decrease in N/O-glycan/cell Surface Sialylation by Down-Regulation of Enzymes in Sialic Acid Biosynthesis in Tandem in a Pancreatic Carcinoma Model. *FEBS J.* 279, 4062–4080. doi:10.1111/febs.12001

Ansari, N. A., Chaudhary, D. K., and Dash, D. (2018). Modification of Histone by Glyoxal: Recognition of Glycated Histone Containing Advanced Glycation Adducts by Serum Antibodies of Type 1 Diabetes Patients. *Glycobiology* 28, 207–213. doi:10.1093/glycob/cwy006

Balmaña, M., Giménez, E., Puerta, A., Llop, E., Figueras, J., Fort, E., et al. (2016). Increased α1-3 Fucosylation of α-1-acid Glycoprotein (AGP) in Pancreatic Cancer. *J. Proteomics* 132, 144–154. doi:10.1016/j.jprot.2015.11.006

Balog, J., Sasi-Szabó, L., Kinross, J., Lewis, M. R., Muirhead, L. J., Veselkov, K., et al. (2013). Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Sci. Translational Med.* 5, 194ra93, 2013 . 194ra93. doi:10.1126/scitranslmed.3005623

Barnett, D., Liu, Y., Partyka, K., Huang, Y., Tang, H., Hostetter, G., et al. (2017). The CA19-9 and Sialyl-TRA Antigens Define Separate Subpopulations of Pancreatic Cancer Cells. *Sci. Rep.* 7, 4020. doi:10.1038/s41598-017-04164-z

Belczacka, I., Pejchinovski, M., Krochmal, M., Magalhães, P., Frantzi, M., Mullen, W., et al. (2019). Urinary Glycopeptide Analysis for the Investigation of Novel Biomarkers. *Prot. Clin. Appl.* 13, 1800111, 2019 . e1800111. doi:10.1002/prca.201800111

Bermingham, M. L., Colombo, M., Mcgurnaghan, S. J., Blackbourn, L. A. K., Vučković, F., Pučić Baković, M., et al. (2018). N-glycan Profile and Kidney Disease in Type 1 Diabetes. *Dia Care* 41, 79–87. doi:10.2337/dc17-1042

Bishop, J. R., Foley, E., Lawrence, R., and Esko, J. D. (2010). Insulin-dependent Diabetes Mellitus in Mice Does Not Alter Liver Heparan Sulfate. *J. Biol. Chem.* 285, 14658–14662. doi:10.1074/jbc.M110.112391

Blache, D., Bourdon, E., Salloignon, P., Lucchi, G., Ducoroy, P., Petit, J.-M., et al. (2015). Glycated Albumin with Loss of Fatty Acid Binding Capacity Contributes to Enhanced Arachidonate Oxygenation and Platelet Hyperactivity: Relevance in Patients with Type 2 Diabetes. *Diabetes* 64, 960–972. doi:10.2337/db14-0879

Borges, C. R., Oran, P. E., Buddi, S., Jarvis, J. W., Schaab, M. R., Rehder, D. S., et al. (2011). Building Multidimensional Biomarker Views of Type 2 Diabetes on the Basis of Protein Microheterogeneity. *Clin. Chem.* 57, 719–728. doi:10.1373/clinchem.2010.156976

Brown, K. A., Melby, J. A., Roberts, D. S., and Ge, Y. (2020). Top-down Proteomics: Challenges, Innovations, and Applications in Basic and Clinical Research. *Expert Rev. Proteomics* 17, 719–733. doi:10.1080/14789450.2020.1855982

Brownlee, M. (1995). Advanced Protein Glycosylation in Diabetes and Aging. *Annu. Rev. Med.* 46, 223–234. doi:10.1146/annurev.med.46.1.223

Buchberger, A. R., Delaney, K., Johnson, J., and Li, L. (2018). Mass Spectrometry Imaging: a Review of Emerging Advancements and Future Insights. *Anal. Chem.* 90, 240–265. doi:10.1021/acs.analchem.7b04733

Bullard, K. M., Cowie, C. C., Lessem, S. E., Saydah, S. H., Menke, A., Geiss, L. S., et al. (2018). Prevalence of Diagnosed Diabetes in Adults by Diabetes Type - United States, 2016. *MMWR Morb. Mortal. Wkly. Rep.* 67, 359–361. doi:10.15585/mmwr.mm6712a2

Cao, Z., Maupin, K., Curnutte, B., Fallon, B., Feasley, C. L., Brouhard, E., et al. (2013). Specific Glycoforms of MUC5AC and Endorepellin Accurately Distinguish Mucinous from Nonmucinous Pancreatic Cysts. *Mol. Cell Proteomics* 12, 2724–2734. doi:10.1074/mcp.M113.030700

Caprioli, R. M., Farmer, T. B., and Gile, J. (1997). Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* 69, 4751–4760. doi:10.1021/ac970888i

Caval, T., Heck, A. J. R., and Reiding, K. R. (2020). Meta-heterogeneity: Evaluating and Describing the Diversity in Glycosylation between Sites on the Same Glycoprotein. *Mol. Cel. Proteomics* 20, 100010. doi:10.1074/mcp.R120.002093

Cech, N. B., and Enke, C. G. (2000). Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides. *Anal. Chem.* 72, 2717–2723. doi:10.1021/ac9914869

Chatterjee, B., and Thakur, S. S. (2018). Investigation of post-translational Modifications in Type 2 Diabetes. *Clin. Proteom* 15, 32. doi:10.1186/s12014-018-9208-y

Chen, B., Brown, K. A., Lin, Z., and Ge, Y. (2018a). Top-Down Proteomics: Ready for Prime Time?. *Anal. Chem.* 90, 110–127. doi:10.1021/acs.analchem.7b04747

Chen, G., Li, H., Qiu, L., Qin, X., Liu, H., and Li, Z. (2014). Change of Fucosylated IgG2 Fc-Glycoforms in Pancreatitis and Pancreatic Adenocarcinoma: a Promising Disease-Classification Model. *Anal. Bioanal. Chem.* 406, 267–273. doi:10.1007/s00216-013-7439-3

Chen, Z., Yu, Q., Hao, L., Liu, F., Johnson, J., Tian, Z., et al. (2018b). Site-specific Characterization and Quantitation of N-Glycopeptides in PKM2 Knockout Breast Cancer Cells Using DiLeu Isobaric Tags Enabled by Electron-Transfer/higher-Energy Collision Dissociation (EThcD). *Analyst* 143, 2508–2519. doi:10.1039/c8an00216a

Chen, Z., Yu, Q., Yu, Q., Johnson, J., Shipman, R., Zhong, X., et al. (2021). In-depth Site-specific Analysis of N-Glycoproteome in Human Cerebrospinal Fluid and Glycosylation Landscape Changes in Alzheimer's Disease. *Mol. Cell Proteomics* 20, 100081. doi:10.1016/j.mcpro.2021.100081

Chernykh, A., Kawahara, R., and Thaysen-Andersen, M. (2021). Towards Structure-Focused Glycoproteomics. *Biochem. Soc. Trans.* 49, 161–186. doi:10.1042/bst20200222

Cho, J., Scragg, R., and Petrov, M. S. (2020). Postpancreatitis Diabetes Confers Higher Risk for Pancreatic Cancer Than Type 2 Diabetes: Results from a Nationwide Cancer Registry. *Dia Care* 43, 2106–2112. doi:10.2337/dc20-0207

Cho, K.-C., Chen, L., Hu, Y., Schnaubelt, M., and Zhang, H. (2019). Developing Workflow for Simultaneous Analyses of Phosphopeptides and Glycopeptides. *ACS Chem. Biol.* 14, 58–66. doi:10.1021/acschembio.8b00902

Chou, C. Y., Chang, C. T., and Chen, C. J. (2020). Analytically Validated Protein Biomarkers of Chronic Pancreatitis and Pancreatic Cancer for Potential Clinical Diagnosis with Mass Spectrometry. *Rapid Commun. Mass. Spectrom.* 34 (Suppl. 1), e8580. doi:10.1002/rcm.8580

Chugh, S., Barkeer, S., Rachagani, S., Nimmakayala, R. K., Perumal, N., Pothuraju, R., et al. (2018). Disruption of C1galt1 Gene Promotes Development and Metastasis of Pancreatic Adenocarcinomas in Mice. *Gastroenterology* 155, 1608–1624. doi:10.1053/j.gastro.2018.08.007

Clift, C. L., Drake, R. R., Mehta, A., and Angel, P. M. (2021). Multiplexed Imaging Mass Spectrometry of the Extracellular Matrix Using Serial Enzyme Digests from Formalin-Fixed Paraffin-Embedded Tissue Sections. *Anal. Bioanal. Chem.* 413, 2709–2719. doi:10.1007/s00216-020-03047-z

Cui, Y., Tabang, D. N., Zhang, Z., Ma, M., Alpert, A. J., and Li, L. (2021). Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Anal. Chem.* 93, 7908–7916. doi:10.1021/acs.analchem.1c00615

Cui, Y., Yang, K., Tabang, D. N., Huang, J., Tang, W., and Li, L. (2019). Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *J. Am. Soc. Mass. Spectrom.* 30, 2491–2501. doi:10.1007/s13361-019-02230-6

D'aronco, S., Crotti, S., Agostini, M., Traldi, P., Chilelli, N. C., and Lapolla, A. (2019). The Role of Mass Spectrometry in Studies of Glycation Processes and Diabetes Management. *Mass. Spec. Rev.* 38, 112–146. doi:10.1002/mas.21576

Delafield, D. G., and Li, L. (2021). Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. *Mol. Cell Proteomics* 20, 100054. doi:10.1074/mcp.R120.002095

Dotz, V., Lemmers, R. F. H., Reiding, K. R., Hipgrave Ederveen, A. L., Lieverse, A. G., Mulder, M. T., et al. (2018). Plasma Protein N-Glycan Signatures of Type 2 Diabetes. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1862, 2613–2622. doi:10.1016/j.bbagen.2018.08.005

Drake, R. R. (2015). Glycosylation and Cancer: Moving Glycomics to the Forefront. *Adv. Cancer Res.* 126, 1–10. doi:10.1016/bs.acr.2014.12.002

Dupont, F. O., Hivert, M.-F., Allard, C., Ménard, J., Perron, P., Bouchard, L., et al. (2014). Glycation of Fetal Hemoglobin Reflects Hyperglycemia Exposure In Utero: Table 1. *Dia Care* 37, 2830–2833. doi:10.2337/dc14-0549

Engle, D. D., Tiriac, H., Rivera, K. D., Pommier, A., Whalen, S., Oni, T. E., et al. (2019). The Glycan CA19-9 Promotes Pancreatitis and Pancreatic Cancer in Mice. *Science* 364, 1156–1162. doi:10.1126/science.aaw3145

Feng, Y., Chen, B., Yu, Q., Zhong, X., Frost, D. C., Ikonomidou, C., et al. (2019a). Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound (SUGAR) Tags: A Probe for Quantitative Glycomic Analysis. *Anal. Chem.* 91, 3141–3146. doi:10.1021/acs.analchem.8b05757

Feng, Y., Li, M., Lin, Y., Chen, B., and Li, L. (2019b). Multiplex Quantitative Glycomics Enabled by Periodate Oxidation and Triplex Mass Defect Isobaric Multiplex Reagents for Carbonyl-Containing Compound Tags. *Anal. Chem.* 91, 11932–11937. doi:10.1021/acs.analchem.9b02736

Freedman, L. P., Gibson, M. C., Ethier, S. P., Soule, H. R., Neve, R. M., and Reid, Y. A. (2015). Reproducibility: Changing the Policies and Culture of Cell Line Authentication. *Nat. Methods* 12, 493–497. doi:10.1038/nmeth.3403

Giménez, E., Balmaña, M., Figueras, J., Fort, E., Bolós, c. d., Sanz-Nebot, V., et al. (2015). Quantitative Analysis of N-Glycans from Human Alfa-Acid-Glycoprotein Using Stable Isotope Labeling and Zwitterionic Hydrophilic Interaction Capillary Liquid Chromatography Electrospray Mass Spectrometry as Tool for Pancreatic Disease Diagnosis. *Analytica Chim. Acta* 866, 59–68. doi:10.1016/j.aca.2015.02.008

Glover, M. S., Yu, Q., Chen, Z., Shi, X., Kent, K. C., and Li, L. (2018). Characterization of Intact Sialylated Glycopeptides and Phosphorylated Glycopeptides from IMAC Enriched Samples by EThcD Fragmentation: Toward Combining Phosphoproteomics and Glycoproteomics. *Int. J. Mass Spectrom.* 427, 35–42. doi:10.1016/j.ijms.2017.09.002

Guerrero, P. E., Duran, A., Ortiz, M. R., Castro, E., Garcia-Velasco, A., Llop, E., et al. (2021). Microfibril Associated Protein 4 (MFAP4) Is a Carrier of the Tumor Associated Carbohydrate Sialyl-Lewis X (sLex) in Pancreatic Adenocarcinoma. *J. Proteomics* 231, 104004. doi:10.1016/j.jprot.2020.104004

Guo, Z., Liu, X., Li, M., Shao, C., Tao, J., Sun, W., et al. (2015). Differential Urinary Glycoproteome Analysis of Type 2 Diabetic Nephropathy Using 2D-LC-MS/MS and iTRAQ Quantification. *J. Transl. Med.* 13, 371. doi:10.1186/s12967-015-0712-9

Gutierrez Reyes, C. D., Jiang, P., Donohoo, K., Atashi, M., and Mechref, Y. S. (2021). Glycomics and Glycoproteomics: Approaches to Address Isomeric Separation of Glycans and Glycopeptides. *J. Sep. Sci.* 44, 403–425. doi:10.1002/jssc.202000878

Hanna-Sawires, R. G., Schiphuis, J. H., Wuhrer, M., Vasen, H. F. A., Van Leerdam, M. E., Bonsing, B. A., et al. (2021). Clinical Perspective on Proteomic and Glycomic Biomarkers for Diagnosis, Prognosis, and Prediction of Pancreatic Cancer. *Ijms* 22, 2655. doi:10.3390/ijms22052655

Harkin, C., Smith, K. W., Cruickshank, F. L., Logan Mackay, C., Flinders, B., Heeren, R. M. A., et al. (2021). On-tissue Chemical Derivatization in Mass Spectrometry Imaging. *Mass. Spec. Rev.*, 1–33. doi:10.1002/mas.21680

Hasehira, K., Furuta, T., Shimomura, O., Asada, M., Oda, T., and Tateno, H. (2021). Quantitative Structural Analysis of Glycans Expressed within Tumors Derived from Pancreatic Cancer Patient-Derived Xenograft Mouse Models. *Biochem. Biophysical Res. Commun.* 534, 310–316. doi:10.1016/j.bbrc.2020.11.087

Heijs, B., Potthoff, A., Soltwisch, J., and Dreisewerd, K. (2020). MALDI-2 for the Enhanced Analysis of N-Linked Glycans by Mass Spectrometry Imaging. *Anal. Chem.* 92, 13904–13911. doi:10.1021/acs.analchem.0c02732

Higai, K., Azuma, Y., Aoki, Y., and Matsumoto, K. (2003). Altered Glycosylation of α1-acid Glycoprotein in Patients with Inflammation and Diabetes Mellitus. *Clinica Chim. Acta* 329, 117–125. doi:10.1016/s0009-8981(02)00427-8

Hofmann, B. T., Schlüter, L., Lange, P., Mercanoglu, B., Ewald, F., Fölster, A., et al. (2015). COSMC Knockdown Mediated Aberrant O-Glycosylation Promotes Oncogenic Properties in Pancreatic Cancer. *Mol. Cancer* 14, 109. doi:10.1186/s12943-015-0386-1

Holst, S., Belo, A. I., Giovannetti, E., Van Die, I., and Wuhrer, M. (2017). Profiling of Different Pancreatic Cancer Cells Used as Models for Metastatic Behaviour Shows Large Variation in Their N-Glycosylation. *Sci. Rep.* 7, 16623. doi:10.1038/s41598-017-16811-6

Huang, J., Liu, X., Wang, D., Cui, Y., Shi, X., Dong, J., et al. (2021). Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Anal. Chem.* 93, 8568–8576. doi:10.1021/acs.analchem.1c01324

Kailemia, M. J., Park, D., and Lebrilla, C. B. (2017). Glycans and Glycoproteins as Specific Biomarkers for Cancer. *Anal. Bioanal. Chem.* 409, 395–410. doi:10.1007/s00216-016-9880-6

Kang, T., Boland, B. B., Jensen, P., Alarcon, C., Nawrocki, A., Grimsby, J. S., et al. (2020). Characterization of Signaling Pathways Associated with Pancreatic β-cell Adaptive Flexibility in Compensation of Obesity-Linked Diabetes in Db/db Mice. *Mol. Cell Proteomics* 19, 971–993. doi:10.1074/mcp.RA119.001882

Keser, T., Tijardović, M., Gornik, I., Lukić, E., Lauc, G., Gornik, O., et al. (2021). High-throughput and Site-specific N-Glycosylation Analysis of Human Alpha-1-Acid Glycoprotein Offers a Great Potential for New Biomarker Discovery. *Mol. Cell Proteomics* 20, 100044. doi:10.1074/mcp.RA120.002433

Kleeff, J., Korc, M., Apte, M., La Vecchia, C., Johnson, C. D., Biankin, A. V., et al. (2016). Pancreatic Cancer. *Nat. Rev. Dis. Primers* 2, 16022. doi:10.1038/nrdp.2016.22

Kontro, H., Joenväärä, S., Haglund, C., and Renkonen, R. (2014). Comparison of Sialylated N -glycopeptide Levels in Serum of Pancreatic Cancer Patients, Acute Pancreatitis Patients, and Healthy Controls. *Proteomics* 14, 1713–1723. doi:10.1002/pmic.201300270

Korwar, A. M., Vannuruswamy, G., Jagadeeshaprasad, M. G., Jayaramaiah, R. H., Bhat, S., Regin, B. S., et al. (2015). Development of Diagnostic Fragment Ion Library for Glycated Peptides of Human Serum Albumin: Targeted Quantification in Prediabetic, Diabetic, and Microalbuminuria Plasma by Parallel Reaction Monitoring, SWATH, and MSE. *Mol. Cell Proteomics* 14, 2150–2159. doi:10.1074/mcp.m115.050518

Korwar, A. M., and Zhang, Q. (2021). Comprehensive Quantification of Carboxymethyllysine-Modified Peptides in Human Plasma. *J. Am. Soc. Mass. Spectrom.* 32, 744–752. doi:10.1021/jasms.0c00443

Kozak, R. P., Royle, L., Gardner, R. A., Fernandes, D. L., and Wuhrer, M. (2012). Suppression of Peeling during the Release of O-Glycans by Hydrazinolysis. *Anal. Biochem.* 423, 119–128. doi:10.1016/j.ab.2012.01.002

Krishnan, S., Whitwell, H., Cuenco, J., Gentry-Maharaj, A., Menon, U., Pereira, S., et al. (2017). Evidence of Altered Glycosylation of Serum Proteins Prior to Pancreatic Cancer Diagnosis. *Ijms* 18, 2670. doi:10.3390/ijms18122670

Lapolla, A., Fedele, D., and Traldi, P. (2000). The Role of Mass Spectrometry in the Study of Non-enzymatic Protein Glycation in Diabetes. *Mass. Spectrom. Rev.* 19, 279–304. doi:10.1002/1098-2787(2000)19:5<279::Aid-mas3>3.0.Co;2-g

Lapolla, A., Molin, L., and Traldi, P. (2013). Protein Glycation in Diabetes as Determined by Mass Spectrometry. *Int. J. Endocrinol.* 2013, 1–11. doi:10.1155/2013/412103

Lee, C.-L., Chiu, P. C. N., Pang, P.-C., Chu, I. K., Lee, K.-F., Koistinen, R., et al. (2011). Glycosylation Failure Extends to Glycoproteins in Gestational Diabetes Mellitus. *Diabetes* 60, 909–917. doi:10.2337/db10-1186

Ligorio, M., Sil, S., Malagon-Lopez, J., Nieman, L. T., Misale, S., Di Pilato, M., et al. (2019). Stromal Microenvironment Shapes the Intratumoral Architecture of Pancreatic Cancer. *Cell* 178, 160–175. e127. doi:10.1016/j.cell.2019.05.012

Liljedahl, L., Pedersen, M. H., Norlin, J., Mcguire, J. N., and James, P. (2016). N-glycosylation Proteome Enrichment Analysis in Kidney Reveals Differences between Diabetic Mouse Models. *Clin. Proteom* 13, 22. doi:10.1186/s12014-016-9123-z

Lim, J.-M., Wollaston-Hayden, E. E., Teo, C., Hausman, D., and Wells, L. (2014). Quantitative Secretome and Glycome of Primary Human Adipocytes during Insulin Resistance. *Clin. Proteomics* 11, 20. doi:10.1186/1559-0275-11-20

Lin, Z., Yin, H., Lo, A., Ruffin, M. T., Anderson, M. A., Simeone, D. M., et al. (2013). Label-free Relative Quantification of Alpha-2-Macroglobulin Site-specific Core-Fucosylation in Pancreatic Cancer by LC-MS/MS. *Electrophoresis* 35, 2108–2115. doi:10.1002/elps.201300376

Liot, S., Balas, J., Aubert, A., Prigent, L., Mercier-Guy, P., Verrier, B., et al. (2021). Stroma Involvement in Pancreatic Ductal Adenocarcinoma: An Overview Focusing on Extracellular Matrix Proteins. *Front. Immunol.* 12, 612271. doi:10.3389/fimmu.2021.612271

Liu, J., Dolikun, M., Štambuk, J., Trbojević-Akmačić, I., Zhang, J., Zhang, J., et al. (2019). Glycomics for Type 2 Diabetes Biomarker Discovery: Promise of Immunoglobulin G Subclass-specific Fragment Crystallizable N-Glycosylation in the Uyghur Population. *OMICS: A J. Integr. Biol.* 23, 640–648. doi:10.1089/omi.2019.0052

Liu, Y., Wang, C., Wang, R., Wu, Y., Zhang, L., Liu, B.-F., et al. (2018). Isomer-specific Profiling of N-Glycans Derived from Human Serum for Potential Biomarker Discovery in Pancreatic Cancer. *J. Proteomics* 181, 160–169. doi:10.1016/j.jprot.2018.04.016

Lu, H., Xiao, K., and Tian, Z. (2021). Benchmark of Site- and Structure-specific Quantitative Tissue N-Glycoproteomics for Discovery of Potential N-Glycoprotein Markers: a Case Study of Pancreatic Cancer. *Glycoconj. J.* 38, 213–231. doi:10.1007/s10719-021-09994-8

Ma, J., and Hart, G. W. (2013). ProteinO-GlcNAcylation in Diabetes and Diabetic Complications. *Expert Rev. Proteomics* 10, 365–380. doi:10.1586/14789450.2013.820536

Majumdar, G., Harrington, A., Hungerford, J., Martinez-Hernandez, A., Gerling, I. C., Raghow, R., et al. (2006). Insulin Dynamically Regulates Calmodulin Gene Expression by Sequential O-Glycosylation and Phosphorylation of Sp1 and its Subcellular Compartmentalization in Liver Cells. *J. Biol. Chem.* 281, 3642–3650. doi:10.1074/jbc.M511223200

Malaker, S. A., Pedram, K., Ferracane, M. J., Bensing, B. A., Krishnan, V., Pett, C., et al. (2019). The Mucin-Selective Protease StcE Enables Molecular and Functional Analysis of Human Cancer-Associated Mucins. *Proc. Natl. Acad. Sci. USA* 116, 7278–7287. doi:10.1073/pnas.1813020116

Mancera-Arteu, M., Giménez, E., Balmaña, M., Barrabés, S., Albiol-Quer, M., Fort, E., et al. (2019). Multivariate Data Analysis for the Detection of Human Alpha-Acid Glycoprotein Aberrant Glycosylation in Pancreatic Ductal Adenocarcinoma. *J. Proteomics* 195, 76–87. doi:10.1016/j.jprot.2019.01.006

Mann, B. F., Goetz, J. A., House, M. G., Schmidt, C. M., and Novotny, M. V. (2012). Glycomic and Proteomic Profiling of Pancreatic Cyst Fluids Identifies Hyperfucosylated Lactosamines on the N-Linked Glycans of Overexpressed Glycoproteins. *Mol. Cell Proteomics* 11, 1–11. 015792. doi:10.1074/mcp.M111.015792

Matsuda, A., Kuno, A., Yoshida, M., Wagatsuma, T., Sato, T., Miyagishi, M., et al. (2020). Comparative Glycomic Analysis of Exosome Subpopulations Derived from Pancreatic Cancer Cell Lines. *J. Proteome Res.* 19, 2516–2524. doi:10.1021/acs.jproteome.0c00200

Maynard, J. C., and Chalkley, R. J. (2021). Methods for Enrichment and Assignment of N-Acetylglucosamine Modification Sites. *Mol. Cell Proteomics* 20, 100031. doi:10.1074/mcp.R120.002206

Mcdowell, C. T., Klamer, Z., Hall, J., West, C. A., Wisniewski, L., Powers, T. W., et al. (2021). Imaging Mass Spectrometry and Lectin Analysis of N-Linked Glycans in Carbohydrate Antigen-Defined Pancreatic Cancer Tissues. *Mol. Cell Proteomics* 20, 100012. doi:10.1074/mcp.RA120.002256

Melby, J. A., Roberts, D. S., Larson, E. J., Brown, K. A., Bayne, E. F., Jin, S., et al. (2021). Novel Strategies to Address the Challenges in Top-Down Proteomics. *J. Am. Soc. Mass. Spectrom.* 32, 1278–1294. doi:10.1021/jasms.1c00099

Mereiter, S., Balmaña, M., Gomes, J., Magalhães, A., and Reis, C. A. (2016). Glycomic Approaches for the Discovery of Targets in Gastrointestinal Cancer. *Front. Oncol.* 6, 55. doi:10.3389/fonc.2016.00055

Miyahara, K., Nouso, K., Morimoto, Y., Kinugasa, H., Kato, H., Yamamoto, N., et al. (2015). Prognostic Value of Altered N-Glycosylation of Circulating Glycoproteins in Patients with Unresectable Pancreatic Cancer Treated with Gemcitabine. *Pancreas* 44, 551–556. doi:10.1097/mpa.0000000000000321

Muralidharan, M., Bhat, V., Bindu, Y. S., and Mandal, A. K. (2019). Glycation Profile of Minor Abundant Erythrocyte Proteome across Varying Glycemic index in Diabetes Mellitus. *Anal. Biochem.* 573, 37–43. doi:10.1016/j.ab.2019.02.026

Nakano, M., Nakagawa, T., Ito, T., Kitada, T., Hijioka, T., Kasahara, A., et al. (2008). Site-specific Analysis of N-Glycans on Haptoglobin in Sera of Patients with Pancreatic Cancer: a Novel Approach for the Development of Tumor Markers. *Int. J. Cancer* 122, 2301–2309. doi:10.1002/ijc.23364

Nedić, O., Lagundžin, D., and Masnikosa, R. (2012). Posttranslational Modifications of the Insulin-like Growth Factor-Binding Protein 3 in Patients with Type 2 Diabetes Mellitus Assessed by Affinity Chromatography. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 904, 93–98. doi:10.1016/j.jchromb.2012.07.028

Nie, S., Lo, A., Wu, J., Zhu, J., Tan, Z., Simeone, D. M., et al. (2014). Glycoprotein Biomarker Panel for Pancreatic Cancer Discovered by Quantitative Proteomics Analysis. *J. Proteome Res.* 13, 1873–1884. doi:10.1021/pr400967x

Nigjeh, E. N., Chen, R., Allen-Tamura, Y., Brand, R. E., Brentnall, T. A., and Pan, S. (2017). Spectral Library-Based Glycopeptide Analysis-Detection of Circulating Galectin-3 Binding Protein in Pancreatic Cancer. *Prot. Clin. Appl.Appl* 11, 1700064. doi:10.1002/prca.201700064

Nouso, K., Amano, M., Ito, Y. M., Miyahara, K., Morimoto, Y., Kato, H., et al. (2013). Clinical Utility of High-Throughput Glycome Analysis in Patients with Pancreatic Cancer. *J. Gastroenterol.* 48, 1171–1179. doi:10.1007/s00535-012-0732-7

Oh, S. S., Galanter, J., Thakur, N., Pino-Yanes, M., Barcelo, N. E., White, M. J., et al. (2015). Diversity in Clinical and Biomedical Research: A Promise yet to Be Fulfilled. *Plos Med.* 12, e1001918, 2015 . e1001918. doi:10.1371/journal.pmed.1001918

Pan, S., Brentnall, T. A., and Chen, R. (2016). Glycoproteins and Glycoproteomics in Pancreatic Cancer. *Wjg* 22, 9288–9299. doi:10.3748/wjg.v22.i42.9288

Pan, S., Brentnall, T. A., and Chen, R. (2020). Proteome Alterations in Pancreatic Ductal Adenocarcinoma. *Cancer Lett.* 469, 429–436. doi:10.1016/j.canlet.2019.11.020

Pan, S., Chen, R., Tamura, Y., Crispin, D. A., Lai, L. A., May, D. H., et al. (2014). Quantitative Glycoproteomics Analysis Reveals Changes in N-Glycosylation Level Associated with Pancreatic Ductal Adenocarcinoma. *J. Proteome Res.* 13, 1293–1306. doi:10.1021/pr4010184

Park, H.-M., Hwang, M. P., Kim, Y.-W., Kim, K.-J., Jin, J. M., Kim, Y. H., et al. (2015). Mass Spectrometry-Based N-Linked Glycomic Profiling as a Means for Tracking Pancreatic Cancer Metastasis. *Carbohydr. Res.* 413, 5–11. doi:10.1016/j.carres.2015.04.019

Patwa, T. H., Zhao, J., Anderson, M. A., Simeone, D. M., and Lubman, D. M. (2006). Screening of Glycosylation Patterns in Serum Using Natural Glycoprotein Microarrays and Multi-Lectin Fluorescence Detection. *Anal. Chem.* 78, 6411–6421. doi:10.1021/ac060726z

Paziewska, A., Polkowski, M., Rubel, T., Karczmarski, J., Wiechowska-Kozlowska, A., Dabrowska, M., et al. (2018). Mass Spectrometry-Based Comprehensive Analysis of Pancreatic Cyst Fluids. *Biomed. Res. Int.* 2018, 1–12. doi:10.1155/2018/7169595

Peng, W., Gutierrez Reyes, C. D., Gautam, S., Yu, A., Cho, B. G., Goli, M., et al. (2021). MS-based Glycomics and Glycoproteomics Methods Enabling Isomeric Characterization. *Mass. Spec. Rev.*, 1–40. doi:10.1002/mas.21713

Porterfield, M., Zhao, P., Han, H., Cunningham, J., Aoki, K., Von Hoff, D. D., et al. (2014). Discrimination between Adenocarcinoma and normal Pancreatic Ductal Fluid by Proteomic and Glycomic Analysis. *J. Proteome Res.* 13, 395–407. doi:10.1021/pr400422g

Pouria, S., Corran, P. H., Smith, A. C., Smith, H. W., Hendry, B. M., Challacombe, S. J., et al. (2004). Glycoform Composition Profiling of O-Glycopeptides Derived from Human Serum IgA1 by Matrix-Assisted Laser Desorption Ionization-Time of Flight-Mass Spectrometry. *Anal. Biochem.* 330, 257–263. doi:10.1016/j.ab.2004.03.053

Prevention, C. F. D. C. A. (2020). *National Diabetes Statistics Report*. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services.2020

Qiu, H., Jin, L., Chen, J., Shi, M., Shi, F., Wang, M., et al. (2020). Comprehensive Glycomic Analysis Reveals that Human Serum Albumin Glycation Specifically Affects the Pharmacokinetics and Efficacy of Different Anticoagulant Drugs in Diabetes. *Diabetes* 69, 760–770. doi:10.2337/db19-0738

Raghav, A., Ahmad, J., and Alam, K. (2017). Nonenzymatic Glycosylation of Human Serum Albumin and its Effect on Antibodies Profile in Patients with Diabetes Mellitus. *PLoS One* 12, e0176970, 2017 . e0176970. doi:10.1371/journal.pone.0176970

Ramos-Martínez, I., Vivanco-Rojas, O., Juárez-Domínguez, B., Hernández-Zimbrón, L., Ochoa-De La Paz, L., Quiroz-Mercado, H., et al. (2021). Abnormal N-Glycosylation of Human Lens Epithelial Cells in Type-2

Diabetes May Contribute to Cataract Progression. *Opth* Vol. 15, 1365–1373. doi:10.2147/opth.S300242

Riley, N. M., Bertozzi, C. R., and Pitteri, S. J. (2021). A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-Based Glycoproteomics. *Mol. Cell Proteomics* 20, 100029. doi:10.1074/mcp.R120.002277

Rudman, N., Gornik, O., and Lauc, G. (2019). Altered N-glycosylation Profiles as Potential Biomarkers and Drug Targets in Diabetes. *FEBS Lett.* 593, 1598–1615. doi:10.1002/1873-3468.13495

Sagini, M. N., Hotz-Wagenblatt, A., and Berger, M. R. (2021). A Subgroup of Lactosyl-Sepharose Binding Proteins Requires Calcium for Affinity and Galactose for Anti-proliferation. *Chemico-Biological Interactions* 334, 109354. doi:10.1016/j.cbi.2020.109354

Sans, M., Feider, C. L., and Eberlin, L. S. (2018). Advances in Mass Spectrometry Imaging Coupled to Ion Mobility Spectrometry for Enhanced Imaging of Biological Tissues. *Curr. Opin. Chem. Biol.* 42, 138–146. doi:10.1016/j.cbpa.2017.12.005

Sarrats, A., Saldova, R., Pla, E., Fort, E., Harvey, D. J., Struwe, W. B., et al. (2010). Glycosylation of Liver Acute-phase Proteins in Pancreatic Cancer and Chronic Pancreatitis. *Prot. Clin. Appl.* 4, 432–448. doi:10.1002/prca.200900150

Sharma, A., Cox, J., Glass, J., Lee, T. J., Kodeboyina, S. K., Zhi, W., et al. (2020). Serum Glycoproteomic Alterations in Patients with Diabetic Retinopathy. *Proteomes* 8, 25. doi:10.3390/proteomes8030025

Shi, Y., Li, Z., Felder, M. A., Yu, Q., Shi, X., Peng, Y., et al. (2019). Mass Spectrometry Imaging of N-Glycans from Formalin-Fixed Paraffin-Embedded Tissue Sections Using a Novel Subatmospheric Pressure Ionization Source. *Anal. Chem.* 91, 12942–12947. doi:10.1021/acs.analchem.9b02995

Shiao, J.-Y., Chang, Y.-T., Chang, M.-C., Chen, M. X., Liu, L.-W., Wang, X.-Y., et al. (2020). Development of Efficient On-Bead Protein Elution Process Coupled to Ultra-high Performance Liquid Chromatography-Tandem Mass Spectrometry to Determine Immunoglobulin G Subclass and Glycosylation for Discovery of Bio-Signatures in Pancreatic Disease. *J. Chromatogr. A* 1621, 461039. doi:10.1016/j.chroma.2020.461039

Shih, H.-C., Chang, M.-C., Chen, C.-H., Tsai, I.-L., Wang, S.-Y., Kuo, Y.-P., et al. (2019). High Accuracy Differentiating Autoimmune Pancreatitis from Pancreatic Ductal Adenocarcinoma by Immunoglobulin G Glycosylation. *Clin. Proteom* 16, 1. doi:10.1186/s12014-018-9221-1

Shon, D. J., Kuo, A., Ferracane, M. J., and Malaker, S. A. (2021). Classification, Structural Biology, and Applications of Mucin Domain-Targeting Proteases. *Biochem. J.* 478, 1585–1603. doi:10.1042/bcj20200607

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654

Singh, S. S., Heijmans, R., Meulen, C. K. E., Lieverse, A. G., Gornik, O., Sijbrands, E. J. G., et al. (2020a). Association of the IgG N-Glycome with the Course of Kidney Function in Type 2 Diabetes. *BMJ Open Diab Res. Care* 8, e001026. doi:10.1136/bmjdrc-2019-001026

Singh, S. S., Naber, A., Dotz, V., Schoep, E., Memarian, E., Slieker, R. C., et al. (2020b). Metformin and Statin Use Associate with Plasma Protein N-Glycosylation in People with Type 2 Diabetes. *BMJ Open Diab Res. Care* 8, e001230. doi:10.1136/bmjdrc-2020-001230

Smilowitz, J. T., Totten, S. M., Huang, J., Grapov, D., Durham, H. A., Lammi-Keefe, C. J., et al. (2013). Human Milk Secretory Immunoglobulin a and Lactoferrin N-Glycans Are Altered in Women with Gestational Diabetes Mellitus. *J. Nutr.* 143, 1906–1912. doi:10.3945/jn.113.180695

Soboleva, A., Mavropulo-Stolyarenko, G., Karonova, T., Thieme, D., Hoehenwarter, W., Ihling, C., et al. (2019). Multiple Glycation Sites in Blood Plasma Proteins as an Integrated Biomarker of Type 2 Diabetes Mellitus. *Ijms* 20, 2329. doi:10.3390/ijms20092329

Soltwisch, J., Kettling, H., Vens-Cappell, S., Wiegelmann, M., Müthing, J., and Dreisewerd, K. (2015). Mass Spectrometry Imaging with Laser-Induced Postionization. *Science* 348, 211–215. doi:10.1126/science.aaa1051

Sookwong, P., Nakagawa, K., Fujita, I., Shoji, N., and Miyazawa, T. (2011). Amadori-glycated Phosphatidylethanolamine, a Potential Marker for Hyperglycemia, in Streptozotocin-Induced Diabetic Rats. *Lipids* 46, 943–952. doi:10.1007/s11745-011-3588-3

Spraggins, J. M., Djambazova, K. V., Rivera, E. S., Migas, L. G., Neumann, E. K., Fuetterer, A., et al. (2019). High-Performance Molecular Imaging with MALDI

Trapped Ion-Mobility Time-Of-Flight (timsTOF) Mass Spectrometry. *Anal. Chem.* 91, 14552–14560. doi:10.1021/acs.analchem.9b03612

Stadlmann, J., Hoi, D. M., Taubenschmid, J., Mechtler, K., and Penninger, J. M. (2018). Analysis of PNGase F-Resistant N-Glycopeptides Using SugarQb for Proteome Discoverer 2.1 Reveals Cryptic Substrate Specificities. *Proteomics* 18, 1700436, 2018 . e1700436. doi:10.1002/pmic.201700436

Tan, Z., Yin, H., Nie, S., Lin, Z., Zhu, J., Ruffin, M. T., et al. (2015). Large-scale Identification of Core-Fucosylated Glycopeptide Sites in Pancreatic Cancer Serum Using Mass Spectrometry. *J. Proteome Res.* 14, 1968–1978. doi:10.1021/acs.jproteome.5b00068

Tanaka-Okamoto, M., Hanzawa, K., Mukai, M., Takahashi, H., Ohue, M., and Miyamoto, Y. (2018). Identification of Internally Sialylated Carbohydrate Tumor Marker Candidates, Including Sda/CAD Antigens, by Focused Glycomic Analyses Utilizing the Substrate Specificity of Neuraminidase. *Glycobiology* 28, 247–260. doi:10.1093/glycob/cwy010

Tanaka-Okamoto, M., Mukai, M., Takahashi, H., Fujiwara, Y., Ohue, M., and Miyamoto, Y. (2017). Various Sulfated Carbohydrate Tumor Marker Candidates Identified by Focused Glycomic Analyses. *Glycobiology* 27, 400–415. doi:10.1093/glycob/cww133

Taniguchi, N., and Kizuka, Y. (2015). Glycans and Cancer. *Adv. Cancer Res.* 126, 11–51. doi:10.1016/bs.acr.2014.11.001

Thomas, D. R., and Scott, N. E. (2021). Glycoproteomics: Growing up Fast. *Curr. Opin. Struct. Biol.* 68, 18–25. doi:10.1016/j.sbi.2020.10.028

Tousi, F., Bones, J., Hancock, W. S., and Hincapie, M. (2013). Differential Chemical Derivatization Integrated with Chromatographic Separation for Analysis of Isomeric SialylatedN-Glycans: A Nano-Hydrophilic Interaction Liquid Chromatography-MS Platform. *Anal. Chem.* 85, 8421–8428. doi:10.1021/ac4018007

Turiák, L., Shao, C., Meng, L., Khatri, K., Leymarie, N., Wang, Q., et al. (2014). Workflow for Combined Proteomics and Glycomics Profiling from Histological Tissues. *Anal. Chem.* 86, 9670–9678. doi:10.1021/ac5022216

Van, J. A. D., Clotet-Freixas, S., Hauschild, A.-C., Batruch, I., Jurisica, I., Elia, Y., et al. (2020). Urinary Proteomics Links Keratan Sulfate Degradation and Lysosomal Enzymes to Early Type 1 Diabetes. *PLoS One* 15, e0233639, 2020 . e0233639. doi:10.1371/journal.pone.0233639

Vreeker, G. C. M., Hanna-Sawires, R. G., Mohammed, Y., Bladergroen, M. R., Nicolardi, S., Dotz, V., et al. (2020). Serum N -Glycome Analysis Reveals Pancreatic Cancer Disease Signatures. *Cancer Med.* 9, 8519–8529. doi:10.1002/cam4.3439

Wagatsuma, T., Nagai-Okatani, C., Matsuda, A., Masugi, Y., Imaoka, M., Yamazaki, K., et al. (2020). Discovery of Pancreatic Ductal Adenocarcinoma-Related Aberrant Glycosylations: A Multilateral Approach of Lectin Microarray-Based Tissue Glycomic Profiling with Public Transcriptomic Datasets. *Front. Oncol.* 10, 338. doi:10.3389/fonc.2020.00338

Wang, S.-H., Wang, T.-F., Wu, C.-H., and Chen, S.-H. (2014). In-depth Comparative Characterization of Hemoglobin Glycation in normal and Diabetic Bloods by LC-MSMS. *J. Am. Soc. Mass. Spectrom.* 25, 758–766. doi:10.1007/s13361-014-0830-2

Wang, Z., Gucek, M., and Hart, G. W. (2008). Cross-talk between GlcNAcylation and Phosphorylation: Site-specific Phosphorylation Dynamics in Response to Globally Elevated O-GlcNAc. *Proc. Natl. Acad. Sci.* 105, 13793–13798. doi:10.1073/pnas.0806216105

West, C. A., Lu, X., Grimsley, G., Norris-Caneda, K., Mehta, A. S., Angel, P. M., et al. (2021). Optimization of Multiple Glycosidase and Chemical Stabilization Strategies for N-Glycan Isomer Detection by Mass Spectrometry Imaging in Formalin-Fixed, Paraffin-Embedded Tissues. *Methods Mol. Biol.* 2271, 303–316. doi:10.1007/978-1-0716-1241-5_21

Whatcott, C. J., Posner, R. G., Von Hoff, D. D., and Han, H. (2012). "Desmoplasia and Chemoresistance in Pancreatic Cancer," in *Pancreatic Cancer and Tumor Microenvironment*. Editors P.J. Grippo and H.G. Munshi (ThiruvananthapuramIndia: Transworld Research Network).

Wilkinson, H., and Saldova, R. (2020). Current Methods for the Characterization ofO-Glycans. *J. Proteome Res.* 19, 3890–3905. doi:10.1021/acs.jproteome.0c00435

Wolters-Eisfeld, G., Mercanoglu, B., Hofmann, B. T., Wolpers, T., Schnabel, C., Harder, S., et al. (2018). Loss of Complex O-Glycosylation Impairs Exocrine Pancreatic Function and Induces MODY8-like Diabetes in Mice. *Exp. Mol. Med.* 50, 1–13. doi:10.1038/s12276-018-0157-3

Xu, S., Sun, F., Tong, M., and Wu, R. (2021). MS-based Proteomics for Comprehensive Investigation of Protein O-GlcNAcylation. *Mol. Omics* 17, 186–196. doi:10.1039/d1mo00025j

Yabu, M., Korekane, H., Takahashi, H., Ohigashi, H., Ishikawa, O., and Miyamoto, Y. (2013). Accumulation of Free Neu5Ac-Containing Complex-type N-Glycans in Human Pancreatic Cancers. *Glycoconj. J.* 30, 247–256. doi:10.1007/s10719-012-9435-9

Yang, S., Onigman, P., Wu, W. W., Sjogren, J., Nyhlen, H., Shen, R.-F., et al. (2018). Deciphering Protein O-Glycosylation: Solid-phase Chemoenzymatic Cleavage and Enrichment. *Anal. Chem.* 90, 8261–8269. doi:10.1021/acs.analchem.8b01834

Yu, Q., Canales, A., Glover, M. S., Das, R., Shi, X., Liu, Y., et al. (2017). Targeted Mass Spectrometry Approach Enabled Discovery ofO-Glycosylated Insulin and Related Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal. Chem.* 89, 9184–9191. doi:10.1021/acs.analchem.7b01926

Zhang, H., Shi, X., Vu, N. Q., Li, G., Li, Z., Shi, Y., et al. (2020a). On-Tissue Derivatization with Girard's Reagent P Enhances N-Glycan Signals for Formalin-Fixed Paraffin-Embedded Tissue Sections in MALDI Mass Spectrometry Imaging. *Anal. Chem.* 92, 13361–13368. doi:10.1021/acs.analchem.0c02704

Zhang, J., Rector, J., Lin, J. Q., Young, J. H., Sans, M., Katta, N., et al. (2017). Nondestructive Tissue Analysis for *Ex Vivo* and *In Vivo* Cancer Diagnosis Using a Handheld Mass Spectrometry System. *Sci. Transl. Med.* 9, eaan3968. doi:10.1126/scitranslmed.aan3968

Zhang, L., and Zhang, Q. (2020). Glycated Plasma Proteins as More Sensitive Markers for Glycemic Control in Type 1 Diabetes. *Prot. Clin. Appl.* 14, 1900104, 2020 . e1900104. doi:10.1002/prca.201900104

Zhang, M., Xu, W., and Deng, Y. (2013). A New Strategy for Early Diagnosis of Type 2 Diabetes by Standard-free, Label-free LC-MS/MS Quantification of Glycated Peptides. *Diabetes* 62, 3936–3942. doi:10.2337/db13-0347

Zhang, T., Van Die, I., Tefsen, B., Van Vliet, S. J., Laan, L. C., Zhang, J., et al. (2020b). Differential O- and Glycosphingolipid Glycosylation in Human Pancreatic Adenocarcinoma Cells with Opposite Morphology and Metastatic Behavior. *Front. Oncol.* 10, 732. doi:10.3389/fonc.2020.00732

Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A., and Lubman, D. M. (2006). Comparative Serum Glycoproteomics Using Lectin Selected Sialic Acid Glycoproteins with Mass Spectrometric Analysis: Application to Pancreatic Cancer Serum. *J. Proteome Res.* 5, 1792–1802. doi:10.1021/pr060034r

Zhao, Q., Jia, T., Cao, Q., Tian, F., and Ying, W. (2018). A Crude 1-DNJ Extract from Home Made Bombyx Batryticatus Inhibits Diabetic Cardiomyopathy-Associated Fibrosis in Db/db Mice and Reduces Protein N-Glycosylation Levels. *Ijms* 19, 1699. doi:10.3390/ijms19061699

Zhou, Q., Fülöp, A., and Hopf, C. (2021). Recent Developments of Novel Matrices and On-Tissue Chemical Derivatization Reagents for MALDI-MSI. *Anal. Bioanal. Chem.* 413, 2599–2617. doi:10.1007/s00216-020-03023-7

Zhou, Y., Lih, T.-S. M., Yang, G., Chen, S.-Y., Chen, L., Chan, D. W., et al. (2020). An Integrated Workflow for Global, Glyco-, and Phospho-Proteomic Analysis of Tumor Tissues. *Anal. Chem.* 92, 1842–1849. doi:10.1021/acs.analchem.9b03753

gQuant, an Automated Tool for Quantitative Glycomic Data Analysis

Jiangming Huang[1,2], Biyun Jiang[1], Mingqi Liu[1], Pengyuan Yang[1,2] and Weiqian Cao[1,3]*

[1]The Fifth People's Hospital, Fudan University, and the Shanghai Key Laboratory of Medical Epigenetics, The International Co-laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, Shanghai, China, [2]Department of Chemistry, Fudan University, Shanghai, China, [3]NHC Key Laboratory of Glycoconjugates Research, Fudan University, Shanghai, China

MALDI-MS-based glycan isotope labeling methods have been effectively and widely used for quantitative glycomics. However, interpretation of the data produced by MALDI-MS is inaccurate and tedious because the bioinformatic tools are inadequate. In this work, we present gQuant, an automated tool for MALDI-MS-based glycan isotope labeling data processing. gQuant was designed with a set of dedicated algorithms to improve the efficiency, accuracy and convenience of quantitation data processing. When tested on the reference data set, gQuant showed a fast processing speed, as it was able to search the glycan data of model glycoproteins in a few minutes and reported more results than the manual analysis did. The reported quantitation ratios matched well with the experimental glycan mixture ratios ranging from 1:10 to 10:1. In addition, gQuant is fully open-source and is coded in Python, which is supported by most operating systems, and it has a user-friendly interface. gQuant can be easily adapted by users for specific experimental designs, such as specific glycan databases, different derivatization types and relative quantitation designs and can thus facilitate fast glycomic quantitation for clinical sample analysis using MALDI-MS-based stable isotope labeling.

Keywords: glycomic quantitation, stable isotope labeling, MALDI-MS analysis, quantitative tool, automated processing

INTRODUCTION

Protein glycosylation plays significant roles in many biological and physiological processes, including cell adhesion, sperm fusion, and protein folding, as well as in protein half-life (Hart and Copeland, 2010; Xu and Ng, 2015). It has also been reported that aberrant glycosylation has a substantial impact on host-cell reactions, infectious diseases such as tuberculosis, the progression of tumors (Lu et al., 2016), autoimmune diseases (Eakin et al., 2016) and prostate cancer (Shah et al., 2015). Therefore, precision quantitation of glycans is greatly needed to better understand glycan functionality, measure different glycan levels and potentially discover glycan biomarkers.

Many efforts have been made to develop mass spectrometry (MS)-based glycan quantitation techniques due to the excellent qualitative ability, sensitivity and high throughput of mass spectrometry (Wuhrer, 2013; Cao et al., 2020). For example, a metabolic labeling strategy was introduced for *in vivo* labeling and glycan quantitation (Orlando et al., 2009); chemical labeling strategies using different isotope labeling reagents, such as 2-aminobenzoic acid, aniline, arginine, and 1-phenyl-3-methyl-5-pyrazolone, were developed for glycan derivatization and quantitation (Prien et al., 2010; Albrecht et al., 2017; Cai et al., 2015; Smith et al., 2017; Wang et al., 2017); and enzymatic ^{18}O labeling was also utilized for glycan relative quantification (Zhang et al., 2015; Cao

et al., 2015). Among the developed techniques, MALDI-MS-based methods have shown high feasibility, efficiency and speed in quantitative glycan analysis and have been widely used. However, MALDI-MS-based glycan quantitation data have mostly been processed manually. This work is tedious and requires expert knowledge of protein glycosylation. Thus, this drawback greatly impedes the development of quantitative glycomics and the understanding of protein glycosylation.

Some studies have been carried out to develop bioinformatic tools to assist glycan MS data interpretation and quantitation. GlycoWorkbench is one of the most popular glycomic analytical tools that can search glycans with given m/z(s). However, it lacks quantitation modules (Ceroni et al., 2008). GlycoReSoft was developed to help annotate and quantitate glycans, but it can only be applied for LC-ESI-MS data (Maxwell et al., 2012). Massytools was mainly designed to quantitate the abundance of glycans from the total sample or monoclonal antibody and report normalized percentage results (Jansen et al., 2015), while LaCytools was designed for LC-ESI-MS-based targeted glycan quantitation (Jansen et al., 2016). Multiglycan was introduced to obtain glycan mass spectra from MALDI-MS or LC-ESI-MS. Multiglycan detects different combination modes of glycans and can be used for glycan quantitation (Yu et al., 2013a; Hu et al., 2015). Although these tools were designed for glycan searches and to assist quantitation (Yu et al., 2013b), the glycan quantitation functions remain underdeveloped, especially for stable isotope labeling-based glycan relative quantitation data. Therefore, a precision and user-friendly tool for glycan isotope labeling-based MS data analysis is still needed.

In this work, we developed a MALDI-MS-based relative quantitation glycan data processing tool, gQuant. gQuant is embedded with well-defined glycan databases and a set of thoroughly designed algorithms to facilitate automated MALDI-MS-based glycan data preprocessing, glycan identification and quantitation ratio calculations. Glycan isotope patterns were studied to aid the precise quantitation of biological samples. Moreover, gQuant was developed using the Python 2.7 and Python 3 programming language and can be used in most operating systems. gQuant is a fully open-source software tool that is fast and has a user-friendly interface to assist glycan MS data processing.

MATERIALS AND METHODS

The development of the gQuant tool for automated glycan quantitation data processing includes two sections: the preparation of a comprehensive glycan database and establishment of the quantitation tool. A set of MS data was further analyzed to test the feasibility of the newly developed glycan tool. Detailed construction and testing of gQuant are listed below.

Glycan Database Preparation

The inherent glycan database in gQuant was constructed by carefully collecting and integrating information from published data and literatures (Ranzinger et al., 2009; Balog

et al., 2012; and; Liu et al., 2014). As reported, glycans from human sources rarely contain N-glycolylneuraminic acid (NeuGc). Therefore, two glycan databases, human-sourced and mammalian-sourced, were developed and contained 344 and 419 nonredundant glycans, respectively. The largest difference between the two databases is that the mammalian glycan database is composed of both N-acetylneuraminic acid (NeuAc) and N-glycolylneuraminic acid (NeuGc), while sialyated glycans in the human glycan database is composed of only N-acetylneuraminic acid (NeuAc). Element compositions, glycans and corresponding molecular weights (both residual and free-ended) were uniform, calculated and well recorded in the database file. In addition, distinct glycan isotope distributions were generated and recorded *via* the online tool MS-ISOTOPE (https://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msisotope). The isotope distributions were further applied for isotope interference calculations to increase quantitation accuracy. It should be noted that such a glycan database can be well extended or customized to perform, for example, the incorporation of O-glycans according to users' needs.

gQuant Tool Development

gQuant was created with a series of algorithms, including spectral preprocessing, glycan mapping, quantitation and ratio calculations. For profiled data, peak picking and centroiding algorithms were applied using Gaussian distribution fitting adapted from pymzML (Kösters et al., 2018). The noise level was also calculated by means of the "median" or "mean" value of the spectra multiplied by a predefined coefficient to facilitate signal-to-noise level estimation and signal filtration. Spectra were then deisotoped, and peaks without sufficient isotopes (by default, the envelope should contain more than three isotopes) were discarded. All satisfied isotope envelopes were recorded and transferred for further glycan matching and quantitation.

The glycan database was also customized for all given charge carriers (adducts, for example; H+, Na+, K+, et al. in positive mode and -H, Na-2H, K-2H, et al. in negative mode) and glycan derivatives (for example, 2-AB, reducing end, or nonderivatized, which was named "free end" in gQuant). After spectral preprocessing and database customization, a molecular-weight-based glycan exhaustive match algorithm was applied to all monoisotopic peaks in the spectra. All possible glycan molecular forms (adducts, derivatizations) were recorded. For stable isotopic labeling-based glycan quantitation data, heavily labeled glycans were also searched, and each matched glycan (light isotopic labeled and heavy isotopic labeled) was tagged with "L" and "H", respectively, in the final results.

If the parameter of delta mass was not defined (set as zero by default), gQuant only output matched glycans, peaks and corresponding peak intensities, and the results can assist users in determining all detected glycans or calculating normalized percentages. If delta mass was not zero, the final glycan relative quantitation ratio calculation was calculated in the tool by **Eq. 1** and **Eq. 2** as shown below:

FIGURE 1 | General pipeline of gQuant for MS-based glycan quantitation using stable isotope labeling.
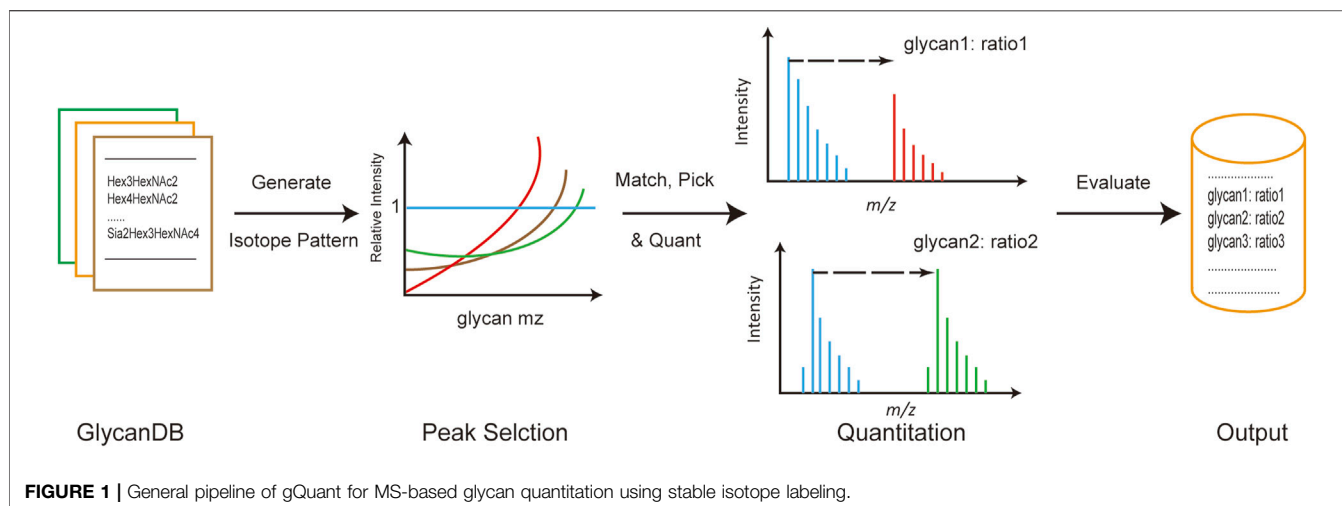
$$\text{Ratio}_{\text{calibr}} = \frac{\text{Intensity}_{\text{light}}}{\text{Intensity}_{\text{heavy}} - \text{Intensity}_{\text{light}} \times \text{calibr}_{\text{ratio}}} \quad (1)$$

$$\text{calibr}_{\text{ratio}} = \frac{\text{Theoretical Abundance}_{\text{isotope for quantitation+delta mass}}}{\text{Theoretical Abundacne}_{\text{isotope for quantitation}}} \quad (2)$$

where delta mass was the round number of delta mass.

To facilitate the usage of gQuant, a user-friendly interface (UI) was also implemented. Detailed information on the UI is presented in the *Results and Discussion* section and **Supplementary material 1**.
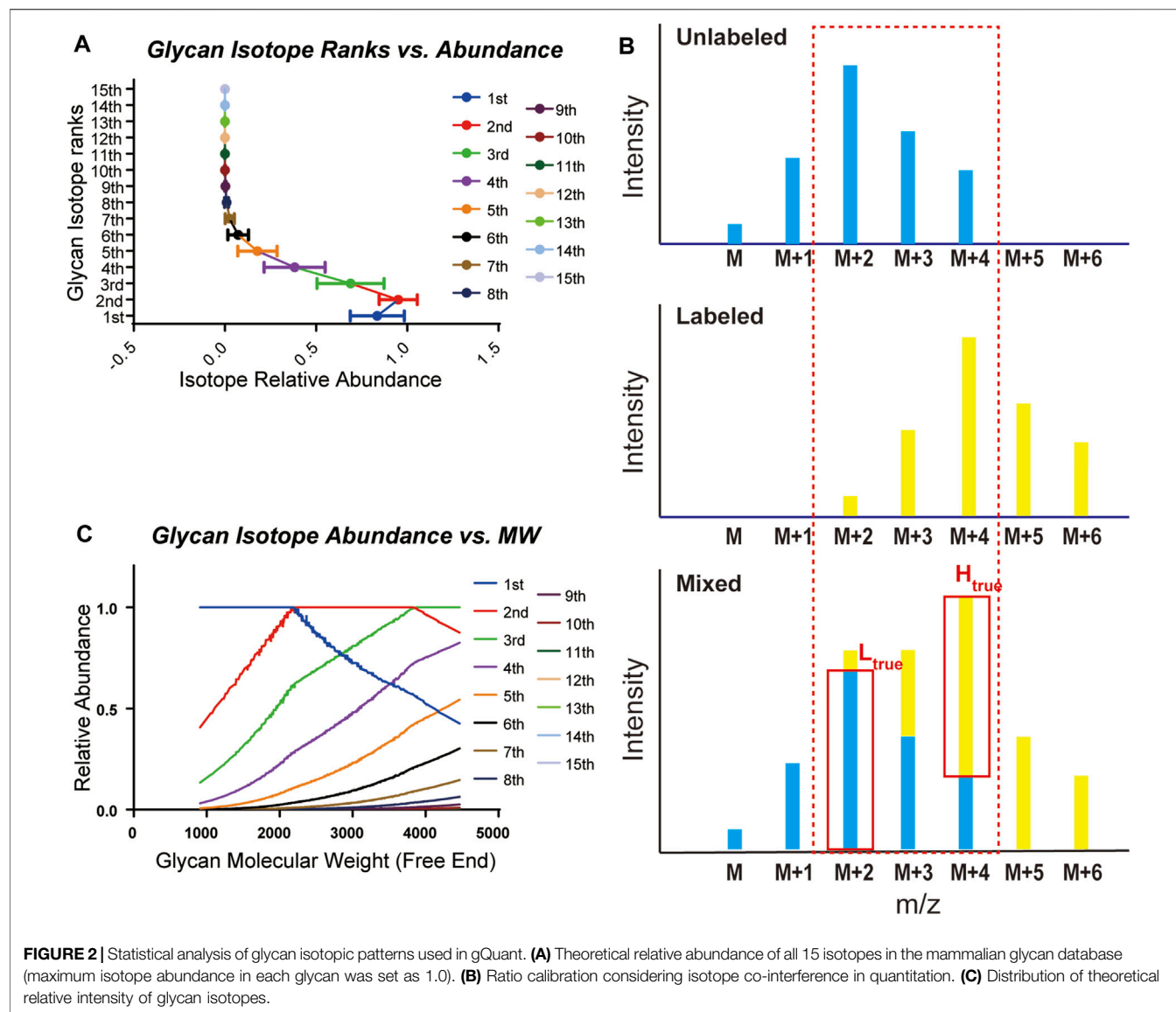
Feasibility Test of gQuant

A series of datasets (GREDIL and PFBHA datasets) were applied to evaluate the performance of gQuant. The GREDIL dataset used $NaBH_4$ to reduce PNGase F-released glycans and incorporated ^{18}O + deuterium to form a delta mass of 3.0 Da between samples, while the PFBHA dataset used o-(2,3,4,5,6-pentafluorobenzyl) hydroxylamine hydrochloride (PFBHA) and PFBHA-2 deuterium for sample glycan derivatization. Detailed methods of the two datasets were well described in previous publications (Cao et al., 2015; Yang et al., 2019). The datasets were adopted as part of the following search parameters. For the PFBHA dataset, raw data were preprocessed by the vendor-provided software Data Explorer 4.3 (AB SIEX, Framingham, MA, United States). Briefly, centroid peaks were generated with the *Centroiding* function in the *Process* menu. Peak lists with m/z and intensity were then exported to standalone ASCII formatted data files via *File* menu→*Export*→*ASCII Spectrum*→*Save*. For the GREDIL dataset, since the vendor software failed to centroid peaks properly, profiled form data were exported by the Shimadzu Biotech Launchpad MALDI-MS application 2.9.3 (Shimadzu, Japan) *via File* menu→*Export*→*ASCII* ... → with *Processed* selected and saved. All data files were then analyzed with the gQuant tool with the m/z tolerance set to 50ppm, adducts of H+, Na+, K+ considered, a positive mode, max charge of 3, signal-

to-noise ratio set as 10 and derivatization type and delta masses set according to corresponding experimental designs.

RESULTS AND DISCUSSION

The general process of developing the gQuant tool is depicted in **Figure 1**. Briefly, integrated glycan databases for human and mammalian samples were carefully collected and extracted from the current glycan database and published literature. As a result, a total of 419 glycan compositions were recorded in the mammalian glycan database, with maximum hexose (Hex), N-acetylhexosamine (HexNAc), N-acetylneuraminic acid/N-glycolylneuraminic acid (NeuAc/NeuGc), and fucose (dHex) values of 12, 7, 4, and 5, respectively, and 344 entries were recorded in the human glycan database (No NeuGc). The maximum glycan molecular weight of the database was approximately 4,500 Da, which covered the most common glycans (**Supplementary Table 1** and **Supplementary Table 2**). Chemical element compositions of each glycan were then generated, and the theoretical isotope envelopes for each composition were calculated. Peak lists were obtained by matching the isotope envelopes with the theoretical glycan m/z under predefined mass tolerance, derivatizations, charge carriers and isotopic labeling mass (according to experimental designs and instrument types). Then, annotated peaks were quantitated by gQuant, and the quantitation ratio (heavy to light) was reported.

Glycans are composed mainly of C, H, N, and O elements and show distinct isotopic patterns. Isotopic patterns can be further utilized to assist glycan identification and quantitation. Therefore, it is important to study the isotopic patterns of glycans to better understand glycan mass spectrometry behaviors such as isotope envelopes. Statistical investigation of glycan isotopic patterns benefited from the well-constructed mammalian glycan database. As shown in **Supplementary Table 1** and **Supplementary Table 2**, by theoretical calculation of glycan elemental compositions, glycans in the

FIGURE 2 | Statistical analysis of glycan isotopic patterns used in gQuant. **(A)** Theoretical relative abundance of all 15 isotopes in the mammalian glycan database (maximum isotope abundance in each glycan was set as 1.0). **(B)** Ratio calibration considering isotope co-interference in quantitation. **(C)** Distribution of theoretical relative intensity of glycan isotopes.

gQuant database reported a maximum of 14 isotopes with abundance. In addition, it can be concluded that the eighth isotopes accounts for less than 1% of the highest isotopes (**Figure 2A**, **Supplementary Table 3**); therefore, a mass interval of higher than 8 Da is normally sufficient to ensure avoidance of glycan isotope interference. For mass intervals smaller than 8 Da, the isotope pattern in the gQuant can be applied to recalibrate the quantitation ratio, as shown in **Figure 2B** and **Equation (2)** described in Methods part. As glycan molecular weight increased, the highest isotope ion peak gradually changed from the first isotope ion peak to the second one at approximately 2,250 and to the third one at 3,850 (**Figure 2C**). Since glycans are highly polar and fragment poorly in MS, it would be meaningful to use more abundant isotopes for quantitation to increase quantitation sensitivity and accuracy. The features have already been incorporated in gQuant to improve the performance of glycan quantitation.

The accuracy and feasibility of gQuant was assessed by quantitating PFBHA datasets (Yang et al., 2019), in which glycans from model glycoproteins were derivatized with PFBHA or PFBHA-2 deuterium. As shown in **Supplementary Table 4**, glycan derivatives were almost always found in the top-ranked results, as the matched mass tolerances were in ascending order. In addition, the reported ratios in the L/H ratio column were close to 1.0, as experimentally designed. In addition to quantitation ratios, gQuant can also report other information, such as glycan compositions, glycan derivatization types, charge carriers, paired m/z values and intensities (**Supplementary Table 5**), thus, greatly facilitating the further collection and understanding of quantitative glycan information. Based on the detailed analysis of the theoretical glycan isotope distribution, the highest isotope-based selection for the quantitation strategy was applied in gQuant to enhance glycan quantification. It should be noted that only the molecular weight
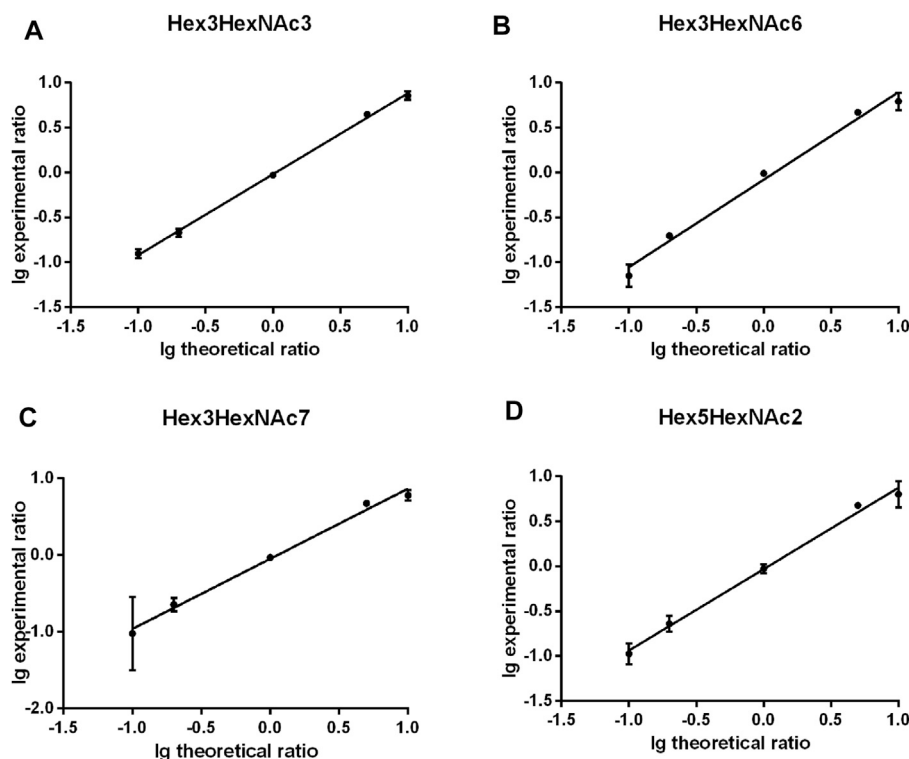
FIGURE 3 | Comparison of gQuant the reported ratio with experimental mixed ratios (ranging from 10:1, 5:1, 1:1 1:5, and 1:10) of four glycans. **(A)** Hex3HexNAc3, **(B)** Hex3HexNAc6, **(C)** Hex3HexNAc7 and **(D)** Hex5HexNAc2.

is used for glycan matching; therefore, gQuant reports all matched glycans. For further filtration of suspicious identifications, we suggest setting proper mass tolerance according to different mass spectrometry instruments and strategies they used. For example, the MS tolerance of 50ppm is recommended for the PFBHA datasets. Further manual check or even tandem mass spectrometry verification may be necessary for some target candidates or some ambiguous results as appropriate.

The quantitation performance of gQuant was further evaluated on differentially labeled glycans with defined ratios (heavy labeled and unlabeled samples with a mass shift of 2.0 Da (2 deuterium vs 2 hydrogen) mixed at 10:1, 5:1, 1:1, 1:5, and 1:10, v/v, respectively) from PFBHA datasets. As shown in **Figure 3** and **Supplementary Table 6**, the quantitation ratios reported by gQuant were consistent with the experimental designs and manual analytical results for four glycans (Hex3HexNAc3 [M + Na]$^+$ 1,331.40601, Hex3HexNAc6 [M + Na]$^+$ 1940.64413, Hex3HexNAc7 [M + Na]$^+$ 2,143.7235, and Hex5HexNAc2 [M + Na]$^+$ 1,452.43229). In addition, gQuant reported more glycans (37 glycans, **Supplementary Table 4**) than manually did (only two glycans were reported). It was also observed that there could be a fixed mass offset in different tests due to the instrumental errors. To reduce the mass shift error, gQuant also offered a simple parameter of calibrated mass to recalibrate the mass error and improve the matching and quantitating accuracy. In addition to test on quantifying glycans from single glycoprotein, gQuant

was further applied to quantitation of glycans from a complex sample, rat serum, the data of which were obtained from GREDIL dataset (Cao et al., 2015). As shown in **Supplementary Table 7**, a total of 15 glycan peaks with calibrated mass, such as m/z 935.39 Hex3HexNAc2 Na$^+$, m/z 1,665.44 Hex5HexNAc4 Na+/ Hex4HexNAcdHex1 K$^+$, were identified and quantified. The reported glycan quantitation ratio was closed to mixed ratio. Identified glycans were relatively consistent with previous publication (Cao et al., 2015). Thus, gQuant showed high accuracy and was informative for glycan quantitation data processing.

The efficiency of gQuant was also tested on the aforementioned dataset. As shown in **Supplementary Table 8**, it took less than 3 min to process a single model protein data file and less than 25 min for complex samples, while it took hours or days to process sample data manually. In addition, gQuant supports the batch running mode, and the running speed can be further accelerated by algorithm optimization for the glycan matching process in the future. Thus, because it is automated and efficient, gQuant can greatly improve the speed and throughput of glycan quantitation analysis.

The common obstacle for the application of most bioinformatic tools is the lack of user-friendly interfaces. To facilitate the usage of gQuant, a convenient user interface was designed and implemented, as shown in **Figure 4**. There are three main panels of gQuant UI: a data file input and output (IO) setting panel, a mass spectrometry parameter
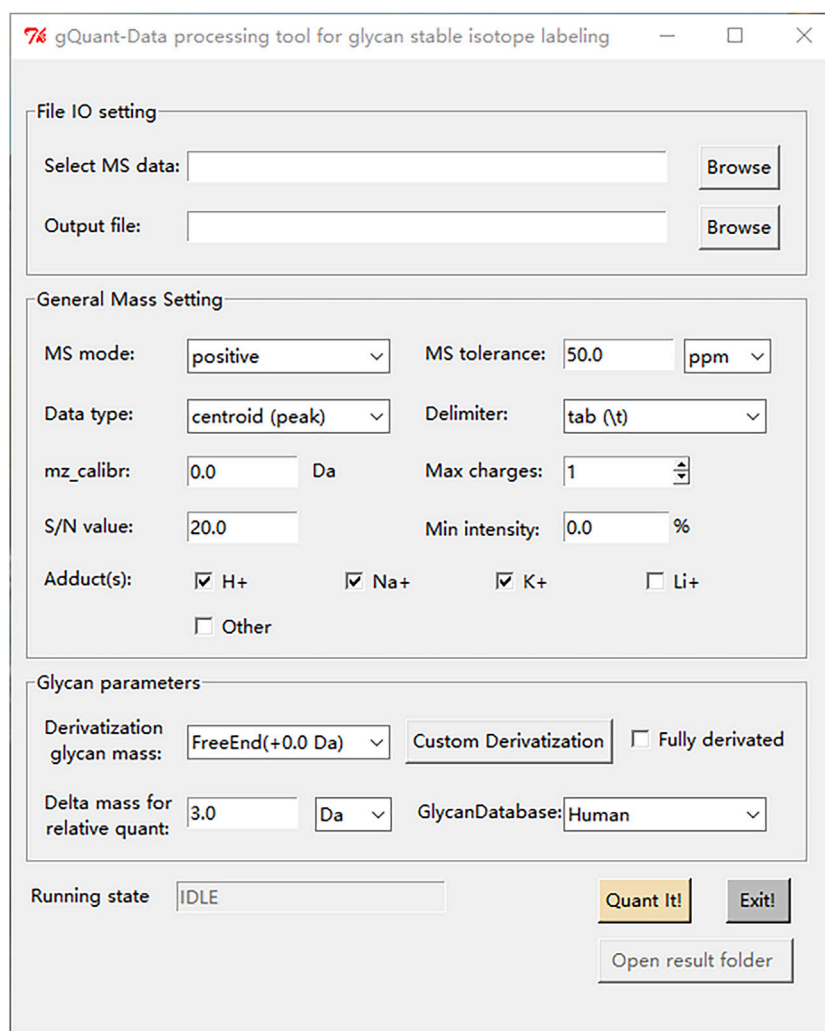
FIGURE 4 | The graphical user interface of gQuant.

configuration panel (*General Mass Setting*), and a glycan related parameter setting panel (*Glycan Parameters*). The data file input and output setting panel enables batched MS data selection and output directory settings; the *General Mass Setting* panel contains parameters such as MS mode (positive or negative), data point types (profile, or centroid), signal-to-noise level setting, charge carrier setting (adducts) checkboxes, etc.; the *Glycan Parameters* panel facilitates glycan derivatization setting either by predefined items (2-AB, reducing end, etc.,) or to define a new derivatization type by "Custom Derivatization"; and it also supports delta mass setting for relative quantitation and glycan database selection. In addition, as mentioned above, gQuant is programmed with Python 2.7 and Python 3, which is expected to be supported by most operating systems, such as Windows, Mac or Linux. Moreover, gQuant is fully open-source software under the Apache 2.0 license, and it can be easily adapted to accommodate different analytical purposes, for example, to modify the glycan database for different sourced samples or for

O-glycan analysis, or to add a function to support LC-MS quantitation data analysis.

CONCLUSION

Glycan quantitation MS data processing remains a tedious and challenging task for researchers. To tackle this problem, we designed and implemented an automated glycan quantitation tool, gQuant, in this work. gQuant is capable of automatically and efficiently processing quantitative glycan mass spectrometry data and reporting all matched glycans and quantitation ratios. gQuant was embedded in N-glycan databases for human (No NeuGc)- or mammalian (Containing NeuGc)-sourced samples. Statistical evaluation of glycan isotopic distributions was performed to study the distinct isotopic patterns of glycans. It was suggested that the glycan isotope interference decreased to less than 1% when the isobaric interval increased above 8 Da gQuant enables automated glycan matching, glycan abundance ratio

calculation and data output. Compared with manual interpretation, this tool showed good accuracy and efficiency. The performance test indicated that gQuant can successfully process glycan quantitation data and report precision ratios, that it reports more glycans than the manual method and that it is much more efficient than the manual method (with a speed as low as several minutes). Although the presented results are all from N-glycan data, gQuant can also be easily adapted for relative quantitation of other types of glycans, such as mucin-type O-glycans, with a specified or user-defined glycan database. In summary, as an open-source, convenient and efficient software tool, gQuant is expected to facilitate glycomic studies and clinical glycan MS data analysis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JH designed the informatics tools, analyzed the data and wrote the manuscript; BY and ML contributed to the manuscript preparation; PY supervised the work; WC supervised the

work, provided scientific suggestions and contributed to the manuscript revision. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.707738/full#supplementary-material

REFERENCES

Albrecht, S., Mittermayr, S., Smith, J., Martín, S. M., Doherty, M., and Bones, J. (2017). Twoplex 12/13 C6 Aniline Stable Isotope and Linkage-Specific Sialic Acid Labeling 2D-LC-MS Workflow for Quantitative N-Glycomics. *Proteomics* 17 (1-2), 1600304. doi:10.1002/pmic.201600304

Balog, C. I. A., Stavenhagen, K., Fung, W. L. J., Koeleman, C. A., McDonnell, L. A., Verhoeven, A., et al. (2012). N-glycosylation of Colorectal Cancer Tissues. *Mol. Cell Proteomics* 11 (9), 571–585. doi:10.1074/mcp.m111.011601

Cai, Y., Jiao, J., Bin, Z., Zhang, Y., Yang, P., and Lu, H. (2015). Glycan Reductive Isotope-Coded Amino Acid Labeling (GRIAL) for Mass Spectrometry-Based Quantitative N-Glycomics. *Chem. Commun.* 51 (4), 772–775. doi:10.1039/c4cc08086f

Cao, W.-Q., Liu, M.-Q., Kong, S.-Y., Wu, M.-X., Huang, Z.-Z., and Yang, P.-Y. (2020). Novel Methods in Glycomics: a 2019 Update. *Expert Rev. Proteomics* 17 (1), 11–25. doi:10.1080/14789450.2020.1708199

Cao, W., Zhang, W., Huang, J., Jiang, B., Zhang, L., and Yang, P. (2015). Glycan Reducing End Dual Isotopic Labeling (GREDIL) for Mass Spectrometry-Based Quantitative N-Glycomics. *Chem. Commun.* 51 (71), 13603–13606. doi:10.1039/c5cc05365j

Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S. M. (2008). GlycoWorkbench: a Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *J. Proteome Res.* 7 (4), 1650–1659. doi:10.1021/pr7008252

Eakin, A. J., Bustard, M. J., McGeough, C. M., Ahmed, T., Bjourson, A. J., and Gibson, D. S. (2016). Siglec-1 and -2 as Potential Biomarkers in Autoimmune Disease. *Prot. Clin. Appl.* 10 (6), 635–644. doi:10.1002/prca.201500069

Hart, G. W., and Copeland, R. J. (2010). Glycomics Hits the Big Time. *Cell* 143 (5), 672–676. doi:10.1016/j.cell.2010.11.008

Hu, Y., Zhou, S., Yu, C.-Y., Tang, H., and Mechref, Y. (2015). Automated Annotation and Quantitation of Glycans by Liquid Chromatography/electrospray Ionization Mass Spectrometric Analysis Using the MultiGlycan-ESI Computational Tool. *Rapid Commun. Mass. Spectrom.* 29 (1), 135–142. doi:10.1002/rcm.7093

Jansen, B. C., Falck, D., de Haan, N., Hipgrave Ederveen, A. L., Razdorov, G., Lauc, G., et al. (2016). LaCyTools: A Targeted Liquid Chromatography-Mass Spectrometry Data Processing Package for Relative Quantitation of Glycopeptides. *J. Proteome Res.* 15 (7), 2198–2210. doi:10.1021/acs.jproteome.6b00171

Jansen, B. C., Reiding, K. R., Bondt, A., Hipgrave Ederveen, A. L., Palmblad, M., Falck, D., et al. (2015). MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality Control Developed for Glycomic and Glycoproteomic MALDI-MS. *J. Proteome Res.* 14 (12), 5088–5098. doi:10.1021/acs.jproteome.5b00658

Kösters, M., Leufken, J., Schulze, S., Sugimoto, K., Klein, J., Zahedi, R. P., et al. (2018). pymzML v2.0: Introducing a Highly Compressed and Seekable Gzip Format. *Bioinformatics (Oxford, England)* 34 (14), 2513–2514. doi:10.1093/bioinformatics/bty046

Liu, M., Zhang, Y., Chen, Y., Yan, G., Shen, C., Cao, J., et al. (2014). Efficient and Accurate Glycopeptide Identification Pipeline for High-Throughput Site-Specific N-Glycosylation Analysis. *J. Proteome Res.* 13 (6), 3121–3129. doi:10.1021/pr500238v

Lu, L. L., Chung, A. W., Rosebrock, T. R., Ghebremichael, M., Yu, W. H., Grace, P. S., et al. (2016). A Functional Role for Antibodies in Tuberculosis. *Cell* 167 (2), 433–443. doi:10.1016/j.cell.2016.08.072

Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., et al. (2012). GlycReSoft: a Software Package for Automated Recognition of Glycans from LC/MS Data. *PloS one* 7 (9), e45474. doi:10.1371/journal.pone.0045474

Orlando, R., Lim, J.-M., Atwood, J. A., 3rd, Angel, P. M., Fang, M., Aoki, K., et al. (2009). IDAWG: Metabolic Incorporation of Stable Isotope Labels for Quantitative Glycomics of Cultured Cells. *J. Proteome Res.* 8 (8), 3816–3823. doi:10.1021/pr8010028

Prien, J. M., Prater, B. D., Qin, Q., and Cockrill, S. L. (2010). Mass Spectrometric-Based Stable Isotopic 2-aminobenzoic Acid Glycan Mapping for Rapid Glycan Screening of Biotherapeutics. *Anal. Chem.* 82 (4), 1498–1508. doi:10.1021/ac902617t

Ranzinger, R., Frank, M., von der Lieth, C.-W., and Herget, S. (2009). Glycome-DB.org: a portal for Querying across the Digital World of Carbohydrate Sequences. *Glycobiology* 19 (12), 1563–1567. doi:10.1093/glycob/cwp137

Shah, P., Wang, X., Yang, W., Toghi Eshghi, S., Sun, S., Hoti, N., et al. (2015). Integrated Proteomic and Glycoproteomic Analyses of Prostate Cancer Cells

Reveal Glycoprotein Alteration in Protein Abundance and Glycosylation*. *Mol. Cell Proteomics* 14 (10), 2753–2763. doi:10.1074/mcp.m115.047928

Smith, J., Mittermayr, S., Váradi, C., and Bones, J. (2017). & Bones, J.Quantitative Glycomics Using Liquid Phase Separations Coupled to Mass Spectrometry. *Analyst* 142 (5), 700–720. doi:10.1039/c6an02715f

Wang, C., Zhang, P., Jin, W., Li, L., Qiang, S., Zhang, Y., et al. (2017). Quantitative O -glycomics Based on Improvement of the One-Pot Method for Nonreductive O -glycan Release and Simultaneous Stable Isotope Labeling with 1-(d 0/d 5)phenyl-3-Methyl-5-Pyrazolone Followed by Mass Spectrometric Analysis. *J. Proteomics* 150, 18–30. doi:10.1016/j.jprot.2016.08.012

Wuhrer, M. (2013). Glycomics Using Mass Spectrometry. *Glycoconj J.* 30 (1), 11–22. doi:10.1007/s10719-012-9376-3

Xu, C., and Ng, D. T. W. (2015). Glycosylation-directed Quality Control of Protein Folding. *Nat. Rev. Mol. Cel Biol.* 16 (12), 742–752. doi:10.1038/nrm4073

Yang, L., Du, X., Peng, Y., Cai, Y., Wei, L., Zhang, Y., et al. (2019). Integrated Pipeline of Isotopic Labeling and Selective Enriching for Quantitative Analysis of N-Glycome by Mass Spectrometry. *Anal. Chem.* 91 (2), 1486–1493. doi:10.1021/acs.analchem.8b04525

Yu, C.-Y., Mayampurath, A., Hu, Y., Zhou, S., Mechref, Y., and Tang, H. (2013b). Automated Annotation and Quantification of Glycans Using Liquid Chromatography-Mass Spectrometry. *Bioinformatics (Oxford, England)* 29 (13), 1706–1707. doi:10.1093/bioinformatics/btt190

Yu, C.-Y., Mayampurath, A., and Tang, H. (2013a). Software Tools for Glycan Profiling. *Methods Mol. Biol. (Clifton, N.J.)* 951, 269–276. doi:10.1007/978-1-62703-146-2_18

Zhang, W., Cao, W., Huang, J., Wang, H., Wang, J., Xie, C., et al. (2015). PNGase F-Mediated Incorporation of 18O into Glycans for Relative Glycan Quantitation. *Analyst* 140 (4), 1082–1089. doi:10.1039/c4an02073a

TiO$_2$ Simultaneous Enrichment, On-Line Deglycosylation, and Sequential Analysis of Glyco- and Phosphopeptides

Cheng Chen[1,2†], Xiaofei Zhang[1†], Xuefang Dong[1], Han Zhou[1], Xiuling Li[1,3]* and Xinmiao Liang[1,3]*

[1]Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Ganjiang Chinese Medicine Innovation Center, Nanchang, China

Reversible protein glycosylation and phosphorylation tightly modulate important cellular processes and are closely involved in pathological processes in a crosstalk dependent manner. Because of their significance and low abundances of glyco- and phosphopeptides, several strategies have been developed to simultaneously enrich and co-elute glyco- and phosphopeptides. However, the co-existence of deglycosylated peptides and phosphopeptides aggravates the mass spectrometry analysis. Herein we developed a novel strategy to analyze glyco- and phosphopeptides based on simultaneous enrichment with TiO$_2$, on-line deglycosylation and collection of deglycosylated peptides, and subsequent elution of phosphopeptides. To optimize on-line deglycosylation conditions, the solution pH, buffer types and concentrations, and deglycosylation time were investigated. The application of this novel strategy to 100 μg mouse brain resulted in 355 glycopeptides and 1,975 phosphopeptides, which were 2.5 and 1.4 folds of those enriched with the reported method. This study will expand the application of TiO$_2$ and may shed light on simultaneously monitoring protein multiple post-translational modifications.

Keywords: simultaneous enrichment, on-line deglycosylation, sequential elution, phosphopeptides, glycopeptides, TiO$_2$

INTRODUCTION

Protein glycosylation and phosphorylation are two of the most ubiquitous and important post-translational modifications (PTMs) and they play vital roles in regulating a variety of physiological and pathological processes. These two types of PTMs rarely work alone but interplay in a crosstalk dependent manner. Increasing lines of evidence indicate that the crosstalk between protein glycosylation and phosphorylation is involved in many important biological events (Hart et al., 2011) and their abnormalities are closely associated with many serious diseases (Liu et al., 2002; Takeda et al., 2015; Ma et al., 2017; Zhang et al., 2017). For example, reciprocal protein glycosylation and phosphorylation co-regulate nutrient sensing, neural development, and cell cycle (Hart et al., 2011); the hyperphosphorylation of tau protein is triggered by its abnormal N-linked glycosylation, which is key to Alzheimer's disease (Losev et al.,

2021). Thus simultaneous monitoring these two PTMs and elucidating of their crosstalk in biological samples, especially for precious and trace of biological samples, have pathological and clinical significance.

In the past decade, PTM proteomics has developed rapidly, benefiting from advances of mass spectrometry (MS) technology and improvement of enrichment strategies. However, it remains challenging to simultaneously analyze glyco- and phosphopeptides, due to their low abundances and the high complexity of biological samples. To date, several materials have been developed for simultaneous enrichment of glyco- and phosphopeptides, including metal oxide affinity chromatography- (MOAC-) based materials (Xu et al., 2016; Xu et al., 2017; Sun et al., 2019), immobilized metal ion affinity chromatography- (IMAC-) based materials (Melo-Braga et al., 2014; Zou et al., 2017; Cho et al., 2019; Wang et al., 2019), and hydrogen bond-based polymer material (Lu et al., 2020). As the representative of MOAC-based materials, TiO$_2$ is the most commonly used for its excellent robustness (Peng et al., 2017), reproducibility (Sun et al., 2019), and commercial availability. The affinity of TiO$_2$ to glycopeptides is based on ligand-exchange and hydrophilic interactions between TiO$_2$ and saccharides (Sheng et al., 2013) and binding of TiO$_2$ toward phosphopeptides is based on Lewis acid-base interaction between TiO$_2$ and phosphate groups (Yan and Deng 2019). In classical TiO$_2$ simultaneous enrichment cases (**Scheme 1A**), the captured glyco- and phosphopeptides are co-eluted (Hu et al., 2018; Palmisano et al., 2012a) and undergo an enzymatic deglycosylation treatment for glycosylation sites identification (Deeb et al., 2014). However, the co-existence of deglycosylated peptides and phosphopeptides will increase the burden of further MS analysis. To reduce the complexity of samples, the two-dimensional enrichment is often employed to address this issue (Melo-Braga et al., 2015), but additional processing steps may lead to a low recovery of targets. Besides, the co-existent phosphopeptides can be hydrolyzed under alkaline deglycosylation conditions (Thompson et al., 2003), and the desalting procedure after deglycosylation will aggravate the loss of PTM-peptides.

Herein, we developed a novel strategy to simultaneously enrich and sequentially analyze glyco- and phosphopeptides, which consists of the simultaneous enrichment of glyco- and phosphopeptides with TiO$_2$ and the on-line deglycosylation to obtain deglycosylated peptides and sequential elution of phosphopeptides. The on-line deglycosylation is key to the success of this strategy. Thus, some key factors of the on-line deglycosylation were investigated and optimized, such as solution pH, buffer concentrations, and deglycosylation time. This work will have a great potential in the simultaneous analysis of the protein glycosylation and other multiple PTMs.

EXPERIMENTS

Reagents and Standards

HPLC-grade acetonitrile (ACN), urea, ammonium hydroxide (NH$_3$·H$_2$O), DL-dithiothreitol (DTT), iodoacetamide (IAA), ammonium formate (HCOONH$_4$),

ammonium acetate (NH$_4$OAc), ammonium bicarbonate (NH$_4$HCO$_3$), formic acid (FA), acetic acid, glycolic acid, [Glu1]-Fibrinopeptide B human (GFB) (internal standard), bovine fetuin (standard glycoprotein), α-casein (standard phosphoprotein), and trypsin were purchased from Sigma-Aldrich (St Louis, MO, United States). Standard phosphopeptide (with sequence of HS*PIAPSSPSPK) was synthesized by Qiangyao Biotechnology Co., Ltd. (Shanghai, China). Trifluoroacetic acid (TFA), acetone, and ethyl alcohol were purchased from Shanghai Macklin Biochemical Co., Ltd. (Shanghai, China). PNGase F was purchased from New England Biolabs (Ipswich, MA, United States). Radioimmunoprecipitation (RIPA) lysis buffer and bicinchoninic acid (BCA) protein assay kit were purchased from Beyotime Biotechnology (Shanghai, China). GELoader was purchased from Eppendorf (Hamburg, Germany). TiO$_2$ was purchased from GL Sciences (Tokyo, Japan). C18HC material was purchased from ACCHROM (Wenling, China). Mouse brains were provided by Dalian Medical University (Dalian, China). Pure water was purified with a Milli-Q system (Millipore, Milford, MA, United States).

Instruments

The peptide samples and TiO$_2$ were mixed in a thermomixer (Qianjun, Shanghai, China). TiO$_2$ was separated from the mixture by centrifuge (Merck, Milford, MA, United States). A Labconco CentriVap system (Labconco, Kansas, MO, United States) was applied to dry samples in specific steps. Determination of protein concentrations was by a microplate reader (Thermo Scientific, San Jose, CA, United States). The qualitative analysis of the standard protein digests was conducted on a nano electrospray ionization quadrupole time-of-flight mass spectrometer (ESI-Q-TOF MS) (Waters, Manchester, United Kingdom). The qualitative analysis of the protein digests extracted from the mouse brains was performed using an Orbitrap Eclipse Tribrid mass spectrometer and a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Scientific, San Jose, CA, United States).

Protein Extraction

A mouse brain tissue was cleaned and cut into pieces. Then the tissue pieces were ground into white powder in a mortar with liquid nitrogen. The tissue powder was mixed with 2 ml ice-cold RIPA lysis buffer and transferred into a 5 ml centrifuge tube. The mixture was placed in an ultrasonic crushing machine on ice for 5 min. The sonication sequential mode was 1 s on and 3 s off, in addition to 30-minute cycles. After the lysis, the mixture was centrifuged at 13,000 g for 30 min at room temperature. The supernatant was collected and a precipitant was added. This mixture was deposited overnight at -20°C. After sedimentation, the sample was centrifuged at 13,000 g for 30 min and the supernatant was removed. The precipitation was washed with 3.6 ml of acetone, then 3.6 ml of anhydrous ethanol, and redissolved in 6 M urea. The concentration of the redissolved protein solution was determined by a bicinchoninic acid (BCA) method (Hussain et al., 2014). The animal experiments were authorized by the Experimental Animal Center of Dalian Medical University.

Tryptic Digestion of the Protein and Sample Desalination

The above protein solution was diluted to 1 mg/ml with 6 M urea. 1 ml of protein solution was mixed with 50 μL of DTT (200 mM) and incubated at 56°C for 45 min. Then 200 μL of IAA (200 mM) was added and the mixture was placed in dark for 30 min. Then, 7.25 ml of 50 mM NH_4HCO_3 aqueous solution and 250 μg trypsin were added in the mixture and incubated at 37°C overnight. Finally, 5 μL FA was added to stop the digestion. Then, the sample was desalted with C18HC packed solid phase extraction microcolumns.

Enrichment of Glyco- and Phosphopeptides from the Mixture of Fetuin and α-Casein Digest

The enrichment was performed following a reported method (Palmisano et al., 2012b) with minor modification. The tryptic digests of fetuin (5 μg) and α-casein (5 μg) were mixed in 50 μL of 80% ACN/5% TFA, with 1 M glycolic acid (loading buffer). The mixture was added with 1 mg TiO_2 material and incubated for 15 min. After removal of the supernatant by centrifugation, the TiO_2 material was washed twice with 50 μL of loading buffer and centrifuged to remove the supernatants. The enriched peptides were used in further experiments.

On-Line Deglycosylation of the Glycopeptides

The TiO_2 materials attached with glyco- and phosphopeptides were mixed with 45 μL of 50 mM NH_4OAc and 5 μL of PNGase F (2,500 U). The resulting solution was incubated for 16 h at 37°C. After centrifugation, the supernatant was collected and desalted for the MS analysis.

Effect of the Solution pH on the On-Line Deglycosylation

The TiO_2 materials attached with glyco- and phosphopeptides were separately suspended in four solutions with different pH values: 50 mM $HCOONH_4$ (pH 3.0), 50 mM NH_4OAc (pH 6.9), 50 mM NH_4HCO_3 (pH 8.3), and 0.1% $NH_3 \cdot H_2O$ (v/v, pH 11.5). For each solution, after incubation for 3 h at 37°C, the supernatant was collected by centrifuge and desalted for the MS analysis.

Effect of the NH₄OAc Concentration on the On-Line Deglycosylation

The TiO_2 materials attached with glyco- and phosphopeptides were, respectively, suspended in 45 μL of NH_4OAc solutions at different concentrations (5, 10, 20, 25, and 50 mM). For each solution, 5 μL of PNGase F (2,500 U) was added and it was incubated at 37°C overnight. After that, the supernatants were removed by centrifugation. The deglycosylated peptides were eluted with 80 μL of 40% ACN/5% TFA. Subsequently, the phosphopeptides were eluted with 80 μL of 5% (v/v) $NH_3 \cdot H_2O$. The elution fractions were vacuum-dried and desalted in 19 μL of 50% ACN/0.1% FA. Before the MS analysis, 1 μL internal standard (1 pmol GFB) was added in each sample.

Effect of the Deglycosylation Time on the On-Line Deglycosylation

6 mg TiO_2 material, which was attached with glyco- and phosphopeptides, was mixed with 174 μL of 50 mM NH_4OAc and 5 μL of PNGase F (2,500 U). The mixture was incubated at 37°C and 30 μL of suspended sample was taken after 6, 9, 12, 24, and 48 h incubation, separately. TiO_2 was isolated by centrifugation, and the deglycosylated peptides and bound phosphopeptides on TiO_2 were sequentially eluted and treated as described above.

Optimization of the Deglycosylated Peptides Elution Conditions

Three experiments were performed to investigate the elution effectivities of different eluents. The experiments had the same simultaneous enrichment and on-line deglycosylation processes described above but varied in the later procedures. After on-line deglycosylation, for the first experiment, the 1 mg TiO_2 material was successively washed with 80 μL of 5 mM NH_4HCO_3, 80 μL of 10 mM NH_4HCO_3, 80 μL of 20 mM NH_4HCO_3, and 80 μL of 5% (v/v) $NH_3 \cdot H_2O$. For the second experiment, the 1 mg TiO_2 material was washed thrice with 20 μL of 50% ACN/1% FA. For the third experiment, the 1 mg TiO_2 material was successively washed with 120 μL of 40% ACN/5% TFA, 120 μL of 5% TFA, and 80 μL of 5% (v/v) $NH_3 \cdot H_2O$. The supernatants of each wash were desalted for the MS analysis.

Application to the Enrichment of Mouse Brain Actual Sample

Simultaneous Enrichment: 100 μg mouse brain lysate was digested by trypsin, desalted, dried, and dissolved in 50 μL of 80% ACN/5% TFA (1 M glycolic acid). The sample was mixed with 2 mg TiO_2 material and incubated for 30 min. After centrifugation, the precipitate was washed twice with 50 μL of loading buffer. The supernatant was removed by centrifugation.

On-Line Deglycosylation: After the simultaneous enrichment, the TiO_2 material was mixed with 49 μL of 50 mM NH_4OAc and 1 μL of PNGase F (500 U). The mixture was incubated at 37°C for 12 h. The supernatant was removed by centrifugation.

Sequential Elution: The deglycosylated peptides and the bound phosphopeptides on TiO_2 were sequentially eluted with 80 μL of 40% ACN/5% TFA and 80 μL of 5% (v/v) $NH_3 \cdot H_2O$. The eluates were desalted and dried for the MS analysis.

The control experiment was performed on another 100 μg mouse brain lysate according to literature (Palmisano et al., 2012b).

Mass Spectrometry Analysis and Data Processing

The Dionex UltiMate 3000 RSLC system for chromatographic separation included a C18 trap column (75 μm × 20 mm, 3 μm) and a C18 analytical column (75 μm × 50 mm, 2 μm). The
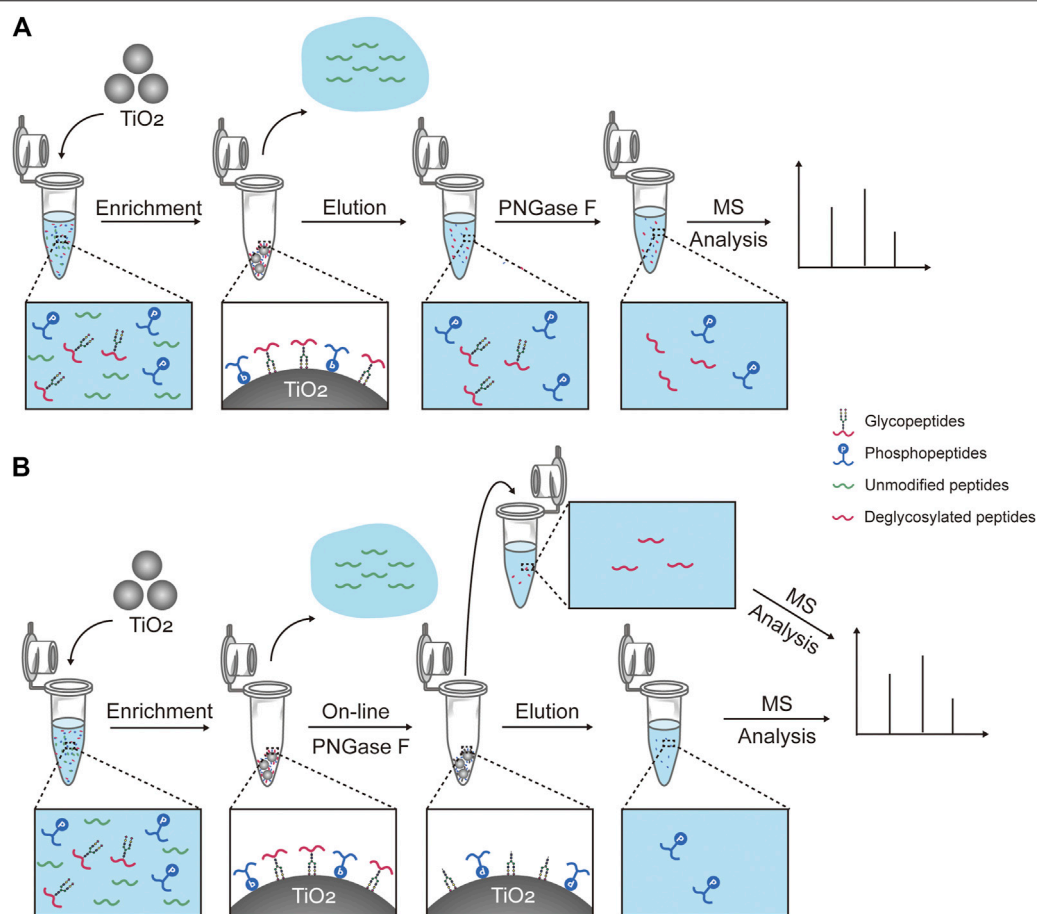
SCHEME 1 | Workflows for capturing and treatment of glyco- and phosphopeptides prior to the mass spectrometry analysis. **(A)** Workflow of one of the reported methods, which consists of the simultaneous enrichment with TiO₂, the co-elution, and the deglycosylation with PNGase F (hereinafter abbreviated to "reported method"). **(B)** Workflow of our strategy, which consists of the simultaneous enrichment with TiO₂, the on-line deglycosylation with PNGase F, and the sequential elution.

injected volume was 9 μL at a flow rate of 300 nL/min. The mobile phase was as follows: phase A was 0.1% FA and phase B was 80% ACN/0.1% FA. The gradient elution was as follows: 1–4% B, 4 min; 4%–8% B, 2 min; 8%–32% B, 104 min; and 32–90% B, 7 min.

The Orbitrap Eclipse Tribrid mass spectrometer was set as follows. For MS1, the spray voltage was 2.1 kV and the capillary temperature of the ion transport was 320°C. The first-stage full scanning range of mass spectrometry was m/z 300–1,500 with a resolution of 120,000. The RF Lens was set at 40%, the AGQ target was set at 300%, and the maximum injection time (MaxIT) was set to 50 ms. For MS2, the resolution was set at 30,000, the AGQ target was set at 100%, the MaxIT was set to 80 ms, the dynamic exclusion was set to 45 s, the isolation window was set to 1.6 Da, the collision energy was set at 30% HCD, and the fixed first mass was fixed to m/z 110. The data-dependent MS/MS was top speed mode with a cycle time of 2 s. The number of microscans to be performed was set at 1 scan s^{-1}.

All the MS raw data were processed by Proteome Discoverer and searched with SEQUEST against the mouse proteins in the UniProt database. The trypsin cleavage with a maximum of two leakage sites was allowed. Carbamidomethyl (C) was set as a fixed modification, and oxidation on methionine (M), acetylation of protein N terminus, phospho-modification (STY), and deamination (N) were set as the variable modifications. The false discovery rate (FDR) was set at 1%. The other conditions were set by default.

RESULTS AND DISCUSSION

The Workflow of the Simultaneous Enrichment, On-Line Deglycosylation, and Sequential Elution Strategy to Analyze Glyco- and Phosphopeptides

In this work, we developed a new strategy to simultaneously enrich and sequentially separate glyco- and phosphopeptides with high efficiency and recovery. Firstly, TiO₂ is used to simultaneously enrich glyco- and phosphopeptides from a

FIGURE 1 | Optimization of the online deglycosylation condition in different solution pH. **(A)** 50 mM HCOONH$_4$ (pH 4.6), **(B)** 50 mM NH$_4$OAc (pH 6.9), **(C)** 50 mM NH$_4$HCO$_3$ (pH 8.3), and **(D)** 0.1% (v/v) NH$_3$·H$_2$O (pH 11.5). The glycopeptides and phosphopeptides were marked with red stars and blue circles, respectively. The signals at m/z ranging from 1,265 to 1,615 in the mass spectra were amplified tenfold (×10), where "×10" represents the magnification times.

complex sample, which are bound to TiO$_2$ with their glycan chains or phosphate groups, respectively (**Scheme 1B**). After the enrichment, unmodified peptides are removed, and glyco- and phosphopeptides are attached on TiO$_2$. Secondly, the on-line deglycosylation using PNGase F is carried out to remove the glycans from the glycopeptides and produce deglycosylated peptides. Thus, the deglycosylated peptides are released from TiO$_2$ and collected. Thirdly, the attached phosphopeptides on TiO$_2$ are eluted. Finally, the deglycosylated peptides and the phosphopeptides are characterized with MS, respectively.

Effect of Solution pH on On-Line Deglycosylation

The solution pH was an important parameter for the on-line deglycosylation. An ideal solution pH should not only work out for the deglycosylation but also have no impact on the phosphopeptide retention on TiO$_2$. Here we investigated the

effect of the solution pH on the glyco- and phosphopeptides retention on TiO$_2$ (**Figure 1**). After the enrichment, the TiO$_2$ materials bound with PTM-peptides were separately resuspended in four solutions: 50 mM HCOONH$_4$ (pH 4.6), 50 mM NH$_4$OAc (pH 6.9), 50 mM NH$_4$HCO$_3$ (pH 8.3), and 0.1% NH$_3$·H$_2$O (v/v, pH 11.5). After incubation for 3 h and centrifugation, the supernatants were collected, desalted, and analyzed with MS. As shown in **Figures 1A,B**, neither glycopeptides nor phosphopeptides can be detected in the mass spectra of the HCOONH$_4$ or NH$_4$OAc treated samples. On the contrary, both glycopeptides (marked with red stars) and phosphopeptides (marked with blue circles) are clearly observed in the mass spectra of the NH$_4$HCO$_3$ and NH$_3$·H$_2$O treated samples (**Figures 1C,D**). These results implied that HCOONH$_4$ and NH$_4$OAc met the basic requirement for the on-line deglycosylation. Considering that the pH value of the NH$_4$OAc solution (pH 6.9) was much closer to the reaction condition pH (pH 7.5) of the PNGase F deglycosylation, this solution pH was chosen in later experiments.

FIGURE 2 | Optimization of the on-line deglycosylation conditions in different NH$_4$OAc concentrations **(A, C)** and deglycosylation times **(B, D)**. The relative abundances of the target deglycosylated peptide and the phosphopeptide obtained after the online deglycosylation **(A)** in different concentrations of NH$_4$OAc or **(B)** for different deglycosylation times. The relative abundances of the **s**tandard phosphopeptide **(C)** in different concentrations of NH$_4$OAc or **(D)** in 50 mM NH$_4$OAc for different times. DGP, deglycosylated glycopeptide; PP, phosphopeptide.

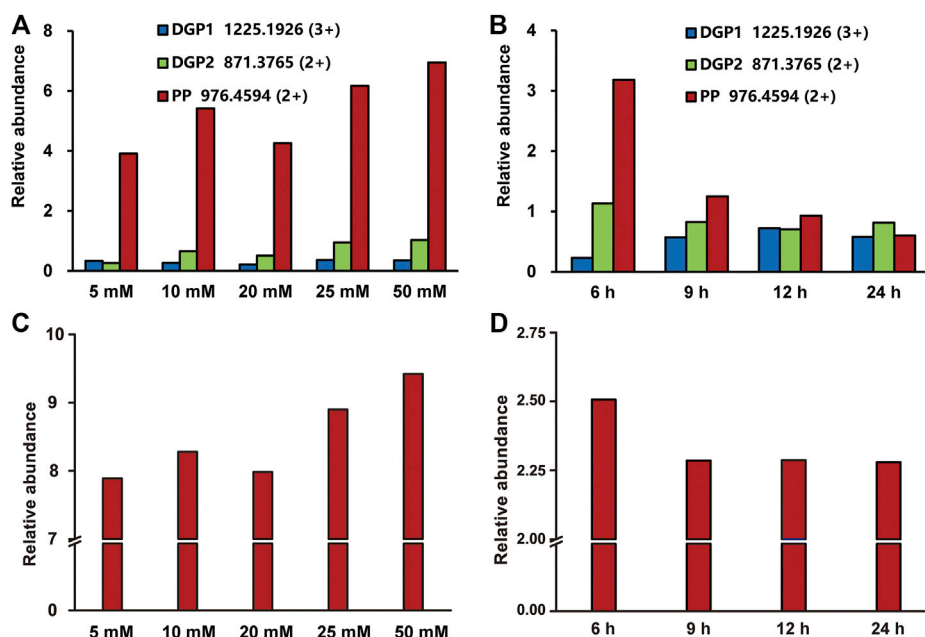## Effect of the NH$_4$OAc Concentration on the On-Line Deglycosylation

The buffer concentrations have an impact on the deglycosylation efficiency by influencing the proton exchange during the enzyme catalysis (Cheison et al., 2011; Liu et al., 2009). Thus, the effect of the NH$_4$OAc concentration on the efficiency of the on-line deglycosylation was further investigated. We measured the relative abundances of the deglycosylated peptides and phosphopeptide in the mass spectra of the elutes after the on-line deglycosylation with NH$_4$OAc in five different concentrations (5, 10, 20, 25, and 50 mM). The deglycosylated peptides at m/z 1,225.1926 (3+) and m/z 871.3765 (2+) and the phosphopeptide at m/z 976.4594 (2+) were chosen as targets of interest. The relative abundances of the target peptides were quantified with GFB as an internal standard. As shown in **Figure 2A**, the relative abundances of the deglycosylated peptide and the phosphopeptide gradually increased with the NH$_4$OAc concentration increasing and reached the maximum in 50 mM NH$_4$OAc during the investigated concentration range. For the deglycosylated peptide, this phenomenon might be attributed to the improved efficiency of proton exchange between PNGase F and the glycopeptides with increased concentrations of NH$_4$OAc, which is consistent with the reported result (Cheison et al., 2011). On the other hand, the enhanced relative abundance of the phosphopeptide might result from the reduced degree of its hydrolysis with the increased NH$_4$OAc concentration. In order to further correlate the relationship between the hydrolysis degree of the

phosphopeptide and the NH$_4$OAc concentration, the relative abundances of the standard phosphopeptide in different concentrations of the NH$_4$OAc solution were tested (**Figure 2C**). When the NH$_4$OAc concentration was low, the relative abundance of the standard phosphopeptide was low, which was ascribed to the high hydrolysis degree of the standard phosphopeptide. As the NH$_4$OAc concentrations increased, the relative abundance of the standard phosphopeptide gradually increased, which was attributed to the decreased hydrolysis degrees of the standard phosphopeptide. It seemed that higher NH$_4$OAc concentrations were favorable for inhibiting the phosphopeptide hydrolysis. These results were in good agreement with that of **Figure 2A**. Taken together, 50 mM NH$_4$OAc was chosen for further on-line deglycosylation.

Effect of the Deglycosylation Time on the On-Line Deglycosylation

Compared with the glycopeptides, the phosphopeptides are more susceptible to external influence and are unstable (Hu et al., 2020). During the deglycosylation process, the phosphopeptides hydrolyze as time goes on, and, therefore, the deglycosylation time was another important factor for the on-line deglycosylation. In order to optimize the deglycosylation time, we measured the relative abundances of the target deglycosylated peptide and the phosphopeptide after the on-line deglycosylation with different times. As shown in **Figure 2B**, the relative
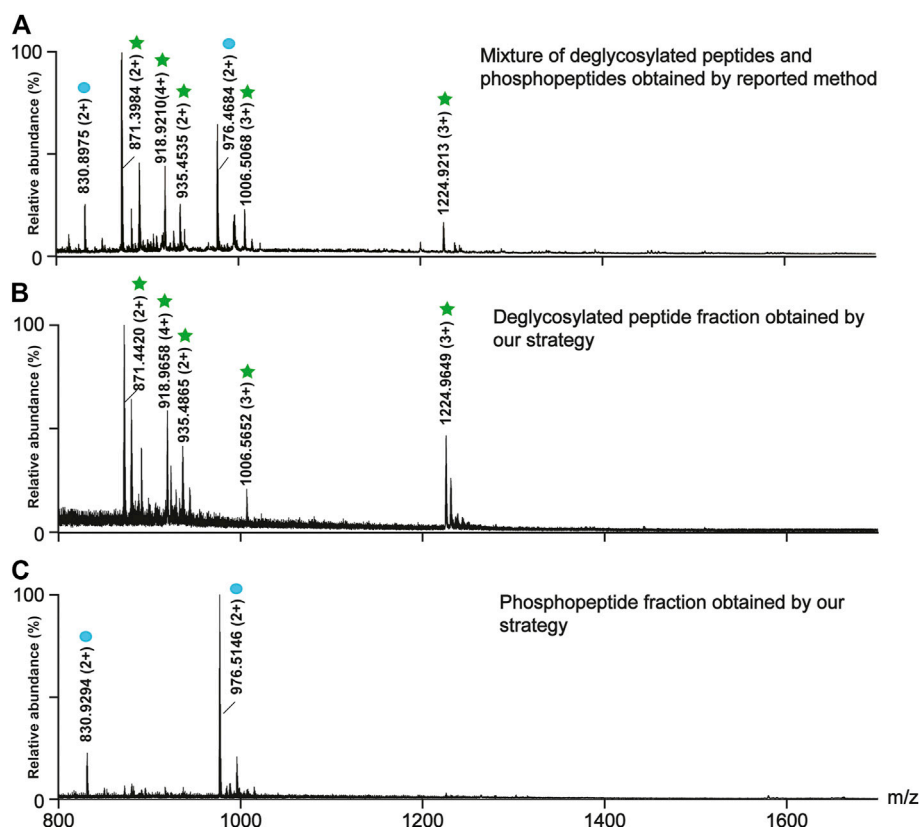
FIGURE 3 | Comparison between the reported method **(A)** and our strategy in the separation efficiency of the deglycosylated peptides **(B)** and the phosphopeptides **(C)**. **(A)** The mass spectra of the mixture of the deglycosylated peptides and the phosphopeptides obtained by the reported method. The mass spectra of **(B)** the deglycosylated peptide fraction and **(C)** the phosphopeptide fraction obtained by our strategy. The deglycosylated peptides and the phosphopeptides are marked with green stars and blue circles, respectively.

abundance of the deglycosylated peptide at m/z 1,225.1926 (3+) increased over time until 12 h was reached, in sharp contrast to the relative abundance of the phosphopeptide which gradually decreased. The latter might result from the hydrolysis of the phosphopeptide (**Figure 2D**). Considering both the phosphopeptide hydrolysis degree and the glycopeptide deglycosylation efficiency, 12 h was chosen as further on-line deglycosylation time.

Optimization of the Elution Conditions for the Deglycosylated Peptides

After the on-line deglycosylation, the released deglycosylated peptides were re-adsorbed on TiO_2, which is line with the previous studies that non-modified peptides tend to be nonspecifically adsorbed on TiO_2 due to Lewis acid-base interaction between the carboxyl groups on the peptide chains and TiO_2 and the hydrophobic interaction between the hydrophobic peptide chains and TiO_2 (Palmisano et al., 2012b). To efficiently elute and collect the absorbed deglycosylated peptides but not the phosphopeptides, the elution efficiencies of three types of eluents were evaluated. The three types of eluents were NH_4HCO_3 solutions with

different concentrations (5, 10, and 20 mM), 50% ACN/1% FA, and 40% ACN/5% TFA. As shown in **Supplementary Figure 1**, 5 and 10 mM NH_4HCO_3 could not elute any PTM-peptides, while 20 mM NH_4HCO_3 could co-elute the deglycosylated peptides and the phosphopeptides. Therefore, NH_4HCO_3 was not suitable for the deglycosylated peptides elution. As to the acidic conditions, after the elution with 50% ACN/1% FA, the deglycosylated peptides were rarely observed (**Supplementary Figure 2**), while all the targeted deglycosylated peptides but not the phosphopeptides could be detected in the eluate of 40% ACN/5% TFA (**Figure 3B**). Afterward, the bound phosphopeptides on TiO_2 were found from the eluate of 5% (v/v) $NH_3 \cdot H_2O$ (**Figure 3C**), which was consistent with the phosphopeptides obtained by using the reported simultaneous enrichment, the co-elution, and the deglycosylation method (**Figure 3A**). Thus, 40% ACN/5% TFA was used as the eluent of the deglycosylated peptides.

Analysis of Glyco- and Phosphopeptides from the Mouse Brain

To examine the effectiveness of our strategy, we applied it to analyze glyco- and phosphopeptides from a 100 μg mouse
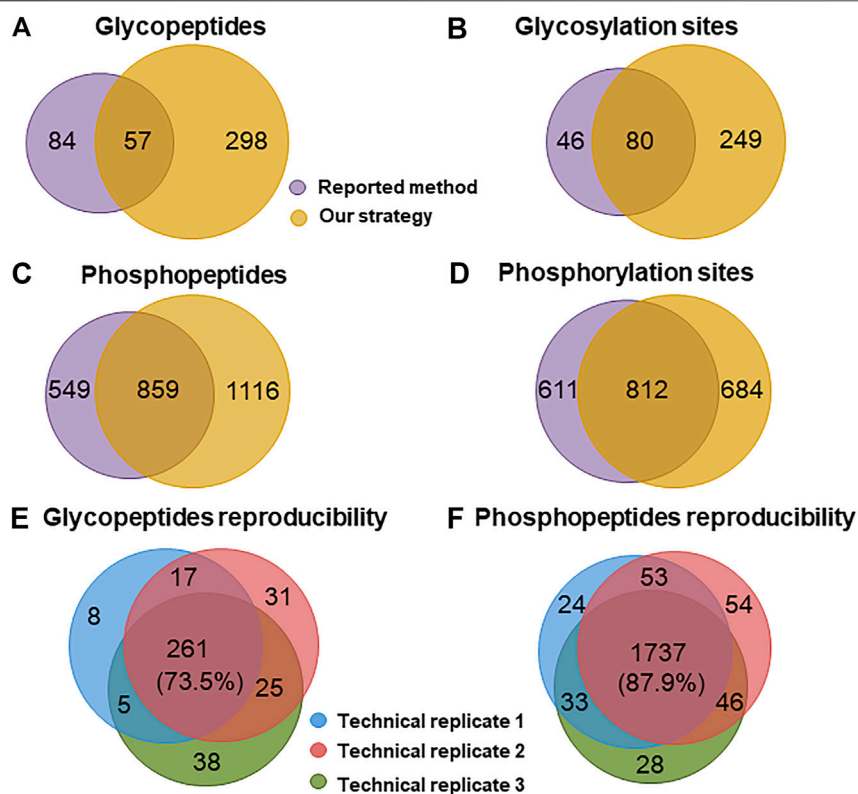
FIGURE 4 | Comparison of the performance for identifying glyco- and phosphopeptides from a mouse brain. Venn diagram analysis of the number of **(A)** glycopeptides, **(B)** glycosylation sites, **(C)** phosphopeptides, and **(D)** phosphorylation sites identified with the reported method and our strategy. Reproducibility for **(E)** glyco- and **(F)** phosphopeptides among three technical replicates of our strategy.

brain. Meanwhile, the reported method (Palmisano et al., 2012b) was carried out for comparison. By using our strategy, a total of 329 glycosylation sites (**Supplementary Table 1**) from 355 glycopeptides (**Supplementary Table 2**) and 1,496 phosphorylation sites (**Supplementary Table 3**) from 1975 phosphopeptides (**Supplementary Table 4**) were identified from three technical replicates. In sharp contrast to these, only 126 glycosylation sites (**Supplementary Table 5**) from 141 glycopeptides (**Supplementary Table 6**) and 1,423 phosphorylation sites (**Supplementary Table 7**) from 1,408 phosphopeptides (**Supplementary Table 8**) were identified from the identical sample using the reported method (**Figures 4A–D**). The numbers of the identified glyco- and phosphopeptides with our strategy were 2.5 and 1.4 folds of those with the reported method, respectively. Moreover, our strategy demonstrated high reproducibility with 73.5 and 87.9% common glyco- and phosphopeptides among three technical replicates (**Figures 4E,F**), respectively. The reproducibility between two technical replicates was much higher, 77.0 ± 2.2% and 91.8 ± 0.3% for glyco- and phosphopeptides, respectively. Besides the high reproducibility, the numbers of the phosphopeptides in the deglycosylated glycopeptide fraction and the phosphopeptide fraction were 67 and 2,298, respectively, suggesting a low degree of overlap

between the deglycosylated glycopeptide and the phosphopeptide fraction in our strategy.

The property differences of the identified PTM-peptides between the reported method and our strategy were further investigated. The molecular weight (Mw) distribution of the identified PTM-peptides and the percentages of the peptides with single-PTM and multiple PTM sites were compared (**Figure 5**).

As shown in **Figure 5A**, the Mw distribution pattern of the identified glycopeptides is consistent between the reported method and our strategy in a higher Mw range (1,500–5,000 Da) is consistent between the reported method and our strategy. However, in the lower Mw range of 500–1,000 Da and 1,000–1,500 Da, the number of the identified glycopeptides with our strategy accounts for 5.1 and 24.5% of the total ones, respectively, in sharp contrast to that of 0 and 7.1% with the reported method. These results indicate that our strategy has advantages in the enrichment and identification of low Mw glycopeptides. This result might be ascribed to the fact that our strategy omits the desalting procedure after the deglycosylation and retains well the low Mw glycopeptides.

As to the percentages of the peptides with single-PTM and multiple PTM sites, the number of identified glycopeptides with two glycosylation sites with our strategy accounts for 24% of the total glycopeptides, compared with only 14% with the reported
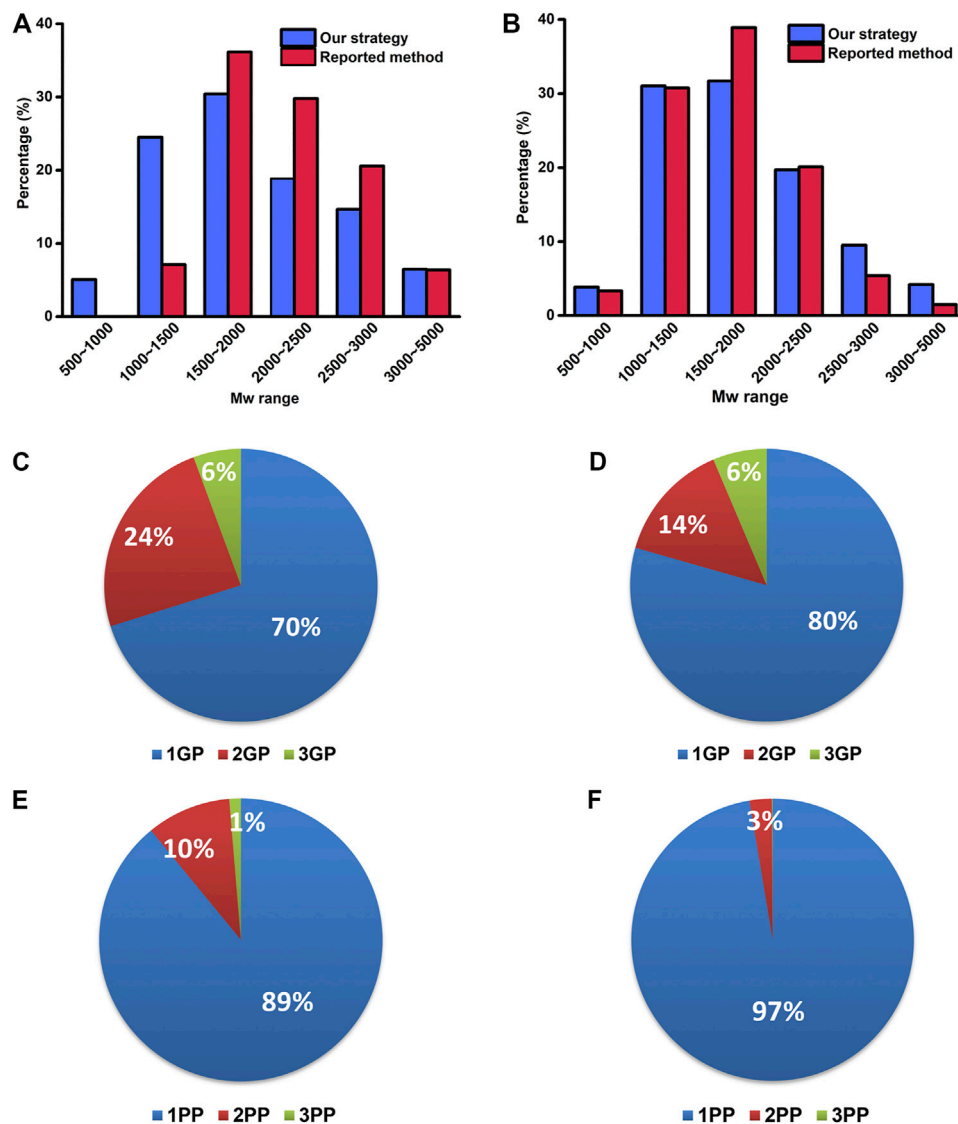
FIGURE 5 | The Mw distribution of identified glyco- **(A)** and phosphopeptides **(B)** by our strategy and the reported method. The percentages of the identified PTM-peptides with single glycosylation/phosphorylation site and multiple glycosylation/phosphorylation sites by our strategy **(C, E)** and the reported method **(D, F)**. 1–3 GP: glycopeptide with 1–3 glycosylation sites; 1–3 PP: phosphopeptide with 1–3 phosphorylation sites.

method (**Figures 5C,D**). Similarly, the number of identified phosphopeptides with two phosphorylation sites with our strategy accounts for 10% of the total phosphopeptides, in sharp contrast to that of 3% with the reported method (**Figures 5E,F**). These results revealed the superiority of our strategy in the identification of the peptides with multiple glycosylation/phosphorylation sites. It is possibly because our strategy reduced the complexity of the samples and increased the number of the identified low-abundance peptides with multiple glycosylation/phosphorylation sites.

The above results indicated that our strategy not only realized the sequential elution of glyco- and phosphopeptides but also significantly increased the numbers of identified glyco- and

phosphopeptides. Our study provided an effective means for the simultaneous characterization of the protein glycosylation and phosphorylation.

CONCLUSION

In this work, we developed a strategy for the analysis of glyco- and phosphopeptides based on the simultaneous enrichment with TiO_2, the on-line deglycosylation, and the sequential elution. The application of this strategy to the mouse brain tissue achieved a higher number of targeted peptides compared with the reported method. Our strategy shows some

advantages in the simultaneous analysis of glyco- and phosphopeptides: 1. The on-line deglycosylation and the sequential elution can separate the deglycosylated peptides and the phosphopeptides into two different fractions, which can reduce the complexity of the samples and improve the coverage of the identified PTM-peptides under the data dependent acquisition (DDA) mode. 2. The reduction of the sample complexity can reduce the ion suppression and increase the number of low-abundance glyco- and phosphopeptides with multiple glycosylation and phosphorylation sites. 3. The elimination of the desalting procedure after the deglycosylation can reduce the loss of low Mw glycopeptides. 4. A neutral online deglycosylation condition can effectively inhibit the hydrolysis of the phosphopeptides and increase the number of identified phosphopeptides.

To sum up, this work will provide a new idea to expand the applications of TiO_2 and tackle the problems in the simultaneous analysis of the protein glycosylation and other multiple PTMs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

REFERENCES

Cheison, S. C., Leeb, E., Letzel, T., and Kulozik, U. (2011). Influence of Buffer Type and Concentration on the Peptide Composition of Trypsin Hydrolysates of β-lactoglobulin. *Food Chem.* 125, 121–127. doi:10.1016/j.foodchem.2010.08.047

Cho, K.-C., Chen, L., Hu, Y., Schnaubelt, M., and Zhang, H. (2019). Developing Workflow for Simultaneous Analyses of Phosphopeptides and Glycopeptides. *ACS Chem. Biol.* 14, 58–66. doi:10.1021/acschembio.8b00902

Deeb, S. J., Cox, J., Schmidt-Supprian, M., and Mann, M. (2014). N-linked Glycosylation Enrichment for In-Depth Cell Surface Proteomics of Diffuse Large B-Cell Lymphoma Subtypes. *Mol. Cell Proteomics* 13, 240–251. doi:10.1074/mcp.M113.033977

Hart, G. W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. (2011). Cross Talk between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* 80, 825–858. doi:10.1146/annurev-biochem-060608-102511

Hu, Y., Jiang, B., Weng, Y., Sui, Z., Zhao, B., Chen, Y., et al. (2020). Bis(Zinc(II)-dipicolylamine)-functionalized Sub-2 μm Core-Shell Microspheres for the Analysis of N-Phosphoproteome. *Nat. Commun.* 11, 6226. doi:10.1038/s41467-020-20026-1

Hu, Y., Shah, P., Clark, D. J., Ao, M., and Zhang, H. (2018). Reanalysis of Global Proteomic and Phosphoproteomic Data Identified a Large Number of Glycopeptides. *Anal. Chem.* 90, 8065–8071. doi:10.1021/acs.analchem.8b01137

Hussain, S., Pezzei, C., Güzel, Y., Rainer, M., Huck, C. W., and Bonn, G. K. (2014). Zirconium Silicate Assisted Removal of Residual Proteins after Organic Solvent Deproteinization of Human Plasma, Enhancing the Stability of the LC-ESI-MS Response for the Bioanalysis of Small Molecules. *Analytica Chim. Acta* 852, 284–292. doi:10.1016/j.aca.2014.09.014

Liu, F., Iqbal, K., Grundke-Iqbal, I., and Gong, C.-X. (2002). Involvement of Aberrant Glycosylation in Phosphorylation of Tau by Cdk5 and GSK-3β. *FEBS Lett.* 530, 209–214. doi:10.1016/s0014-5793(02)03487-7

Liu, Y., Salas-Solano, O., and Gennaro, L. A. (2009). Investigation of Sample Preparation Artifacts Formed during the Enzymatic Release of N-Linked Glycans Prior to Analysis by Capillary Electrophoresis. *Anal. Chem.* 81, 6823–6829. doi:10.1021/ac9010588

Losev, Y., Frenkel-Pinter, M., Abu-Hussien, M., Viswanathan, G. K., Elyashiv-Revivo, D., Geries, R., et al. (2021). Differential Effects of Putative N-Glycosylation Sites in Human Tau on Alzheimer's Disease-Related Neurodegeneration. *Cell. Mol. Life Sci.* 78, 2231–2245. doi:10.1007/s00018-020-03643-3

Lu, Q., Chen, C., Xiong, Y., Li, G., Zhang, X., Zhang, Y., et al. (2020). High-efficiency Phosphopeptide and Glycopeptide Simultaneous Enrichment by Hydrogen Bond-Based Bifunctional Smart Polymer. *Anal. Chem.* 92, 6269–6277. doi:10.1021/acs.analchem.9b02643

Ma, X., Li, H., He, Y., and Hao, J. (2017). The Emerging Link between O-GlcNAcylation and Neurological Disorders. *Cel. Mol. Life Sci.* 74, 3667–3686. doi:10.1007/s00018-017-2542-9

Melo-Braga, M. N., Ibáñez-Vea, M., Larsen, M. R., and Kulej, K. (2015). Comprehensive Protocol to Simultaneously Study Protein Phosphorylation, Acetylation, and N-Linked Sialylated Glycosylation. *Methods Mol. Biol.* 1295, 275–292. doi:10.1007/978-1-4939-2550-6_21

Melo-Braga, M. N., Schulz, M., Liu, Q., Swistowski, A., Palmisano, G., Engholm-Keller, K., et al. (2014). Comprehensive Quantitative Comparison of the Membrane Proteome, Phosphoproteome, and Sialiome of Human Embryonic and Neural Stem Cells. *Mol. Cell Proteomics* 13, 311–328. doi:10.1074/mcp.M112.026898

Palmisano, G., Jensen, S. S., Le Bihan, M.-C., Lainé, J., McGuire, J. N., Pociot, F., et al. (2012a). Characterization of Membrane-Shed Microvesicles from Cytokine-Stimulated β-Cells Using Proteomics Strategies. *Mol. Cell Proteomics* 11, 230–243. doi:10.1074/mcp.M111.01273

Palmisano, G., Parker, B. L., Engholm-Keller, K., Lendal, S. E., Kulej, K., Schulz, M., et al. (2012b). A Novel Method for the Simultaneous Enrichment, Identification, and Quantification of Phosphopeptides and Sialylated

ETHICS STATEMENT

The animal study was reviewed and approved by the Biological Research Ethics Committee of Dalian Medical University. A written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

CC and XZ carried out the experiments and wrote the manuscript. XD participated in the optimization of the enrichment methods. HZ contributed to the manuscript revision. XLL and XML contributed to the study design and the manuscript revision.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.703176/full#supplementary-material

Glycopeptides Applied to a Temporal Profile of Mouse Brain Development. *Mol. Cell Proteomics* 11, 1191–1202. doi:10.1074/mcp.M112.017509

Peng, Y., Le, Z., Wen, M., Zhang, D., Chen, Z., Wu, H. B., et al. (2017). Mesoporous single-crystal-like TiO2 Mesocages Threaded with Carbon Nanotubes for High-Performance Electrochemical Energy Storage. *Nano Energy* 35, 44–51. doi:10.1016/j.nanoen.2017.03.003

Sheng, Q., Li, X., Yin, W., Yu, L., Ke, Y., and Liang, X. (2013). Retention Mechanism and Enrichment of Glycopeptides on Titanium Dioxide. *Anal. Methods* 5, 7072–7080. doi:10.1039/c3ay41294f

Sun, N., Wang, J., Yao, J., Chen, H., and Deng, C. (2019). Magnetite Nanoparticles Coated with Mercaptosuccinic Acid-Modified Mesoporous Titania as a Hydrophilic Sorbent for Glycopeptides and Phosphopeptides Prior to Their Quantitation by LC-MS/MS. *Microchim Acta* 186, 159. doi:10.1007/s00604-019-3274-3

Takeda, S., Wegmann, S., Cho, H., Devos, S. L., Commins, C., Roe, A. D., et al. (2015). Neuronal Uptake and Propagation of a Rare Phosphorylated High-Molecular-Weight Tau Derived from Alzheimer's Disease Brain. *Nat. Commun.* 6, 8490. doi:10.1038/ncomms9490

Thompson, A. J., Hart, S. R., Franz, C., Barnouin, K., Ridley, A., and Cramer, R. (2003). Characterization of Protein Phosphorylation by Mass Spectrometry Using Immobilized Metal Ion Affinity Chromatography with On-Resin β-Elimination and Michael Addition. *Anal. Chem.* 75, 3232–3243. doi:10.1021/ac034134h

Wang, Z., Wang, J., Sun, N., and Deng, C. (2019). A Promising Nanoprobe Based on Hydrophilic Interaction Liquid Chromatography and Immobilized Metal Affinity Chromatography for Capture of Glycopeptides and Phosphopeptides. *Analytica Chim. Acta* 1067, 1–10. doi:10.1016/j.aca.2019.04.010

Xu, D., Gao, M., Deng, C., and Zhang, X. (2016). Synthesis of Bifunctional TiO2@SiO2-B(OH)2@Fe3O4@TiO2 sandwich-like Nanosheets for Sequential Selective Enrichment of Phosphopeptides and Glycopeptides for Mass Spectrometric Analysis. *Anal. Bioanal. Chem.* 408, 5489–5497. doi:10.1007/s00216-016-9647-0

Xu, D., Yan, G., Gao, M., Deng, C., and Zhang, X. (2017). Highly Selective SiO2-NH2@TiO2 Hollow Microspheres for Simultaneous Enrichment of Phosphopeptides and Glycopeptides. *Anal. Bioanal. Chem.* 409, 1607–1614. doi:10.1007/s00216-016-0101-0

Yan, Y., and Deng, C. (2019). Recent Advances in Nanomaterials for Sample Pre-treatment in Phosphoproteomics Research. *Trends Anal. Chem.* 120, 115655. doi:10.1016/j.trac.2019.115655

Zhang, X., Qiao, Y., Wu, Q., Chen, Y., Zou, S., Liu, X., et al. (2017). The Essential Role of YAP O-GlcNAcylation in High-Glucose-Stimulated Liver Tumorigenesis. *Nat. Commun.* 8, 15280. doi:10.1038/ncomms15280

Zou, X., Jie, J., and Yang, B. (2017). Single-Step Enrichment of N-Glycopeptides and Phosphopeptides with Novel Multifunctional Ti4+-Immobilized Dendritic Polyglycerol Coated Chitosan Nanomaterials. *Anal. Chem.* 89, 7520–7526. doi:10.1021/acs.analchem.7b01209

Deciphering the O-Glycosylation of HKU1 Spike Protein With the Dual-Functional Hydrophilic Interaction Chromatography Materials

Yun Cui[1†], Xuefang Dong[2†], Xiaofei Zhang[2], Cheng Chen[2], Dongmei Fu[1*], Xiuling Li[2*] and Xinmiao Liang[2]

[1]School of Biological Engineering, Dalian Polytechnic University, Dalian, China, [2]Key Lab of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

HKU1 is a human beta coronavirus and infects host cells *via* highly glycosylated spike protein (S). The N-glycosylation of HKU1 S has been reported. However, little is known about its O-glycosylation, which hinders the in-depth understanding of its biological functions. Herein, a comprehensive study of O-glycosylation of HKU1 S was carried out based on dual-functional histidine-bonded silica (HBS) materials. The enrichment method for O-glycopeptides with HBS was developed and validated using standard proteins. The application of the developed method to the HKU1 S1 subunit resulted in 46 novel O-glycosylation sites, among which 55.6% were predicted to be exposed on the outer protein surface. Moreover, the O-linked glycans and their abundance on each HKU1 S1 site were analyzed. The obtained O-glycosylation dataset will provide valuable insights into the structure of HKU1 S.

Keywords: HKU1, spike glycoprotein, enrichment, O-glycosylation sites, O-glycosylation abundance

INTRODUCTION

The human HKU1 coronavirus (CoV) was first discovered in Hong Kong in 2004 and found to cause prevalent respiratory diseases (Woo et al., 2005). HKU1 is a kind of beta coronavirus (β-CoV), which includes the severe acute respiratory syndrome (SARS-CoV), Middle East respiratory syndrome (MERS-CoV), and SARS-CoV-2 (Christian et al., 2004; Woo et al., 2009; Zaki et al., 2012; Hoffmann et al., 2020). The CoV spike (S) protein is a large type I transmembrane glycoprotein, and it mediates virus entry to the host cells (Heald-Sargent and Gallagher, 2012). The S protein has two subunits: the S1 subunit is responsible for receptor binding, whereas the S2 subunit facilitates membrane fusion (Millet and Whittaker, 2015; Kirchdoerfer et al., 2016). Specifically, S1 contains two independent domains: an amino (N)-terminal domain (NTD) and a carboxy (C)-terminal domain (CTD) (Peng et al., 2011). Several β-CoVs, including mouse hepatitis virus, human CoV OC43, and bovine CoV (BCoV), use their NTDs to bind receptor protein (Peng et al., 2011; Peng et al., 2012). By contrast, HKU1 uses its CTD to bind to receptors (Qian et al., 2015), similar to SARS-CoV, MERS-CoV, and SARS-CoV-2 (Li et al., 2003; Mou et al., 2013; Hoffmann et al., 2020). Glycosylation contributes significantly to the conformation of the S protein and therefore profoundly affects receptor binding (Fung and Liu, 2018). The S protein of HKU1 is highly N-glycosylated, and 29 N-glycosylation sites have been deciphered (Watanabe et al., 2020). The glycan shield density of the HKU1 S protein is considerably higher than that of SARS-CoV and MERS-CoV (Watanabe et al., 2020).

Except for N-glycosylation, viral O-glycosylation plays pivotal roles in viral entry, propagation, and immune recognition (Bagdonaite et al., 2015; Iversen et al., 2016; Olofsson et al., 2016; Stone et al., 2016). The O-glycosylation of viral surface proteins on human cytomegalovirus, Epstein–Barr virus (Bagdonaite et al., 2016), and hepatitis C virus (Brautigam et al., 2012) have been extensively reported. More importantly, the O-linked glycans on human immunodeficiency virus type 1 can shield against one category of broadly neutralizing antibodies (Silver et al., 2020). In addition, the O-glycosylation on viral glycoproteins can be developed as the epitopes for the potential development of subunit vaccines (Olofsson et al., 2016). However, the O-glycosylation of HKU1 is scarcely reported, and the related virology research on O-glycosylation of S protein is severely hindered.

The identification of O-glycosylation is more challenging than that of N-glycosylation, owing to the lack of conserved O-glycosylation site sequon and consistent O-linked glycan cores, inefficient O-linked glycan-specific glycosidases, and extremely low O-glycosylation stoichiometry (You et al., 2018). Thus far, hydrazide chemistry (Yang et al., 2017), "SimpleCell" method (Steentoft et al., 2011), and lectin affinity chromatography methods (Anan et al., 2019) have been adopted to enrich O-glycopeptides. However, hydrazide chemistry method always destroys the intact glycan structure during the oxidation step. The "SimpleCell" technology blocks the natural elongation of O-linked glycans but eliminates their heterogeneity (Steentoft et al., 2011). The lectin affinity chromatography method can only enrich individual O-linked glycan and lacks universality (Narimatsu et al., 2019; Singh et al., 2020).

Hydrophilic interaction liquid chromatography (HILIC) has been widely adopted to enrich N-glycopeptides with no bias to glycan structures (Cao et al., 2014; Hoffmann et al., 2016; Shao et al., 2016; You et al., 2018; Qing et al., 2020). In our previous work (Dong et al., 2017), dual-functional histidine-bonded silica (HBS) HILIC materials were prepared, and they demonstrated the selective enrichment of N-glycopeptides from human serum (Dong et al., 2017; Qin et al., 2019). Thus, it was expected that HBS materials can be applied for the enrichment of O-glycopeptides from HKU1 S. To achieve this goal, we first developed the enrichment method of O-glycopeptides based on HBS by optimizing different enrichment conditions with bovine fetuin as the model glycoprotein. This newly developed method was further validated by enriching O-glycopeptides from a mixture of bovine fetuin and albumin bovine serum digests, and commercial ZIC-HILIC materials were used for comparison. Finally, O-glycosylation of HKU1 S1 was comprehensively characterized including the O-glycosylation site identification, glycosylation site distribution, exposure ratio prediction, and O-linked glycan analysis. We believe that deciphering O-glycosylation will provide a significant complement to glycosylation for HKU1 S.

MATERIALS AND METHODS

Reagents and Materials
Bovine fetuin, albumin bovine serum (BSA), trypsin, elastase, and chemical reagents of iodoacetamide (IAA), 1,4-dithiothreitol (DTT), acetic acid (HAc), ammonium bicarbonate (NH_4HCO_3), ammonia water ($NH_3 \cdot H_2O$), urea, and the zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) materials were obtained from Sigma (St. Louis, MO). HKU S1 (expressed from HEK293 cells, the purity > 93.2%) was purchased from Sino Biological. PNGase F was purchased from New England Biolabs (Ipswich, MA). Acetonitrile (ACN, HPLC grade) was from Merck (Darmstadt, Germany). Formic acid (FA) was obtained from Honeywell (Shanghai, China). Trifluoroacetic acid (TFA) was obtained from Macklin (Shanghai, China). Pure water used in all experiments was purified with a Milli-Q system (Millipore, Milford, MA). GELoader tips were purchased from Eppendorf (Hamburg, Germany). C18 AQ materials were obtained from Acchrom (Beijing, China). Histidine-bonded silica (HBS) materials were homemade.

Digestion of Proteins
Each protein (BSA, bovine fetuin, and HKU1 S1) of 1 mg was denatured with 100 μl 6 M urea in 50 mM NH_4HCO_3 for 3 h, and then 20 μl DTT (200 mM) was added for reduction at 56°C for 45 min. After adding 40 μl IAA (200 mM) in dark for 30 min, fivefold volume of 50 mM NH_4HCO_3 was added to the solution and then mixed with different enzymes. Trypsin was added for BSA digestion at an enzyme/protein mass ratio of 1:25 (w/w). Bovine fetuin was first digested by elastase at an enzyme/protein ratio of 1:40 (w/w), and then PNGase F was added at 37°C overnight to remove N-glycans. HKU1 S1 was first digested by trypsin and chymotrypsin with the enzyme/protein ratio of 1:20 (w/w), and then PNGase F was used to remove N-glycans at 37°C overnight. Finally, the protein digests were collected and lyophilized to dryness.

Enrichment of O-Glycopeptides From Protein Digests
Optimization of HBS-Based Enrichment Conditions for O-Glycopeptides
Although the HBS-based enrichment method for N-glycopepitdes has been established, the strategy for O-glycopeptide enrichment has not been developed. Thus, we investigated the effect of different ACN contents (from 50 to 80%, v/v), pH value (adjusting by FA or NH_4HCO_3), and types of acid additive (FA, HAc, and TFA) on the O-glycopeptide enrichment efficiency. Bovine fetuin was selected as model protein, and the typical O-glycopeptides with m/z 1300.2193 (3+) and 1440.0979 (2+), and non-glycosylated peptides with m/z 1122.5524 (1+) and 1213.5851 (1+) were selected to evaluate the enrichment performance. The detailed information of the peptide sequence and the glycan structure of the typical O-glycopeptides is shown in **Supplementary Table S1**, signed with green.

Enrichment of O-Glycopeptides From Bovine Fetuin Digests With Optimized Method
One milligram of HBS materials suspended in 20 μl of ACN was packed into a GELoader tip. The tip was activated with 50% ACN/ 0.1% TFA (30 μl × 3) and equilibrated with 80% ACN/0.1% TFA

(30 μl × 3) successively. Then, 10 μg bovine fetuin digests in 80% ACN/0.1% TFA was loaded on the HBS materials and then washed with 20 μl of 80% ACN/0.1% TFA twice. Subsequently, the materials were eluted with 20 μl of 10% $NH_3 \cdot H_2O$, then the eluent was collected and dried to remove $NH_3 \cdot H_2O$, followed by redissolving in 20 μl of 50% ACN/0.1% FA for ESI Q-TOF MS analysis.

Enrichment of O-Glycopeptides From Mixed Standard Protein Digests

The digest mixture of fetuin and BSA at a mass ratio of 1:20, 1:200 (10 μg bovine fetuin) was mixed with 1 mg HBS in 500 μl and 2 mg HBS in 4 ml of 80% ACN/0.1% TFA, respectively. The obtained solution was shaken for 40 min, followed by centrifugation at 10,000 g for 2 min. Then, the supernatant was removed, and the precipitation was washed with 80% ACN/0.1% TFA (250 μl × 3 for ratio of 1:20, and 1 ml × 3 for ratio of 1:200, respectively) to remove the non-glycosylated peptides. Subsequently, the precipitation was transferred into a GELoader tip, respectively, and eluted with 30 μl of 10% $NH_3 \cdot H_2O$. The eluent was collected and dried to remove $NH_3 \cdot H_2O$, followed by redissolving in 30 μl of 50% ACN/0.1% FA for analysis by ESI Q-TOF MS.

In comparison, the enrichment with ZIC-HILIC was carried out as previously described (Huang et al., 2020), with relevant modification. The digest mixture of fetuin and BSA at a ratio of 1:20 (w/w) was mixed with 1 mg ZIC-HILIC in 500 μl of 80% ACN/0.2% TFA. The obtained solution was shaken for 10 min, followed by centrifugation at 10,000 g for 2 min. After removing the supernatant, the precipitation was washed with 80% ACN/0.2% TFA (250 μl × 3) to remove the non-glycosylated peptides. Then, the precipitation was transferred into a GELoader tip and eluted with 20 μl of 30% ACN/2% FA to obtain the O-glycopeptides.

Enrichment of O-Glycopeptides From HKU1 S1

Five milligrams of HBS was suspended in ACN and packed into a GELoader tip. The tip was activated with 50% ACN/0.1% TFA (30 μl × 3) and equilibrated with 80% ACN/0.1% TFA (30 μl × 3) successively. Then, 50 μg HKU1 S1 digests in 100 μl of 80% ACN/0.1% TFA was loaded on the HBS. The materials were washed with 80% ACN/0.1% TFA (30 μl × 3), and subsequently eluted with 40 μl of 10% $NH_3 \cdot H_2O$. The eluent was collected and dried for further liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

Mass Spectrometry Analysis

The enriched O-glycopeptides from bovine fetuin digests were analyzed on a nano-ESI-Q-TOF mass spectrometer (Waters, Manchester, United Kingdom) with collision-induced dissociation (CID) in a positive mode. Full scan MS data were obtained at m/z 600–1700.

The enriched bovine fetuin and HKU1 S1 O-glycopeptides were separated and characterized using Q-Exactive Orbitrap coupled with Accela 600 HPLC system (Thermo, CA, United States), respectively. For the separation of peptides with reverse-phase liquid chromatography, 0.1% FA (pH 2.59)

aqueous solution and ACN/0.1% FA were used as mobile phases A and B, respectively. The analytical column with an inner diameter of 75 μm was packed in-house with C18 AQ particles (3 μm, 120 Å) to 12 cm length. The flow rate was set at 600 nl/min. Gradient elution was performed with 2–8% B in 0.2 min, 8–50% B in 45 min, 50–90% B in 0.5 min, and 90% B in 5 min. Full mass scans were carried out on the Orbitrap with acquisition range from m/z 500 to 1500 (R = 70,000 at m/z 400). The 20 most intense ions from the full scan were selected for fragmentation *via* higher-energy collisional dissociation (HCD) in the ion trap. The dynamic exclusion function was set as follows: repeat count 1, repeat duration 30 s, and exclusion duration of 60 s.

Data Analysis

All the RAW data files obtained from Orbitrap were searched against the database, using Byonic software (version 3.6.0, Protein Metrics, Inc.). The mass tolerance for precursors and fragment ions was set at 10 and 20 ppm, respectively. The O-glycans database was composed with 15 common O-glycans according to Zhao et al. (2020): [HexNAc(1)Hex(1), HexNAc(1)Hex(1)Fuc(1), HexNAc(2)Hex(1), HexNAc(1)Hex(1)NeuAc(1), HexNAc(2) Hex(2), HexNAc(2)Hex(1)NeuAc(1), HexNAc(1)Hex(1) NeuAc(2), HexNAc(2)Hex(2)NeuAc(1), HexNAc(3)Hex(1) NeuAc(1), HexNAc(2)Hex(2)Fuc(1)NeuAc(1), HexNAc(2) Hex(2)NeuAc(2), HexNAc(2)Hex(1)Fuc(1), HexNAc(1)Hex(2), HexNAc(1) NeuAc(1), and HexNAc(1)Hex(2)NeuAc(1)]. The fixed modification was carbamidomethyl (C), and variable modifications included oxidation (M), acetyl (protein N-term), and deamidation (N). Trypsin and chymotrypsin were set as the specific proteolytic enzymes with up to two missed cleavages allowed. Peptides with charge states of 2, 3, and 4 were chosen for further fragmentation. The FDR were all set as <1%. Moreover, the data were searched against reverse and contaminant sequences.

RESULTS AND DISCUSSION

Optimization of HBS-Based Enrichment Conditions for O-Glycopeptides

In the HILIC mode, the content of organic concentration determines the elution strength of the solvent, which affects solute retention on the stationary phase (Buszewski and Noga, 2012). Herein, the effect of different ACN contents (50–80%) on the O-glycopeptide enrichment on HBS was investigated under the same pH condition (containing 1% FA). As shown in **Figure 1A**, the non-glycosylated peptides were reserved from 50 to 80% ACN fractions, and O-glycopeptides were reserved in the elution fraction only. The result demonstrated that the retention of peptides on HBS possesses notable characteristics of HILIC, and O-glycopeptides can be retained strongly on HBS.

HBS with dual-functional characteristics displays hydrophilicity and switchable surface charge at different pH (Dong et al., 2017). To investigate the effect of pH on the enrichment of O-glycopeptides, we evaluated acidic, neutral, and basic ACN/H_2O solutions. Compared with the acidic condition (**Figure 1A**), the co-elution of O-glycopeptides and
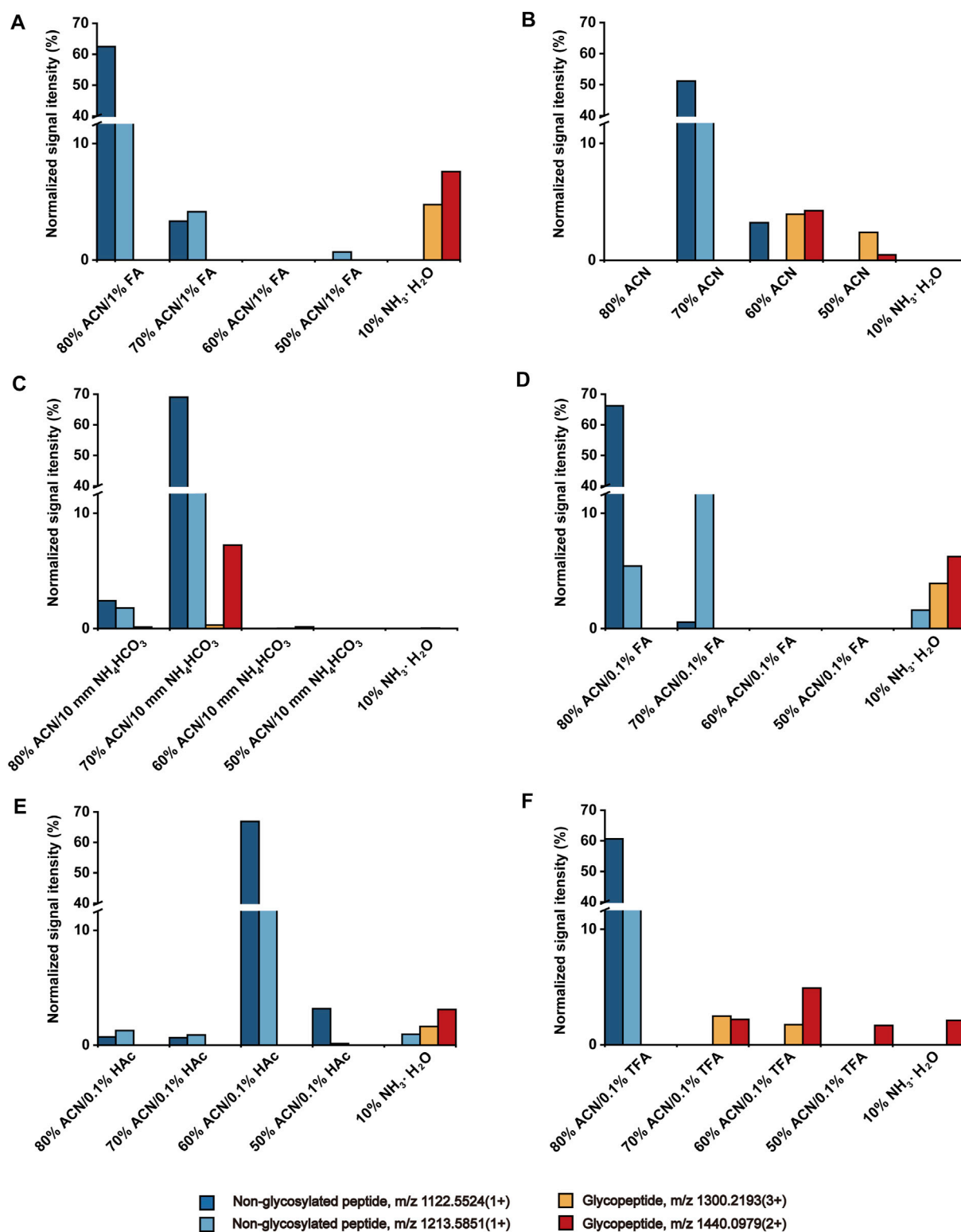
FIGURE 1 | Effect of ACN content **(A)**, solution pH **(B, C)**, and types of acid additive **(D–F)** to the retention of peptides on HBS materials.

non-glycosylated peptides was observed in neutral 60% ACN fraction (**Figure 1B**). When the solution was adjusted to a basic condition with 10 mM NH$_4$HCO$_3$ (**Figure 1C**), the

O-glycopeptide with m/z 1300.2193 (3+) was almost undetectable even after the elution of 10% NH$_3$·H$_2$O. We collected the loading effluent and used the HBS materials to
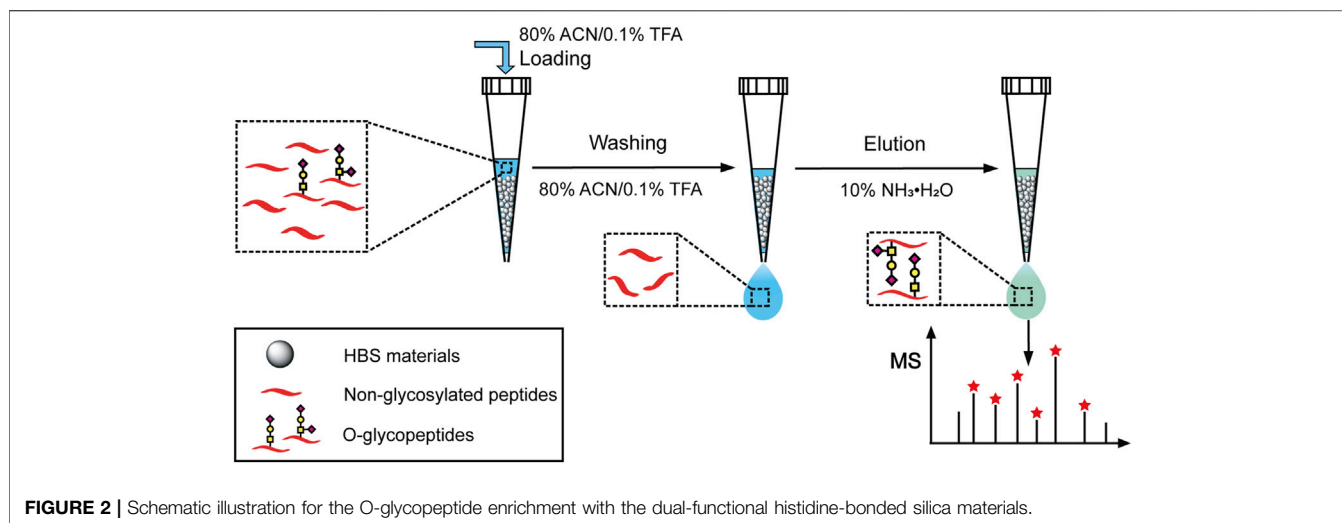
FIGURE 2 | Schematic illustration for the O-glycopeptide enrichment with the dual-functional histidine-bonded silica materials.

enrich the O-glycopeptides again. The O-glycopeptides with m/z 1300.2193 (3+) and 1440.0979 (2+) can be observed after the enrichment (**Supplementary Figure S1**). This finding demonstrated that the basic condition is unsuitable for O-glycopeptide enrichment with HBS materials because the positively charged O-glycopeptides are captured by the hydrophilic interaction of HBS materials under acidic condition, whereas HBS materials and O-glycopeptides are both negatively charged under basic condition, and O-glycopeptides cannot be captured on HBS. Therefore, an acidic solution is optimal for the O-glycopeptide enrichment on HBS.

Given that the acidic solution is favorable to the enrichment of O-glycopeptides on HBS, three types of acid additives (FA, HAc, and TFA) with 0.1% v/v to the solution were evaluated and compared. As shown in **Figures 1D,E**, non-glycosylated peptide m/z 1213.5851 (1+) was co-enriched with O-glycopeptides in the eluted solution, indicating that neither FA nor HAc is an optimal additive for O-glycopeptide enrichment. When the solution was added with TFA (**Figure 1F**), the non-glycosylated peptides flowed out to the 80% ACN fraction completely. In addition, O-glycopeptides occurred in the following fractions without any co-enrichment of non-glycosylated peptides. Although 1% FA as an acid additive facilitated the enrichment of O-glycopeptides significantly (**Figure 1A**), the efficiency of the removal of non-glycosylated peptides was greater with 0.1% TFA addition. Consequently, 0.1% TFA was selected as the acid additive for the following study.

Enrichment of O-Glycopeptides From Bovine Fetuin Digests

Bovine fetuin, a glycoprotein containing sialylated N-linked and O-linked glycans, was used to evaluate the specificity and selectivity of HBS materials for O-glycopeptide enrichment. Based on the above optimized conditions, a process for O-glycopeptide enrichment was developed (**Figure 2**). The bovine fetuin digests in 80% ACN/0.1% TFA were loaded onto

HBS, washed twice with 80% ACN/0.1% TFA to remove the non-glycosylated peptides, and then eluted with 10% $NH_3 \cdot H_2O$. With this optimized method, 32 O-glycopeptides were identified from the bovine fetuin digests (**Supplementary Figure S2A**). **Supplementary Table S1** shows the details of these enriched O-glycopeptides. Further investigation was carried out with the digest mixture of bovine fetuin and BSA at different mass ratios to evaluate the enrichment selectivity of HBS to O-glycopeptides. No O-glycopeptide signal was observed from the desalted digest mixture at a ratio of 1:20 (w/w) without any enrichment (**Supplementary Figure S2B**). By comparison, 28 O-glycopeptides were identified after HBS enrichment from the same ratio of the digest mixture (**Figure 3A**). Commercial ZIC-HILIC was also used for the enrichment of O-glycopeptides, and 13 O-glycopeptides were detected from the same ratio of 1:20 (w/w) after enrichment (**Figure 3B**). The high ratio of the digest mixture was further investigated for the enrichment on HBS, and at the mass ratio of 1:200 (w/w), the HBS materials still showed a high selectivity, with 24 O-glycopeptides identified (**Figure 3C**). These results demonstrated that HBS materials have outstanding anti-interferential abilities, good selectivity, and specificity to O-glycopeptides.

Validation of the Enrichment Method

In addition to selectivity, the reproducibility, recovery, limit of detection (LOD), and adsorption capacity are important parameters required to assess the developed enrichment method. The reproducibility of the optimized method was evaluated with bovine fetuin. The number of enriched O-glycopeptides was 31, 32, and 32 for three replicates. The recovery was measured by using the stable-isotope dimethyl labeling method (Boersema et al., 2009), and the recovery of typical two O-glycopeptides from bovine fetuin was over 93.9% (**Supplementary Table S2**), higher than that of ZIC-HLIC 84.3% (**Supplementary Table S3**). Given the lack of a standard O-glycopeptide, a standard sialylated glycopeptide (m/z 1433.2025) was used to test the LOD (S/N = 3), which reached 6.88 fmol/μl (**Supplementary Figure S3**). In addition,
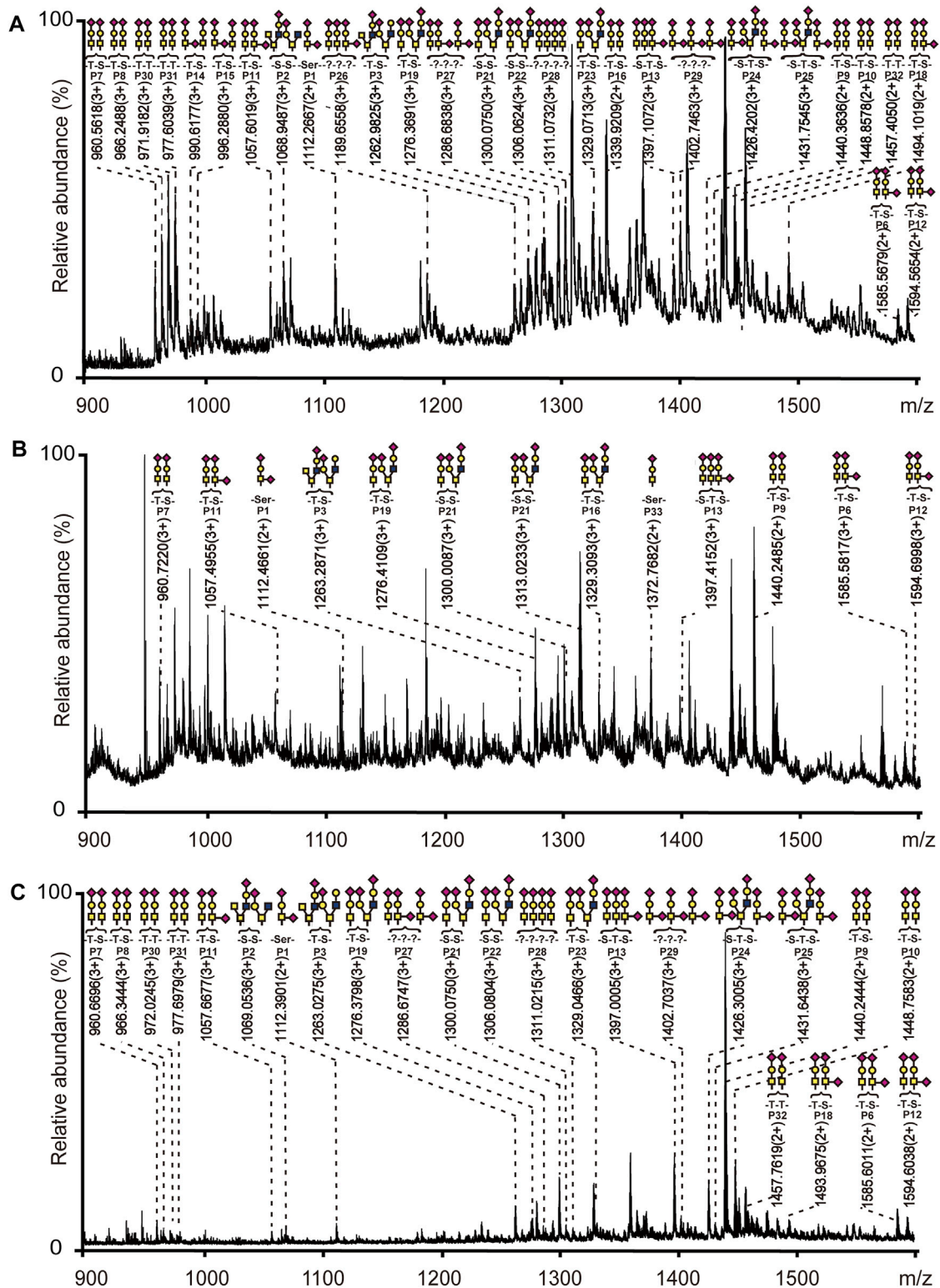
FIGURE 3 | Mass spectra of digest mixture after enrichment with HBS or ZIC-HILIC materials. **(A)** Bovine fetuin and BSA at a mass ratio of 1:20 enriched with HBS. **(B)** Bovine fetuin and BSA at a mass ratio of 1:20 enriched with ZIC-HILIC. **(C)** Bovine fetuin and BSA at a mass ratio of 1:200 enriched with HBS. Glycopeptides are marked with their glycan structures: ☐, GalNAc; ☐, GlcNAc; ◯, galactose; and ◆, N-acetylneuraminic acid (Neu5Ac). The detailed information about the peptide sequences and glycosylation is shown in **Supplementary Table S1**.

FIGURE 4 | Schematic workflow for deciphering the O-glycosylation of HKU1 S1 with the dual-functional hydrophilic interaction chromatography materials. The PDB code is 5I08 for the structure of HKU1 S1.
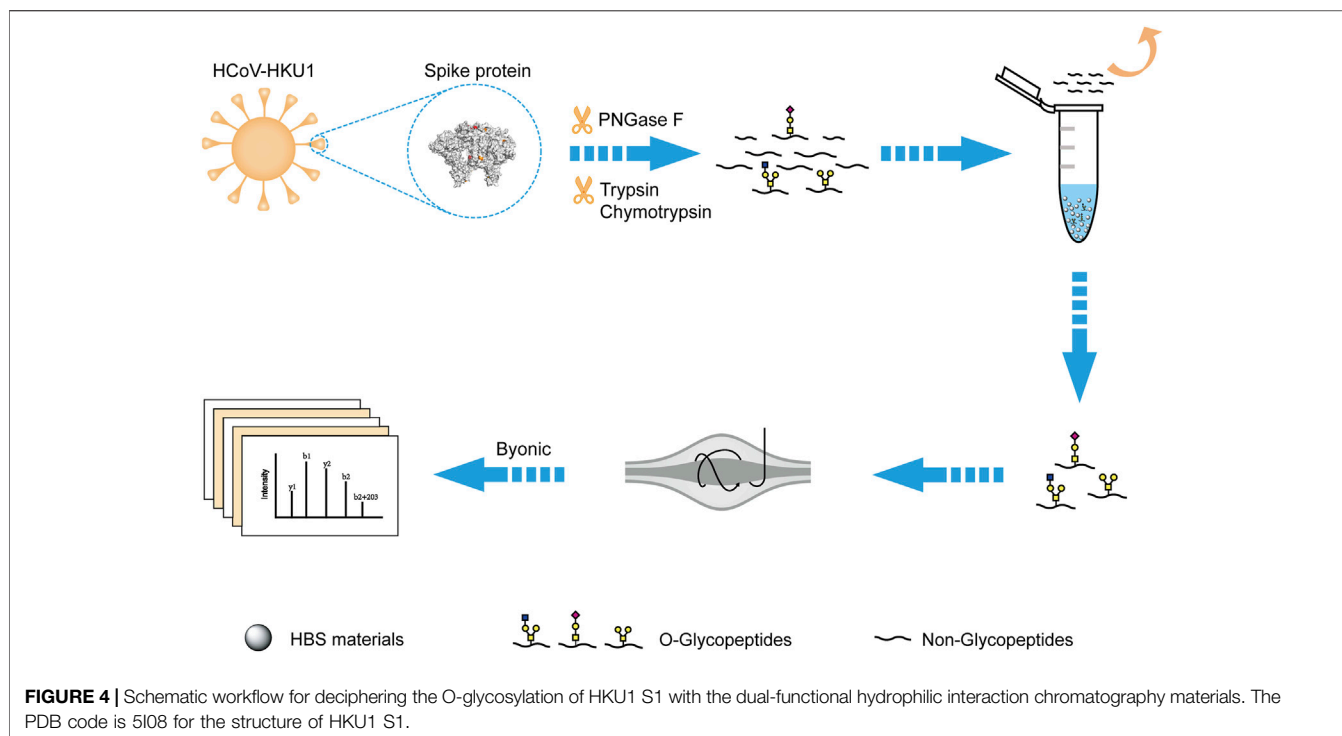
the adsorption capacity of HBS for bovine fetuin was 201.6 mg/g (**Supplementary Figure S4**). Thus, the optimized method in this study can be applied for the O-glycopeptide enrichment in complex samples.

Comprehensive O-Glycosylation Analysis of HKU1 S Protein

Novel Strategy for Deciphering the O-Glycosylation of HKU1 S

Inspired by the O-glycopeptide enrichment efficiency on HBS, we developed a novel strategy for deciphering the O-glycosylation of HKU1 S. As shown in **Figure 4**, the recombinant HKU1 S1 was digested by trypsin, chymotrypsin, and PNGase F successively for digestion into a peptide sample and removal of N-linked glycans. The O-glycopeptides can be captured by the dual-functional HBS materials in the acidic condition and then released in the basic condition. The enriched O-glycopeptides were analyzed by LC-MS/MS. Searching the acquired data against the Byonic provided the identification information for further analysis. The efficiency of PNGase F to remove N-glycans was validated with bovine fetuin digests, and the result is shown in **Supplementary Figure S5**. The N-glycopeptides were successfully removed after PNGase F digestion. The detailed information of the typical N-glycopeptides and de-Nglycan peptides of bovine fetuin is shown in **Supplementary Table S4**. HCD and electron transfer dissociation are supplementary fragmentation types of O-glycosylation characterization in MS/MS (Yang et al., 2019; Riley et al., 2020). However, given the instrument limitation, we only used HCD to fragment the

O-linked glycopeptides in this study. The stepped collision energy for HCD 20–30%, was set for sufficient fragmentation, and the MS/MS spectra were validated manually.

O-Glycosylation Site Identification and Distribution on HKU1 S

Based on the developed novel strategy, the study for the O-glycosylation of HKU1 S1 was carried out, and 46 O-glycosylation sites were identified (**Figure 5A**). Among the identified O-glycosylation sites, 18 were unambiguously identified. **Supplementary Figure S6** shows the corresponding LC-MS/MS b and y product ion fragments. All the O-glycosylation sites of HKU1 S1 were reported for the first time in this study. Compared with the 25 reported total O-glycosylation sites on SARS-CoV-2 S protein (Bagdonaite et al., 2021), the number of O-glycosylation sites on HKU1 S1 was higher. Furthermore, the distribution of O-glycosylation sites on two functional domains, namely, NTD and CTD, was investigated. CTD was reported as the receptor binding domain (RBD) of HKU1. A total of 14 and 22 O-glycosylation sites were distributed on NTD and RBD, respectively. These results showed that the O-glycosylation sites were not evenly but region-specifically distributed on HKU1 S1.

O-Glycosylation Analysis of HKU1 S1

The O-glycosylation site ratio of HKU1 S1, exposure degree of O-glycosylation on the HKU1 S1 outer surface, and exposure ratio on RBD were investigated. We calculated the O-glycosylation site ratio of HKU1 S1 by dividing the total number of amino acids by the number of O-glycosylation sites. The calculated O-glycosylation site ratio was 6.2% (46/
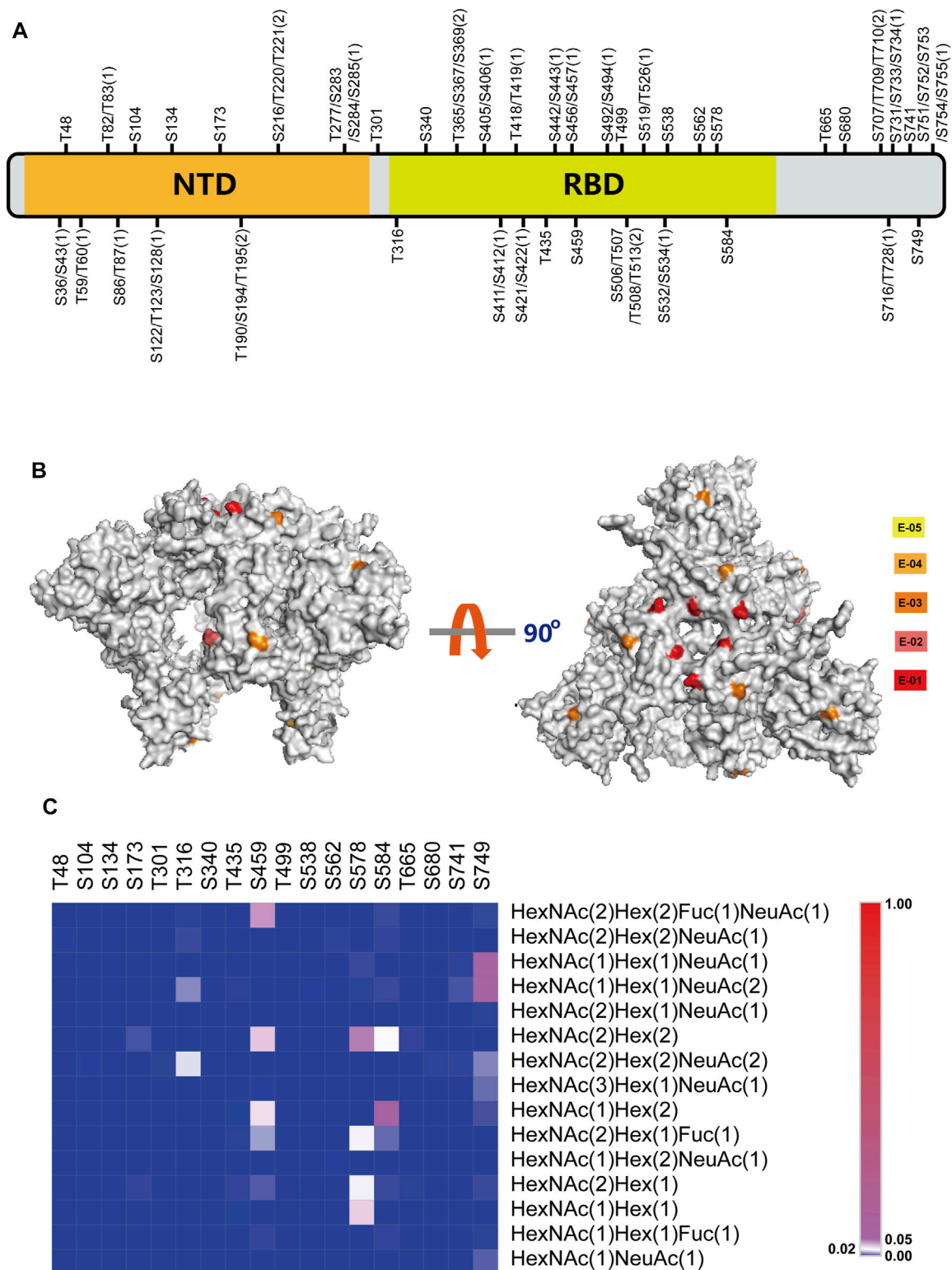
FIGURE 5 | Comprehensive O-glycosylation analysis of HKU1 S1 protein. **(A)** Schematic representation of O-glycosylation sites (OGSs) identified on HKU1 S1. The ambiguous identified OGSs are labeled with potential S/T with the possible number of OGSs in bracket. N-terminal domain (NTD) is labeled in orange, and the receptor-binding domain (RBD) is labeled in lime green. Detailed information about the peptide sequences and glycosylation is shown in **Supplementary Table S5**. **(B)** The normalized O-glycosylation abundance on unambiguous OGSs mapped on HKU1 S1 (PBD code: 5I08). The normalized O-glycosylation abundance was calculated by dividing the O-glycosylation abundance on each O-glycosylation site by the total O-glycosylation abundance of HKU1 S1. Five orders of magnitudes were selected to label the normalized sialylated O-glycosylation abundance with yellow (E-05), bright orange (E-04), orange (E-03), deep salmon (E-02), and red (E-01), respectively. **(C)** The distribution of site-specific O-linked glycans on the individual O-glycosylation site.

747), whereas the N-glycosylation site ratio was 3.8% (29/747). Based on the protein surface accessibility and secondary structure predictions (http://www.cbs.dtu.dk/services/NetSurfP-1.1/), 20 O-glycosylation sites were predicted to be exposed on the HKU1 S1 outer surface (**Supplementary Table S6**), among which 10 were unambiguously exposed. The exposure degree of O-glycosylation sites on HKU1 S1 was 55.6% (10/18), which was slightly higher than that of SARS-CoV-2 S (52.4%, Bagdonaite et al., 2021). Particularly, the exposure ratio of O-glycosylation sites on RBD of HKU1 was 60% (6/10).

Mapping of Relative Normalized Abundances of O-Glycosylation on the 3D Model of HKU1 S1

To further explore the specific-site O-glycosylation on different sites, we investigated the relative normalized abundance of O-glycosylation on HKU1 S1 with visual models (PDB ID: 5I08). The normalized O-glycosylation abundance was calculated by dividing the O-glycosylation abundance on each O-glycosylation site by the total O-glycosylation abundance of HKU1 S1. We defined five magnitudes to represent different normalized O-glycosylation abundances as denoted in the keys (**Figure 5B**). As shown in **Figure 5B**, the RBD on the "head" of the subunit exhibited abundant O-glycosylation, especially the binding domain with receptor, but less exposed O-glycosylation on NTD.

Heat Map of the Total Relative Normalized Abundance of O-Glycosylation on HKU1 S1

In addition to the identification of O-glycosylation sites on the HKU1 S1 protein, the O-linked glycans were recognized. After searching the database with the 15 most common O-linked glycans, the distribution and normalized abundance of site-specific O-linked glycans of HKU1 S1 were mapped (**Figure 5C**). HexNAc(2)Hex(2), HexNAc(1)Hex(2), HexNAc(1)Hex(1)NeuAc(1), and HexNAc(1)Hex(1)NeuAc(2) were the top four most abundant O-linked glycans. Furthermore, the RBD displayed more O-glycosylation abundance on certain sites, such as S459, S578, and S584. K80 is the key residue for the HKU1 S protein to bind to 9-O-acetylated sialic acids from host cells (Hulswit et al., 2019). We identified that T82 or T83, which is adjacent to K80, is O-glycosylated. This may suggest hints between O-glycosylation and receptor binding. Overall, abundant O-glycosylation occurs on HKU1 S1, which also exhibits the micro- and macro-heterogeneity of O-glycosylation.

CONCLUSION

In summary, a comprehensive study of O-glycosylation of the HKU1 S protein S1 subunit was carried out, and 46 O-glycosylation sites were identified, among which 18 were unambiguously identified. All of the O-glycosylation sites were reported for the first time in this study. The novel O-glycosylation information will give insights to the microstructure of the HKU1 S protein, thus facilitating the development of a potential HKU1 vaccine.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ProteomeXchange, PXD025967.

AUTHOR CONTRIBUTIONS

DF, XL (6th author), and XL (7th author) secured funding, contributed to the experimental design, and revised the manuscript. YC and XD finished most of the experiments and wrote the draft of the manuscript. XZ and CC participated in O-glycopeptide enrichment from standard proteins and method validation.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.707235/full#supplementary-material

REFERENCES

Anan, G., Yoneyama, T., Noro, D., Tobisawa, Y., Hatakeyama, S., Sutoh Yoneyama, M., et al. (2019). The Impact of Glycosylation of Osteopontin on Urinary Stone Formation. *Int. J. Mol. Sci.* 21, 884–895. doi:10.1128/JVI.79.2.884–895.200510.3390/ijms21010093

Bagdonaite, I., Nordén, R., Joshi, H. J., Dabelsteen, S., Nyström, K., Vakhrushev, S. Y., et al. (2015). A Strategy for O-Glycoproteomics of Enveloped Viruses-The O-Glycoproteome of Herpes Simplex Virus Type 1. *Plos Pathog.* 11, e1004784. doi:10.1371/journal.ppat.1004784

Bagdonaite, I., Nordén, R., Joshi, H. J., King, S. L., Vakhrushev, S. Y., Olofsson, S., et al. (2016). Global Mapping of O-Glycosylation of Varicella Zoster Virus, Human Cytomegalovirus, and Epstein-Barr Virus. *J. Biol. Chem.* 291, 12014–12028. doi:10.1074/jbc.M116.721746

Bagdonaite, I., Thompson, A. J., Wang, X., Søgaard, M., Fougeroux, C., Frank, M., et al. (2021). Site-specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells. *Viruses* 13, 551. doi:10.3390/v13040551

Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. R. (2009). Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat. Protoc.* 4, 484–494. doi:10.1038/nprot.2009.21

Bräutigam, J., Scheidig, A. J., and Egge-Jacobsen, W. (2012). Mass Spectrometric Analysis of Hepatitis C Viral Envelope Protein E2 Reveals Extended Microheterogeneity of Mucin-type O-Linked Glycosylation. *Glycobiology* 23, 453–474. doi:10.1093/glycob/cws171

Buszewski, B., and Noga, S. (2012). Hydrophilic Interaction Liquid Chromatography (HILIC)-a Powerful Separation Technique. *Anal. Bioanal. Chem.* 402, 231–247. doi:10.1007/s00216-011-5308-5

Cao, L., Yu, L., Guo, Z., Shen, A., Guo, Y., and Liang, X. (2014). N-glycosylation Site Analysis of Proteins from *Saccharomyces cerevisiae* by Using Hydrophilic Interaction Liquid Chromatography-Based Enrichment, Parallel Deglycosylation, and Mass Spectrometry. *J. Proteome Res.* 13, 1485–1493. doi:10.1021/pr401049e

Christian, M. D., Poutanen, S. M., Loutfy, M. R., Muller, M. P., and Low, D. E. (2004). Severe Acute Respiratory Syndrome. *Clin. Infect. Dis.* 38, 1420–1427. doi:10.1086/420743

Dong, X., Qin, H., Mao, J., Yu, D., Li, X., Shen, A., et al. (2017). In-depth Analysis of Glycoprotein Sialylation in Serum Using a Dual-Functional Material with superior Hydrophilicity and Switchable Surface Charge. *Anal. Chem.* 89, 3966–3972. doi:10.1021/acs.analchem.6b04394

Fung, T. S., and Liu, D. X. (2018). Post-translational Modifications of Coronavirus Proteins: Roles and Function. *Future Virol.* 13, 405–430. doi:10.2217/fvl-2018-0008

Heald-Sargent, T., and Gallagher, T. (2012). Ready, Set, Fuse! the Coronavirus Spike Protein and Acquisition of Fusion Competence. *Viruses* 4, 557–580. doi:10.3390/v4040557

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271–280. doi:10.1016/j.cell.2020.02.052

Hoffmann, M., Marx, K., Reichl, U., Wuhrer, M., and Rapp, E. (2016). Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol. Cell Proteomics.* 15, 624–641. doi:10.1074/mcp.M115.053546

Huang, J., Jiang, B., Zhao, H., Wu, M., Kong, S., Liu, M., et al. (2020). Development of a Computational Tool for Automated Interpretation of Intact O-Glycopeptide Tandem Mass Spectra from Single Proteins. *Anal. Chem.* 92, 6777–6784. doi:10.1021/acs.analchem.0c01091

Hulswit, R. J. G., Lang, Y., Bakkers, M. J. G., Li, W., Li, Z., Schouten, A., et al. (2019). Human Coronaviruses OC43 and HKU1 Bind to 9-O-Acetylated Sialic Acids via a Conserved Receptor-Binding Site in Spike Protein Domain A. *Proc. Natl. Acad. Sci. USA.* 116, 2681–2690. doi:10.1073/pnas.1809667116

Iversen, M. B., Reinert, L. S., Thomsen, M. K., Bagdonaite, I., Nandakumar, R., Cheshenko, N., et al. (2016). An Innate Antiviral Pathway Acting before Interferons at Epithelial Surfaces. *Nat. Immunol.* 17, 150–158. doi:10.1038/ni.3319

Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., et al. (2016). Pre-fusion Structure of a Human Coronavirus Spike Protein. *Nature* 531, 118–121. doi:10.1038/nature17200

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., et al. (2003). Angiotensin-converting Enzyme 2 Is a Functional Receptor for the SARS Coronavirus. *Nature* 426, 450–454. doi:10.1038/nature02145

Millet, J. K., and Whittaker, G. R. (2015). Host Cell Proteases: Critical Determinants of Coronavirus Tropism and Pathogenesis. *Virus. Res.* 202, 120–134. doi:10.1016/j.virusres.2014.11.021

Mou, H., Raj, V. S., van Kuppeveld, F. J. M., Rottier, P. J. M., Haagmans, B. L., and Bosch, B. J. (2013). The Receptor Binding Domain of the New Middle East Respiratory Syndrome Coronavirus Maps to a 231-residue Region in the Spike Protein that Efficiently Elicits Neutralizing Antibodies. *J. Virol.* 87, 9379–9383. doi:10.1128/JVI.01277-13

Narimatsu, Y., Joshi, H. J., Schjoldager, K. T., Hintze, J., Halim, A., Steentoft, C., et al. (2019). Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics. *Mol. Cell Proteomics.* 18, 1396–1409. doi:10.1074/mcp.ra118.001121

Olofsson, S., Blixt, O., Bergström, T., Frank, M., and Wandall, H. H. (2016). Viral O-GalNAc Peptide Epitopes: a Novel Potential Target in Viral Envelope Glycoproteins. *Rev. Med. Virol.* 26, 34–48. doi:10.1002/rmv.1859

Peng, G., Sun, D., Rajashankar, K. R., Qian, Z., Holmes, K. V., and Li, F. (2011). Crystal Structure of Mouse Coronavirus Receptor-Binding Domain Complexed with its Murine Receptor. *Proc. Natl. Acad. Sci.* 108, 10696–10701. doi:10.1073/pnas.1104306108

Peng, G., Xu, L., Lin, Y.-L., Chen, L., Pasquarella, J. R., Holmes, K. V., et al. (2012). Crystal Structure of Bovine Coronavirus Spike Protein Lectin Domain. *J. Biol. Chem.* 287, 41931–41938. doi:10.1074/jbc.M112.418210

Qian, Z., Ou, X., Góes, L. G. B., Osborne, C., Castano, A., Holmes, K. V., et al. (2015). Identification of the Receptor-Binding Domain of the Spike Glycoprotein of Human Betacoronavirus HKU1. *J. Virol.* 89, 8816–8827. doi:10.1128/JVI.03737-14

Qin, H., Dong, X., Mao, J., Chen, Y., Dong, M., Wang, L., et al. (2019). Highly Efficient Analysis of Glycoprotein Sialylation in Human Serum by Simultaneous Quantification of Glycosites and Site-specific Glycoforms. *J. Proteome Res.* 18, 3439–3446. doi:10.1021/acs.jproteome.9b00332

Qing, G., Yan, J., He, X., Li, X., and Liang, X. (2020). Recent Advances in Hydrophilic Interaction Liquid Interaction Chromatography Materials for Glycopeptide Enrichment and Glycan Separation. *Trac Trends Anal. Chem.* 124, 115570. doi:10.1016/j.trac.2019.06.020

Riley, N. M., Malaker, S. A., Driessen, M. D., and Bertozzi, C. R. (2020). Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J. Proteome Res.* 19, 3286–3301. doi:10.1021/acs.jproteome.0c00218

Shao, W., Liu, J., Yang, K., Liang, Y., Weng, Y., Li, S., et al. (2016). Hydrogen-bond Interaction Assisted Branched Copolymer HILIC Material for Separation and N-Glycopeptides Enrichment. *Talanta* 158, 361–367. doi:10.1016/j.talanta.2016.05.034

Silver, Z. A., Antonopoulos, A., Haslam, S. M., Dell, A., Dickinson, G. M., Seaman, M. S., et al. (2020). Discovery of O-Linked Carbohydrate on HIV-1 Envelope and its Role in Shielding against One Category of Broadly Neutralizing Antibodies. *Cel Rep.* 30, 1862–1869. doi:10.1016/j.celrep.2020.01.056

Singh, Y., Rodriguez Benavente, M. C., Al-Huniti, M. H., Beckwith, D., Ayyalasomayajula, R., Patino, E., et al. (2020). Positional Scanning MUC1 Glycopeptide Library Reveals the Importance of PDTR Epitope Glycosylation for Lectin Binding. *J. Org. Chem.* 85, 1434–1445. doi:10.1021/acs.joc.9b02396

Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T.-B. G., Kong, Y., Bennett, E. P., et al. (2011). Mining the O-Glycoproteome Using Zinc-finger Nuclease-Glycoengineered SimpleCell Lines. *Nat. Methods.* 8, 977–982. doi:10.1038/nmeth.1731

Stone, J. A., Nicola, A. V., Baum, L. G., and Aguilar, H. C. (2016). Multiple Novel Functions of Henipavirus O-Glycans: the First O-Glycan Functions Identified in the Paramyxovirus Family. *Plos Pathog.* 12, e1005445. doi:10.1371/journal.ppat.1005445

Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., et al. (2020). Vulnerabilities in Coronavirus Glycan Shields Despite Extensive Glycosylation. *Nat. Commun.* 11, 2688. doi:10.1038/s41467-020-16567-0

Woo, P. C. Y., Lau, S. K. P., Chu, C., Chan, K., Tsoi, H., Huang, Y., et al. (2005). Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J. Virol.* 79, 884–895. doi:10.1128/JVI.79.2.884-895.200510.1128/jvi.79.2.884-895.2005

Woo, P. C. Y., Lau, S. K. P., Yip, C. C. Y., Huang, Y., and Yuen, K. (2009). More and More Coronaviruses: Human Coronavirus HKU1. *Viruses* 1, 57–71. doi:10.3390/v1010057

Yang, L., Sun, Z., Zhang, L., Cai, Y., Peng, Y., Cao, T., et al. (2019). Chemical Labeling for fine Mapping of IgG N-Glycosylation by ETD-MS. *Chem. Sci.* 10, 9302–9307. doi:10.1039/C9SC02491C

Yang, S., Hu, Y., Sokoll, L., and Zhang, H. (2017). Simultaneous Quantification of N- and O-Glycans Using a Solid-phase Method. *Nat. Protoc.* 12, 1229–1244. doi:10.1038/nprot.2017.034

You, X., Qin, H., and Ye, M. (2018). Recent Advances in Methods for the Analysis of Protein O-Glycosylation at Proteome Level. *J. Sep. Sci.* 41, 248–261. doi:10.1002/jssc.201700834

Zaki, A. M., Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2012). Isolation of a Novel Coronavirus from a Man with Pneumonia in saudi arabia. *N. Engl. J. Med.* 367, 1814–1820. doi:10.1056/NEJMoa1211721

Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., et al. (2020). Virus-receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell. Host. Micr.* 28, 586–601. doi:10.1016/j.chom.2020.08.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

O-Glycosylation Landscapes of SARS-CoV-2 Spike Proteins

Yong Zhang[1]*, Wanjun Zhao[2], Yonghong Mao[3], Yaohui Chen[3], Shanshan Zheng[1], Wei Cao[1], Jingqiang Zhu[2], Liqiang Hu[1], Meng Gong[1], Jingqiu Cheng[1]* and Hao Yang[1]*

[1]Key Laboratory of Transplant Engineering and Immunology, MOH, Frontiers Science Center for Disease-related Molecular Network, Institutes for Systems Genetics, West China Hospital, Sichuan University, Chengdu, China, [2]Department of Thyroid Surgery, West China Hospital, Sichuan University, Chengdu, China, [3]Institute of Thoracic Oncology, West China Hospital, Sichuan University, Chengdu, China

The densely glycosylated spike (S) proteins that are highly exposed on the surface of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) facilitate viral attachment, entry, and membrane fusion. We have previously reported all the 22 *N*-glycosites and site-specific *N*-glycans in the S protein protomer. Herein, we report the *O*-glycosylation landscapes of SARS-CoV-2 S proteins, which were characterized through high-resolution mass spectrometry. Following digestion with trypsin and trypsin/Glu-C, and de-*N*-glycosylation using PNGase F, we determined the GalNAc-type *O*-glycosylation pattern of S proteins, including *O*-glycosites and the six most common *O*-glycans occupying them, *via* Byonic identification and manual validation. Finally, 255 intact *O*-glycopeptides composed of 50 peptides sequences and 43 *O*-glycosites were discovered by higher energy collision-induced dissociation (HCD), and three *O*-glycosites were confidently identified by electron transfer/higher energy collision-induced dissociation (EThcD) in the insect cell-expressed S protein. Most glycosites were modified by non-sialylated *O*-glycans such as HexNAc(1) and HexNAc(1)Hex (1). In contrast, in the human cell-expressed S protein S1 subunit, 407 intact *O*-glycopeptides composed of 34 peptides sequences and 30 *O*-glycosites were discovered by HCD, and 11 *O*-glycosites were unambiguously assigned by EThcD. However, the measurement of O-glycosylation occupancy hasn't been made. Most glycosites were modified by sialylated *O*-glycans such as HexNAc(1)Hex (1)NeuAc (1) and HexNAc(1)Hex (1)NeuAc (2). Our results reveal that the SARS-CoV-2 S protein is an *O*-glycoprotein; the *O*-glycosites and *O*-glycan compositions vary with the host cell type. These comprehensive *O*-glycosylation landscapes of the S protein are expected to provide novel insights into the viral binding mechanism and present a strategy for the development of vaccines and targeted drugs.

Keywords: SARS-CoV-2, spike protein, O-glycosylation, mass spectrometry, EThcD fragmentation

INTRODUCTION

The spike (S) protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an extensively *N*-glycosylated protein (Watanabe et al., 2020) that protrudes from the virus surface and binds to the angiotensin-converting enzyme 2 (ACE2) receptor on host cells to mediate cell entry (Wrapp et al., 2020). All 22 *N*-glycosites and *N*-glycans attached to asparagine (Asn, N) in a recombinant S protein protomer expressed in human and insect cells have been identified using

high-resolution liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Lenza et al., 2020; Rosenbalm et al., 2020; Walls et al., 2020; Xu et al., 2020; Yan et al., 2020; Zhang et al., 2020; Wang et al., 2021; Zhou et al., 2021). These *N*-glycosites are preferentially distributed in two functional subunits responsible for receptor binding (S1 subunit) and membrane fusion (S2 subunit) (Zhang et al., 2020). Site-specific *N*-glycosylation analysis can provide valuable insights into the infection mechanism and present a strategy for the development of vaccines (Grant et al., 2020).

Unlike *N*-glycosylation, *O*-glycosylation is initiated by the *α*-glycosidic attachment of *N*-acetylgalactosamine (GalNAc) to the hydroxyl group of serine (Ser, S) or threonine (Thr, T), which contains eight types of core structures (Core-1 to Core-8 *O*-glycans), and is involved in a variety of biological functions, such as the mediation of pathogenic binding to human receptors (Mayr et al., 2018; Shajahan et al., 2020a). Moreover, *O*-glycosylation can influence proteolysis during antigen processing, which could prevent the formation of glycopeptides for further presentation to major histocompatibility complex (MHC) and the elicitation of immune response (Wolfert and Boons, 2013). The S protein *O*-glycosites of SARS-CoV-2 have been predicted using computational analysis (Uslupehlivan and Sener, 2020), and Shajahan et al. (2020) identified two *O*-glycosites (T323 and S325) using LC-MS/MS (Shajahan et al., 2020b). However, *O*-glycosylation often occurs in a cluster. Hence, we believe that there are many *O*-glycosites that have not been discovered as deciphering protein *O*-glycosylation remains a big challenge. The comprehensive *O*-glycosylation analysis cannot be performed without appropriate sample preprocessing, analysis methods, and software (King et al., 2017; Qin et al., 2017; Yang et al., 2017; Yang et al., 2018; Ye et al., 2019; Park et al., 2020; Dong et al., 2021).

In the present study, we characterized the intact *O*-glycopeptides of recombinant SARS-CoV-2 S proteins expressed in human and insect cells, using LC-MS/MS. Based on a complementary enzyme digestion strategy, we identified large-scale *O*-glycosites and their corresponding *O*-glycans in the recombinant S proteins. The heterogeneity by different glycoforms of S protein S1 subunits expressed in human and insect cells was resolved and compared. Detailed *O*-glycosylation profiles of S proteins are complementary to the *N*-glycosylation profiles and may help in the development of vaccines and therapeutic drugs.

EXPERIMENTAL SECTION

Materials and Chemicals
Dithiothreitol (DTT), iodoacetamide (IAA), formic acid (FA), trifluoroacetic acid (TFA), Tris base, and urea were purchased from Sigma (St. Louis, MO, United States). Acetonitrile (ACN) was purchased from Merck (Darmstadt, Germany). Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) materials were purchased from Fresh Bioscience (Shanghai,

China). The C18 and C8 membrane were purchased from Agela Technologies (Tianjin, China). Recombinant SARS-CoV-2 S protein (S1+S2 ECD, His tag) expressed by insect cells (High Five) *via* a baculovirus, and S protein (S1, His tag) expressed by human embryonic kidney (HEK293) cells were purchased from Sino Biological (Beijing, China). Codon-optimized DNA sequences encoding the SARS-CoV-2 S protein subunits were cloned into pCMV3-C-His and a baculovirus vector with a poly-histidine tag at the C terminus for recombinant expression of these proteins in human and insect cells, respectively. Sequencing-grade trypsin and Glu-C were obtained from Enzyme & Spectrum (Beijing, China). A quantitative colorimetric peptide assay kit was purchased from Thermo Fisher Scientific (Waltham, MA, United States). Deionized water was prepared using a Milli-Q system (Millipore, Bedford, MA, United States). All other chemicals and reagents of the best available grade were purchased from Sigma-Aldrich or Thermo Fisher Scientific.

Protein digestion
Recombinant S proteins were proteolyzed using an in-solution protease digestion protocol. In brief, 50 μg of protein were dissolved in 100 μl of 50 mM NH_4HCO_3 buffer (pH = 8.5) and heated to denature for 10 min at 95°C. After reduction by DTT (20 mM) for 45 min at 56°C and alkylation with IAA (50 mM) for 1 h at 25°C in the dark, 2 μg of protease (trypsin or trypsin/Glu-C (w/w = 1:1)) was added to the tube and incubated for 16 h at 37°C. Peptides were loaded in a pipette tip which was packed with a C18 membrane. After washing three times using 70 μl of 2% acetonitrile/98% water/0.1% formic acid. Peptides bound to the C18 membrane were eluted three times with 70 μl of 80% acetonitrile/20% water/0.1% formic acid. The peptide concentration was determined using a peptide assay kit, based on the absorbance measured at 480 nm. The peptide mixtures were freeze-dried for further analysis.

Enrichment of Intact Glycopeptides and N-Glycan Removal
Intact *N*- and *O*-glycopeptides were enriched with ZIC-HILIC materials. Specifically, 20 μg of peptides was suspended in 100 μl of 80% ACN/0.2% TFA solution. 5 mg of Zic-HILIC was washed three times for 10 min each with 0.1% TFA and 80% ACN/0.2% TFA, and 2 mg of processed ZIC-HILIC materials was added to the peptide solution and incubated for 2 h at 37°C. Finally, the mixture was transferred to a 200 μl pipette tip packed with a C8 membrane, and washed twice with 80% ACN/0.2% TFA. After enrichment, intact glycopeptides were eluted thrice with 70 μl of 0.1% TFA, and dried using a SpeedVac concentrator. The enriched intact glycopeptides were digested using 1 U PNGase F dissolved in 50 μl of 50 mM NH_4HCO_3 for 2 h at 37°C. The reaction was terminated by adding 0.1% FA. The de-*N*-glycopeptides and *O*-glycopeptides were dried using a SpeedVac concentrator for further analysis.

Liquid chromatography-Tandem Mass Spectrometry Analysis

All the samples were analyzed using higher energy collision-induced dissociation (HCD) in mass spectrometry (Orbitrap Fusion Lumos mass spectrometer). In brief, intact O-glycopeptides and de-N-glycopeptides were dissolved in 0.1% FA and separated on a column (ReproSil-Pur C18-AQ, 1.9 μm, 75 μm inner diameter, 20 cm length; Dr Maisch) over a 78 min gradient (buffer A, 0.1% FA in water; buffer B, 0.1% FA in 80% ACN) at a flow rate of 300 nL/min. MS1 was analyzed with a scan range (m/z) of 350–1,550 at an Orbitrap resolution of 120,000. The RF lens, AGC target, maximum injection time, and exclusion duration were 30%, 1.0e6, 50 ms, and 15 s, respectively. MS2 was analyzed with an isolation window (m/z) of two at an Orbitrap resolution of 15,000. The AGC target, maximum injection time, and HCD type were 5.0e4, 80 ms, and 35%, respectively. For further verification, the same samples were analyzed using electron transfer/higher energy collision-induced dissociation (EThcD) in mass spectrometry (Orbitrap Fusion Lumos mass spectrometer). The MS1 was analyzed with a mass range of 400–1,600 at a resolution of 120,000 at 200 m/z. The RF Lens was set as 30% and the maximum injection time (MIT) was 100 ms. The MS2 was analyzed in quadrupole mode and the isolation window was 2 m/z. The EThcD collision energy type was 35%. The MIT was set at 250 ms and cycle time was set at 3 s.

Data analysis

Raw data files were searched against the SARS-CoV-2 S protein sequence using Byonic™ software (version 3.6.0, Protein Metrics, Inc.), with the mass tolerance for precursors and fragment ions set at ±10 and ±20 ppm, respectively. HCD or EThcD was chosen as the fragmentation type. Two missed cleavage sites were subjected to trypsin or trypsin/Glu-C digestion. The fixed modification was carbamidomethyl (C), and the variable modifications included oxidation (M), acetyl (protein N-term), and de-amidation (N). In addition, the six most common O-glycans (HexNAc(1) with mass of 203.079 Da; HexNAc(2) with mass of 406.159 Da; HexNAc(1)Hex (1) with mass of 365.132 Da; HexNAc(2)Hex (1) with mass of 568.212 Da; HexNAc(1)Hex (1)NeuAc (1) with mass of 656.228 Da; and HexNAc(1)Hex (1)NeuAc (2) with mass of 947.323 Da) were specified as O-glycan modifications for intact O-glycopeptides. We then added the protein database and the decoy database. All other parameters were set to the default values, and protein groups were filtered using a 1% false discovery rate, based on the number of hits obtained for the searches against the databases. Stricter quality control methods for intact O-glycopeptide identification were implemented; they required a score of not less than 300, and at least six amino acids to be identified. Furthermore, all the glycopeptide-spectrum matches (GPSMs) were examined manually by checking the oxonium ions and b/y/c/z ions to ensure the correct identification of the glycopeptides and their glycan compositions, and distinguish the confident O-glycosites and their linked glycans from the uncertain glycosites within a specific glycopeptide. In addition, these O-glycosites had to be identified repeatedly at least twice.

Model building based on the Cryo-EM structure (PDB: 6VSB) of the SARS-CoV-2 S protein was performed using PyMOL.

RESULTS AND DISCUSSION

Strategy for Intact O-Glycopeptide Analysis

Our previous study, as well as others, have revealed site-specific N-glycosylation of recombinant S proteins (Watanabe et al., 2020; Zhang et al., 2020). Two or more potential O-glycosites have also been discovered using LC-MS/MS in recent studys (Shajahan et al., 2020b; Bagdonaite et al., 2021). However, comprehensive O-glycosylation analysis of the SARS-CoV-2 S protein has not been performed. In the present study, we aimed to characterize the O-glycosylation landscapes of SARS-CoV-2 recombinant S proteins by analysis of intact O-glycopeptides, including potential O-glycosites and their linked O-glycans.

The strategy for intact O-glycopeptide analysis is shown in **Figure 1A**. The recombinant SARS-CoV-2 S proteins were digested using trypsin or a mixture of trypsin and Glu-C to cover as many potential O-glycosites as possible. Then, intact glycopeptides were enriched using ZIC-HILIC (Pohlentz et al., 2016), and de-N-glycosylated with PNGase F to avoid interference from non-glycopeptides and N-glycopeptides. Finally, intact O-glycopeptides were analyzed using a high-resolution mass spectrometer, and their mass spectra were characterized using Byonic™ software and validated manually (Zhang et al., 2018). It is worth remarking that O-glycosylation assignment to a specific amino acid by Byonic™ is not always confident when multiple Ser/Thr residues are present within the glycopeptide, especially when using HCD-type MS2 fragmentation. These O-glycosites were classified into the potential sites in this study.

The S protein expressed in insect cells contained 1,209 amino acids (residues 16–1,213), including 94 Thr and 92 Ser residues regarded as potential O-glycosites. The spike protein S1 subunit expressed in human cells contained 681 amino acids (residues 16–685), including 57 Thr and 50 Ser residues as potential O-glycosites (**Supplementary Figure S1**). Combined digestion strategy can improve glycosite identification and glycoprotein sequence coverage (Chen et al., 2011). To evaluate our method based on MS analysis and data analysis by Byonic in this study, we first analyzed the two previously reported O-glycosites, T323 and S325. The spike protein subunits S1 and S2 expressed on human cells were digested by trypsin and/or chymotrypsin, and analyzed by stepped HCD product triggered CID (HCD-pd-CID) without glycopeptide enrichment and PNGase F digestion (Shajahan et al., 2020b). As shown in **Figure 1B**, Byonic™ analysis disclosed the presence of the O-glycopeptide [320]VQPTESIVR[328] with an uncertain O-glycosite at T323 based on the b/y ions with or without glycan retention. It is worth noting that S325 was an alternative glycosite in this peptide because only the "~y7+203" ion with glycan retention was detected, although Byonic™ tended to assign the T323 according to the y4/y5 ions without the linked glycan. However, the presence of ~y6/~y7/~y8/~b7 ions without the

FIGURE 1 | Comprehensive *O*-glycosylation profiling of SARS-CoV-2 spike proteins. **(A)** SARS-CoV-2 spike proteins expressed in insect or human cells were digested using trypsin or a mixture of trypsin and Glu-C. After ZIC-HILIC enrichment and PNGase F digestion, intact *O*-glycopeptides were analyzed using a high-resolution mass spectrometer, and their spectra were characterized using Byonic™ software and validated manually. **(B)** HCD-MS/MS spectrum of reported representative *O*-glycopeptide ^{320}VQPTESIVR^{328} with deduced GalNAcGal glycan detected on site Thr323 or Ser325 of human spike protein subunit 1. **(C)** HCD-MS/MS spectrum of this *O*-glycopeptide with deduced GalNAcGalNeuAc glycan detected on site Thr323 or Ser325.

glycan indicates the fact that the b/y ions produced by HCD tends to lose their glycans (Pap et al., 2018). Without sufficient b/y ions with glycan retention, the confident *O*-glycosite cannot be determined in the glycopeptide. Shajahan et al. has reported that T323 seems a predominantly occupied site in an *O*-glycopepitde with the same peptide sequence ^{320}VQPTESIVR^{328} through HCD fragmentation (Shajahan et al., 2020b), suggesting T323 is a high-probability glycosite. Similarly, the representative HCD-MS/MS spectra in **Figure 1C** revealed the presence of an uncertain *O*-glycosite at S325, which also could be T323 due to high frequency loss of entire glycan in HCD fragmentation. Hence, both T323 and S325 are uncertain *O*-glycosites and could not be confident identified by the HCD-MS/MS spectra in **Figures 1B,C**. These results indicate that our strategy is feasible for *O*-glycosylation profiling.

Comprehensive O-Glycosylation Profiling of Recombinant SARS-CoV-2 S Protein Expressed in Insect Cells

The S protein produced by the baculovirus insect cell expression system contained 186 potential *O*-glycosites. Using our aforementioned strategy, a total of 255 intact O-glycopeptides composed of 50 peptides backbones and 43 uncertain *O*-glycosites were discovered by HCD (**Supplementary Table S1** and **Figure S2**). In these glycopeptides, 40 potential *O*-glycosites, except S477, T572, and T732 were found repeatedly using trypsin alone. Using trypsin combined with Glu-C, three more *O*-glycosites were discovered, although another three *O*-glycosites (S325, T333, and T1066) were missed due to combinational digestion (**Figure 2A**). Hence, although trypsin digestion can yield good identification results,
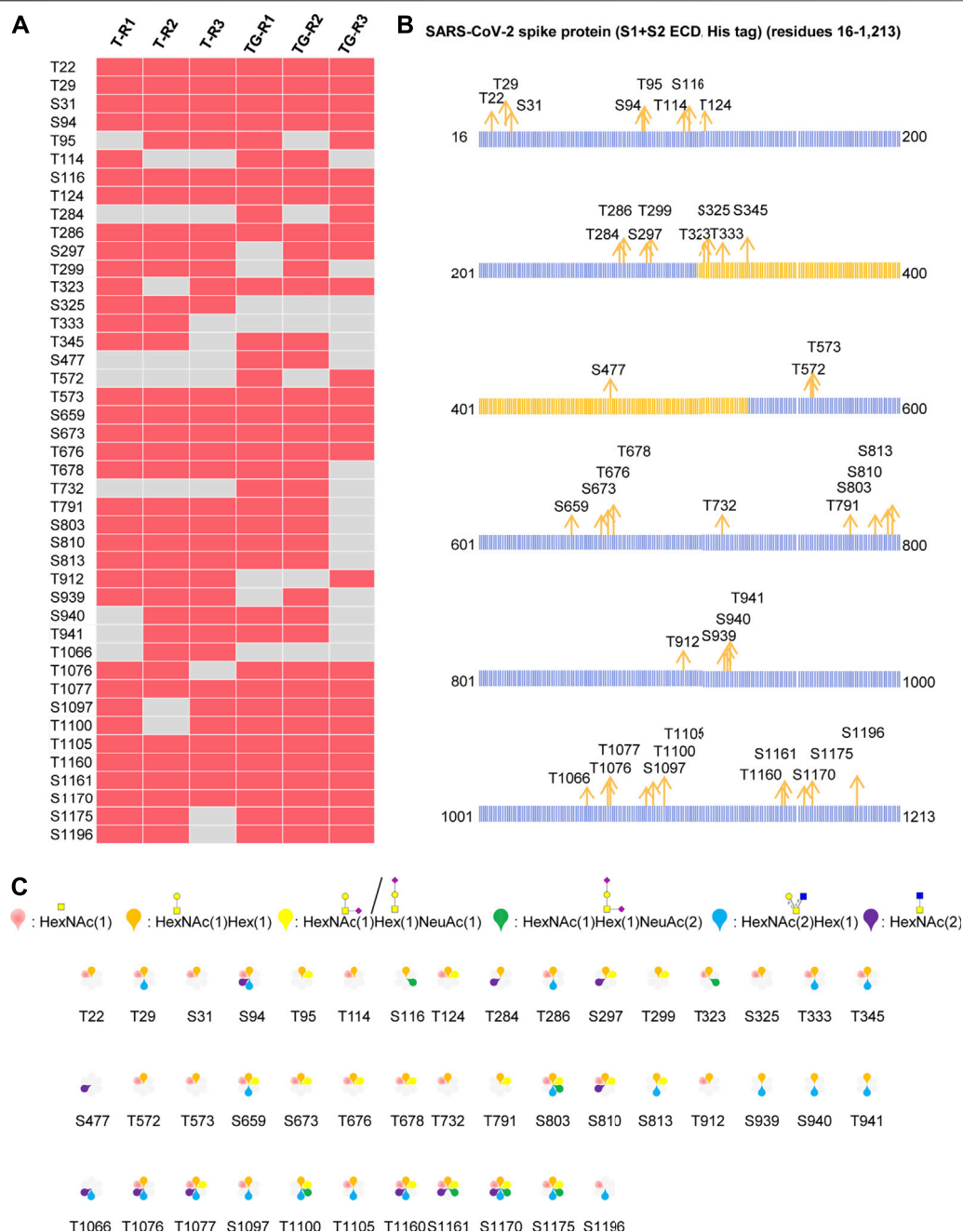
FIGURE 2 | Comprehensive O-glycosylation characterization of recombinant SARS-CoV-2 S protein (S1+S2 ECD, His tag) expressed in insect cells. (A) Uncertain O-glycosites identified using trypsin (T) or typsin/Glu-C (TG) in three replicates. (B) Mapping of identified O-glycosites to amino acid sequences. RBD is highlighted in yellow. (C) O-glycan compositions on each site.

trypsin combined with Glu-C digestion should be considered as complementary step because that some suitable glycopeptides can be easily detected by mass spectrometry. Furthermore, we mapped these O-glycosites to the amino sequences, and found that the O-glycosites appeared in several areas, especially in the N- and C-termini of the S protein (**Figure 2B**). It is notable that the O-glycosites T323, S325, T333, S345, and S477 were located in the receptor-binding domain (RBD). These results indicate that

the SARS-CoV-2 S protein is an O-glycoprotein with a large number of O-glycosites. In addition, the number of O-glycosylated Thr residues 25) was higher than that of O-glycosylated Ser residues 18) (**Figure 2B**). This result is consistent with those of previous studies on O-glycosites (Zhang et al., 2018). Finally, a global O-glycan composition analysis was performed (**Figure 2C**). Six O-glycan compositions were identified on these sites, including
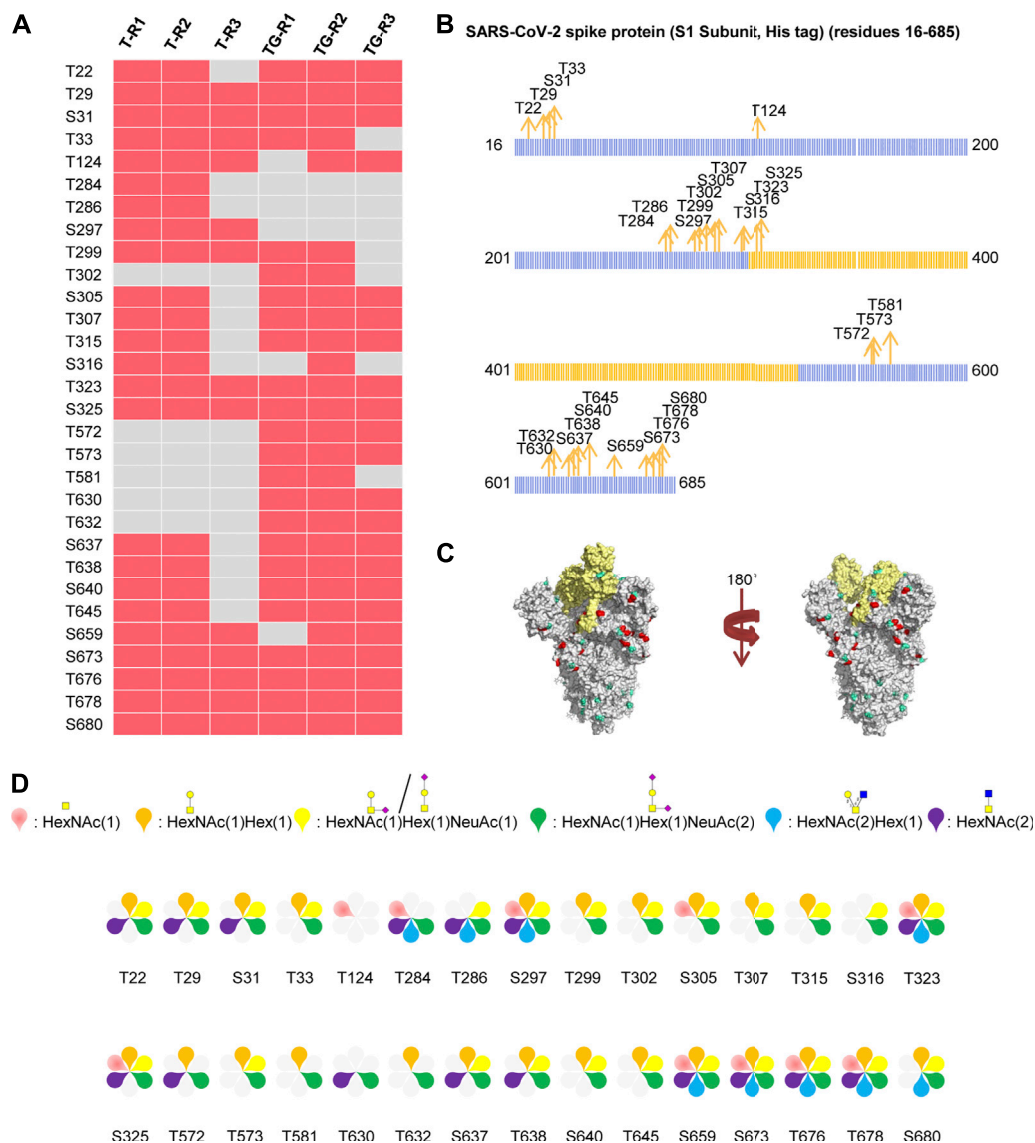
FIGURE 3 | Comprehensive O-glycosylation characterization of SARS-CoV-2 S protein (S1, His tag) expressed in human cells. **(A)** Uncertain O-glycosites identified using trypsin (T) or typsin/Glu-C (TG) in three replicates. **(B)** Mapping of identified O-glycosites to amino acid sequences. RBD is highlighted in yellow. **(C)** O-glycosites (red) and N-glycosites (blue) in three-dimensional structure of SARS-CoV-2 S protein trimers (PDB code: 6VSB). RBD is highlighted in yellow. **(D)** O-glycan compositions on each site.

HexNAc(1), HexNAc(2), HexNAc(1)Hex (1), HexNAc(2)Hex (1), HexNAc(1)Hex (1)NeuAc (1), and HexNAc(1)Hex (1) NeuAc (2). Regarding the frequency of these glycans on different glycosites, occupancies with HexNAc(1)Hex (1), HexNAc(1), HexNAc(2)Hex (1), HexNAc(1)Hex (1)NeuAc (1), HexNAc(2), and HexNAc(1)Hex (1)NeuAc (2) compositions were found on 40, 30, 21, 18, 11, and seven glycosites, respectively. Moreover, most glycosites contained at least two types of O-glycans, a majority of which were non-sialylated (**Figure 2C**). It's worth noting that NeuAc would rarely found on insect O-glycans, because that insect cells generally lack adequate levels of the glycosyltransferases to synthesize sialylated products, especially if there's not NeuAc

oxonium ion in the MS2 spectrum (**Figure 2C**). When that happens, the spectrum may be a half-right identification (right peptide with wrong O-glycans) although with high Byonic score. These results indicate the O-glycans appeared on the recombinant SARS-CoV-2 S protein expressed in insect cells.

Comprehensive O-Glycosylation Profiling of Recombinant SARS-CoV-2 S Protein Expressed in Human Cells

The recombinant SARS-CoV-2 S protein S1 subunit produced by the human cell expression system was used for analysis of the
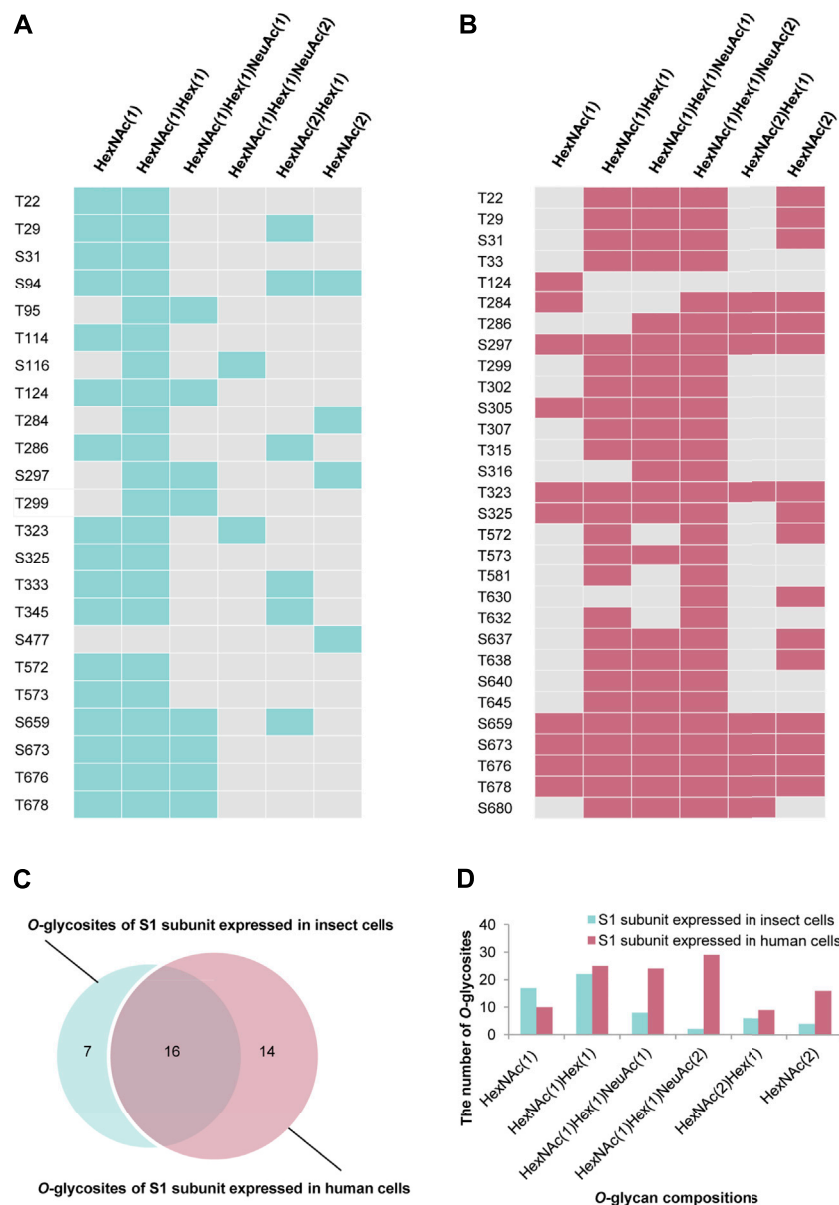
FIGURE 4 | Comparison of O-glycosylation modifications of S1 subunits expressed in insect or human cells. **(A)** O-glycan compositions in each glycosite of S1 subunit expressed in insect cells. **(B)** O-glycan compositions in each glycosite of S1 subunit expressed in human cells. **(C)** Comparison of O-glycosites of S1 subunits expressed in different expression systems. **(D)** Number of S1 subunit O-glycosites attached by each type of O-glycan composition.

O-glycans, as the O-glycan compositions in insect cells could be different from those in human cells. Using our aforementioned strategy, 407 intact O-glycopeptides composed of 34 peptides backbones and 30 uncertain O-glycosites (20 Thr and 10 Ser residues) were discovered by HCD (**Supplementary Table S2** and **Figure S3**). 24 and 27 uncertain O-glycosites were found repeatedly using trypsin and a mixture of trypsin/Glu-C, respectively. The trypsin combined with Glu-C digestion can increase the number of identified O-glycosites (**Figure 3A**). The results showed that the two digestion methods were complementary for O-glycosite identification. Furthermore, we mapped these 30 O-glycosites to the amino sequences. We found

that the O-glycosites mainly appeared at the S1 subunit and RBD (**Figures 3B,C**). It is notable that two conserved O-glycosites, T323 and S325, were located in the RBD of the S1 subunit, and may play a critical role in viral binding with hACE2 receptors (Andersen et al., 2020; Hoffmann et al., 2020). A global O-glycan composition analysis of the S1 subunit was performed. O-glycan occupancies with HexNAc(1)Hex (1)NeuAc (2), HexNAc(1)Hex (1), HexNAc(1)Hex (1)NeuAc (1), HexNAc(2), HexNAc(1), and HexNAc(2)Hex (1) were found on 29, 25, 24, 16, 10, and nine glycosites, respectively. O-glycans on most glycosites were sialylated (**Figure 3D**). These results indicate the more complex O-glycosylation and the heterogeneity of O-glycan
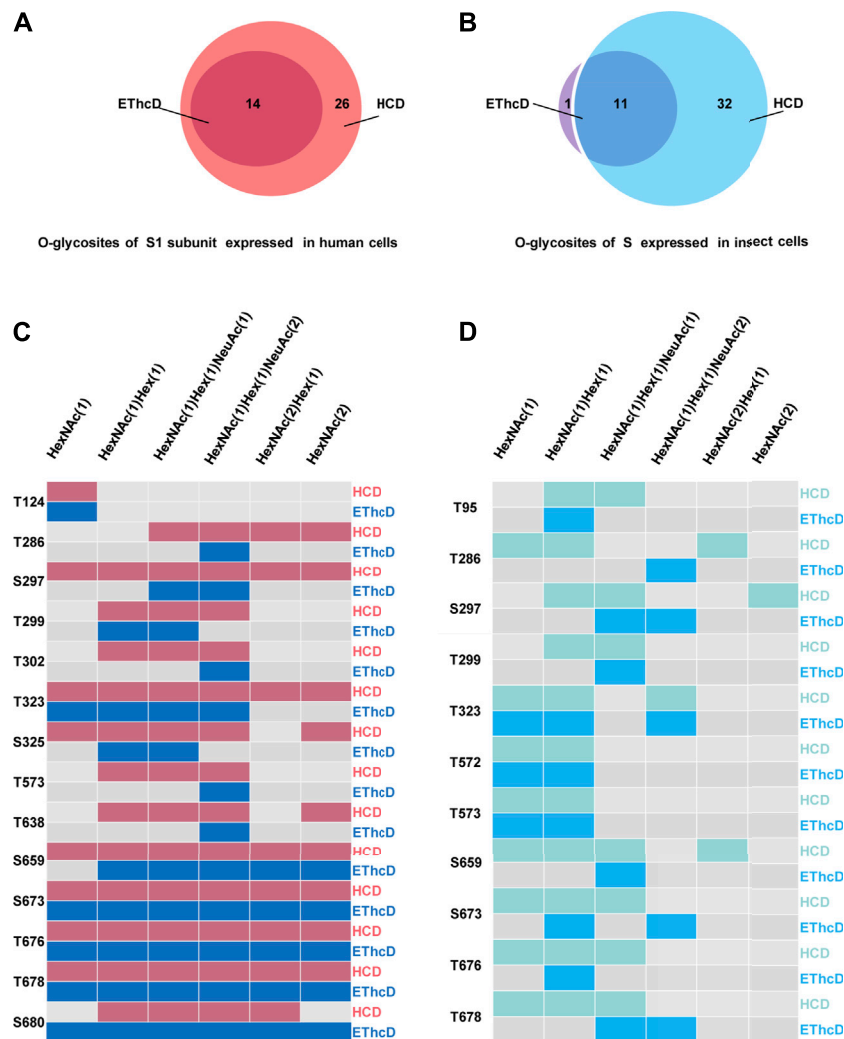
FIGURE 5 | Systems analysis of O-glycosylation of spike protein via EThcD and HCD mass spectrometry. **(A)** Identified O-glycosites of S1 subunit expressed in human cells using HCD and EThcD. **(B)** Identified O-glycosites of S expressed in insect cells using HCD and EThcD. **(C)** The same O-glycosites of S1 subunit expressed in human cells identified using both HCD and EThcD. **(D)** The same O-glycosites of S expressed in insect cells identified using both HCD and EThcD.

compositions on the recombinant SARS-CoV-2 S protein expressed in human cells.

O-Glycosylation Landscapes of S1 Subunits Expressed in Insect and Human Cells

Based on the above findings, we further compared the O-glycosylation landscapes of the S1 subunits expressed in insect and human cells. 23 uncertain O-glycosites were present in the S1 subunit expressed in insect cells (**Figure 4A**). In contrast, 30 O-glycosites were present in the S1 subunit expressed in human cells (**Figure 4B**). In addition, 16 common O-glycosites (T22, T29, S31, T124, T284, T286, S297, T299, T323, S325, T572, T573, S659, S673, T676, and T678) were discovered in the S1 subunits expressed in both insect and human cells, including the two sites, T323 and S325, located in the RBD. Seven and 14 unique O-glycosites were found in the

insect and human cell–produced S1 subunits, respectively (**Figure 4C**). It's worth noting that HCD can identify the intact O-glycopeptide confidently while usually failing to distinguish a specific O-glycosite from multiple uncertain glycosites within a glycopeptide. More validation experiments were needed to compare the differences between the two S proteins. Furthermore, the number of S1 subunit O-glycosites occupied by each type of O-glycan compositions was very different. Most O-glycosites of the insect cell–produced S1 subunit contained HexNAc(1)Hex (1) and HexNAc(1). On the other hand, most O-glycosites of the human cell–produced S1 subunit contained HexNAc(1)Hex (1)NeuAc (2), HexNAc(1) Hex (1)NeuAc (1), and HexNAc(1)Hex (1) (**Figure 4D**). These results implied that the O-glycosite and O-glycan compositions varied with the host cell type, which could be taken into account when using the recombinant proteins for vaccine and drug development.

Identification of Intact O-Glycopeptides Using EThcD

Electron transfer dissociation (ETD) can produce extensive fragmentation of the peptide backbone, enabling sequencing of the peptide, while preserving glycans on the peptide backbone (Hogan et al., 2005; Myers et al., 2013). However, ETD frequently leads to incomplete fragmentation and massive residual precursor ions, and is usually combined with HCD or CID (Kolbowski et al., 2017). Our previous research had shown that EThcD can provide a more complete fragmentation of O-glycopeptides than HCD or ETD alone (Frese et al., 2013), leading to better O-glycosylation site localization (Zhang et al., 2018). Hence, all the above samples were reanalyzed in triplicate via EThcD. A total of 12 O-glycosites of S expressed in insect cells and 14 O-glycosites of S1 subunit expressed in human cells were assigned (**Supplementary Table S3, Figure S4, S5**). Compared with HCD identification results, EThcD identified less O-glycosites. 14 out of 30 O-glycosites identified via HCD on S1 subunit expressed in human cells were identified by EThcD (**Figure 5A**). By further checking c/z ions with glycan retention, 11 O-glycosites (T124, T302, T323, S325, T573, T638, S659, T678, S673, T676, S680) were assigned confidently (**Supplementary Figure S5**). Similarly, EThcD enabled identification of 12 O-glycosites on S protein produced in insect cells, in which only one O-glycosite S680 was not found by HCD (**Figure 5B**). By further checking c/z ions which contain O-glycans, three O-glycosites (T95, T323, T573) can be assigned confidently (**Supplementary Figure S4**). Furthermore, the O-glycan types attached to each O-glycosite were analyzed and many O-glycans identified by HCD, especially the more complex O-glycans such as HexNAc(1)Hex (1)NeuAc (2), were verified by EThcD since ETD preferentially retained the intact glycan moities (**Figure 5C** and **Figure 5D**). In addition, HexNAc(2)Hex (1) and HexNAc(2) were not verified by EThcD in the S expressed in insect cells, which suggested that the two O-glycans might be misidentified by HCD due to fragmentation of glycans. All of these results indicate that with both HCD and EThcD, we can identify a large number of intact O-glycopeptides of spike proteins. However, more verification experiments were still needed to identify more confident O-glycosites. These results proved that the SARS-CoV-2 S protein is a glycoprotein decorated with various O-glycans.

Recently, there are a few O-glycosites of S protein that are consistently identified by different groups. For example, Bagdonaite et al. used an O-glycoproteomic workflow based on in-gel digestion, de-N-glycosylation and desialylation strategy to map O-glycosites on S protein expressed in insect cell or human cell, and in total 25 O-glycosites were identified (Bagdonaite et al., 2021). There are some differences on the O-glycosites reported between their work and our reports. The reasons include using different expression cell strain, the same recombinant S protein from the same cell strain but cultured and processed in different labs or vendors, different sample preparation procedures, different mass spectrometer or analytical method, different software to process the data, even different identification criteria and threshold. They may lead to significant variations in glycosylation analysis. Even so, these

methods and data may be useful for the development of vaccines and targeted drugs.

CONCLUSIONS

In this study, we profiled a comprehensive O-glycosylation pattern of SARS-CoV-2 S proteins using optimized experimental procedure and HCD and EThcD mass spectrometry. There are 255 intact O-glycopeptides composed of 50 peptides sequences and 43 uncertain O-glycosites were discovered by HCD in insect cell–expressed S protein, and most of them were non-sialylated. There are three O-glycosites were confidently identified by EThcD. In contrast, in human S protein, 407 intact O-glycopeptides composed of 34 peptides sequences and 30 uncertain O-glycosites were discovered by HCD, 11 O-glycosites were unambiguously assigned by EThcD, and most of them were sialylated. However, the measurement of O-glycosylation occupancy hasn't been made. Our results revealed that the SARS-CoV-2 S protein was modified by O-glycans, and that the O-glycosite and O-glycan compositions varied with the host cell type.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HY, MG, JZ, YC, and JC directed and designed research; YZ and WZ directed and performed analyses of mass spectrometry data; YM, SZ, LH, and WC adapted algorithms and software for data analysis; YZ and WZ coordinated acquisition, distribution and quality evaluation of samples; YZ and HY wrote the article.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.689521/full#supplementary-material

REFERENCES

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The Proximal Origin of SARS-CoV-2. *Nat. Med.* 26 (4), 450–452. doi:10.1038/s41591-020-0820-9

Bagdonaite, I., Thompson, A. J., Wang, X., Søgaard, M., Fougeroux, C., Frank, M., et al. (2021). Site-Specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells. *Viruses* 13 (4). doi:10.3390/v13040551

Chen, Y., Cao, J., Yan, G., Lu, H., and Yang, P. (2011). Two-step Protease Digestion and Glycopeptide Capture Approach for Accurate Glycosite Identification and Glycoprotein Sequence Coverage Improvement. *Talanta* 85 (1), 70–75. doi:10.1016/j.talanta.2011.03.029

Dong, X., Chen, C., Yan, J., Zhang, X., Li, X., and Liang, X. (2021). Comprehensive O-Glycosylation Analysis of the SARS-CoV-2 Spike Protein with Biomimetic Trp-Arg Materials. *Anal. Chem.* 93 (30), 10444–10452. doi:10.1021/acs.analchem.0c04634

Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F. M., Mechtler, K., Heck, A. J. R., et al. (2013). Unambiguous Phosphosite Localization Using Electron-Transfer/higher-Energy Collision Dissociation (EThcD). *J. Proteome Res.* 12 (3), 1520–1525. doi:10.1021/pr301130k

Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020). Analysis of the SARS-CoV-2 Spike Protein Glycan Shield Reveals Implications for Immune Recognition. *Sci. Rep.* 10 (1), 14991. doi:10.1038/s41598-020-71748-7

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181 (2), 271–280. e8. doi:10.1016/j.cell.2020.02.052

Hogan, J. M., Pitteri, S. J., Chrisman, P. A., and McLuckey, S. A. (2005). Complementary Structural Information from a TrypticN-Linked Glycopeptide via Electron Transfer Ion/Ion Reactions and Collision-Induced Dissociation. *J. Proteome Res.* 4 (2), 628–632. doi:10.1021/pr049770q

King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., et al. (2017). Characterizing the O-Glycosylation Landscape of Human Plasma, Platelets, and Endothelial Cells. *Blood Adv.* 1 (7), 429–442. doi:10.1182/bloodadvances.2016002121

Kolbowski, L., Mendes, M. L., and Rappsilber, J. (2017). Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer. *Anal. Chem.* 89 (10), 5311–5318. doi:10.1021/acs.analchem.6b04935

Lenza, M. P., Oyenarte, I., Diercks, T., Quintana, J. I., Gimeno, A., Coelho, H., et al. (2020). Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and Their Interactions with Human Lectins. *Angew. Chem. Int. Ed.* 59 (52), 23763–23771. doi:10.1002/anie.202011015

Mayr, J., Lau, K., Lai, J. C. C., Gagarinov, I. A., Shi, Y., McAtamney, S., et al. (2018). Unravelling the Role of O-Glycans in Influenza A Virus Infection. *Sci. Rep.* 8 (1), 16382–16412. doi:10.1038/s41598-018-34175-3

Myers, S. A., Daou, S., Affar, E. B., and Burlingame, A. (2013). Electron Transfer Dissociation (ETD): The Mass Spectrometric Breakthrough Essential for O -GlcNAc Protein Site Assignments-A Study of the O -GlcNAcylated Protein Host Cell Factor C1. *Proteomics* 13 (6), 982–991. doi:10.1002/pmic.201200332

Pap, A., Klement, E., Hunyadi-Gulyas, E., Darula, Z., and Medzihradszky, K. F. (2018). Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. *J. Am. Soc. Mass. Spectrom.* 29 (6), 1210–1220. doi:10.1007/s13361-018-1945-7

Park, G. W., Lee, J. W., Lee, H. K., Shin, J. H., Kim, J. Y., and Yoo, J. S. (2020). Classification of Mucin-type O-Glycopeptides Using Higher-Energy Collisional Dissociation in Mass Spectrometry. *Anal. Chem.* 92 (14), 9772–9781. doi:10.1021/acs.analchem.0c01218

Pohlentz, G., Marx, K., and Mormann, M. (2016). Characterization of Protein N-Glycosylation by Analysis of ZIC-HILIC-Enriched Intact Proteolytic Glycopeptides. *Methods Mol. Biol.* 1394, 163–179. doi:10.1007/978-1-4939-3341-9_12

Qin, H., Cheng, K., Zhu, J., Mao, J., Wang, F., Dong, M., et al. (2017). Proteomics Analysis of O-GalNAc Glycosylation in Human Serum by an Integrated Strategy. *Anal. Chem.* 89 (3), 1469–1476. doi:10.1021/acs.analchem.6b02887

Rosenbalm, K. E., Tiemeyer, M., Wells, L., Aoki, K., and Zhao, P. (2020). Glycomics-informed Glycoproteomic Analysis of Site-specific Glycosylation for SARS-CoV-2 Spike Protein. *STAR Protoc.* 1 (3), 100214. doi:10.1016/j.xpro.2020.100214

Shajahan, A., Archer-Hartmann, S., Supekar, N. T., Gleinich, A. S., Heiss, C., and Azadi, P. (2020). Comprehensive Characterization of N- and O- Glycosylation of SARS-CoV-2 Human Receptor Angiotensin Converting Enzyme 2. *Glycobiology* 31, 410–424. doi:10.1093/glycob/cwaa101

Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* 30 (12), 981–988. doi:10.1093/glycob/cwaa042

Uslupehlivan, M., and Sener, E. (2020). Glycoinformatics Approach for Identifying Target Positions to Inhibit Initial Binding of SARS-CoV-2 S1 Protein to the Host Cell. *bioRxiv*. doi:10.1101/2020.03.25.007898

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181 (2), 281–292. doi:10.1016/j.cell.2020.02.058

Wang, Y., Wu, Z., Hu, W., Hao, P., and Yang, S. (2021). Impact of Expressing Cells on Glycosylation and Glycan of the SARS-CoV-2 Spike Glycoprotein. *ACS Omega* 6 (24), 15988–15999. doi:10.1021/acsomega.1c01785

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 369 (6501), 330–333. doi:10.1126/science.abb9983

Wolfert, M. A., and Boons, G.-J. (2013). Adaptive Immune Activation: Glycosylation Does Matter. *Nat. Chem. Biol.* 9 (12), 776–784. doi:10.1038/nchembio.1403

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* 367 (6483), 1260–1263. doi:10.1126/science.abb2507

Xu, W., Wang, M., Yu, D., and Zhang, X. (2020). Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles during Global Transmission Course of COVID-19. *Front. Immunol.* 11, 565278. doi:10.3389/fimmu.2020.565278

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2. *Science* 367 (6485), 1444–1448. doi:10.1126/science.abb2762

Yang, W., Ao, M., Hu, Y., Li, Q. K., and Zhang, H. (2018). Mapping the O-Glycoproteome Using Site-specific Extraction of O-Linked Glycopeptides (EXoO). *Mol. Syst. Biol.* 14 (11), e8486. doi:10.15252/msb.20188486

Yang, W., Shah, P., Hu, Y., Toghi Eshghi, S., Sun, S., Liu, Y., et al. (2017). Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal. Chem.* 89 (21), 11193–11197. doi:10.1021/acs.analchem.7b03641

Ye, Z., Mao, Y., Clausen, H., and Vakhrushev, S. Y. (2019). Glyco-DIA: a Method for Quantitative O-Glycoproteomics with In Silico-boosted Glycopeptide Libraries. *Nat. Methods* 16 (9), 902–910. doi:10.1038/s41592-019-0504-x

Zhang, Y., Zhao, W., Mao, Y., Chen, Y., Wang, S., Zhong, Y., et al. (2020). Site-specific N-Glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins. *Mol. Cell Proteomics* 2020, 013276. doi:10.1074/mcp.RA120.002295

Zhang, Y., Xie, X., Zhao, X., Tian, F., Lv, J., Ying, W., et al. (2018). Systems Analysis of Singly and Multiply O -glycosylated Peptides in the Human Serum Glycoproteome via EThcD and HCD Mass Spectrometry. *J. Proteomics* 170, 14–27. doi:10.1016/j.jprot.2017.09.014

Zhou, D., Tian, X., Qi, R., Peng, C., and Zhang, W. (2021). Identification of 22 N-Glycosites on Spike Glycoprotein of SARS-CoV-2 and Accessible Surface Glycopeptide Motifs: Implications for Vaccination

and Antibody Therapeutics. *Glycobiology* 31 (1), 69–80. doi:10.1093/glycob/cwaa052

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

HepParser: An Intelligent Software Program for Deciphering Low-Molecular-Weight Heparin Based on Mass Spectrometry

Hui Wang[1,2†], Yu Wang[1,2†], Meijie Hou[1,2], Chunming Zhang[1,3], Yaojun Wang[4], Zhendong Guo[2,5], Dongbo Bu[1,2], Yan Li[2,5], Chuncui Huang[2,5*] and Shiwei Sun[1,2*]

[1]Key Lab of Intelligent Information Processing, State Key Lab of Computer Architecture, Big-data Academy, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Phil Rivers Technology, Beijing, China, [4]College of Information and Electrical Engineering, China Agricultural University, Beijing, China, [5]Institute of Biophysics, Chinese Academy of Sciences, Beijing, China

Low-molecular-weight heparins (LMWHs) are considered to be the most successful carbohydrate-based drugs because of their wide use as anticoagulants in clinics. The efficacy of anticoagulants made by LMWHs mainly depends on the components and structures of LMWHs. Therefore, deciphering the components and identifying the structures of LMWHs are critical to developing high-efficiency anticoagulants. However, most LMWHs are mixtures of linear polysaccharides which are comprised of several disaccharide repeating units with high similarity, making it extremely challenging to separate and decipher each component in LMWHs. Here, we present a new algorithm named hepParser to decipher the main components of LMWHs automatically and precisely based on the liquid chromatography/mass spectrometry (LC/MS) data. When tested on the general LMWH using hepParser, profiling of the oligosaccharides with different degrees of polymerization (dp's) was completed with high accuracy within 1 minute. When compared with the results of GlycReSoft on heparan sulfate samples, hepParser achieved more comprehensive and reasonable results automatically.

Keywords: LMWHs, mass spectrometry, glycosaminoglycans, isotopic distribution, computational method

INTRODUCTION

Heparin is a complex, linear polysaccharide, which belongs to the family of glycosaminoglycans (GAGs). Most heparins are comprised of ~25 disaccharide repeating units of a glucuronic acid residue (GlcA) or iduronic acid residue (IdoA) 1,4 linked to a glucosamine residue (GlcN), with various substitution patterns of sulfation at the 2-O-position of the hexuronic acid residue (HexA), the 3-O-position, the 6-O-position, and/or the N-position of GlcN, and N-acetylation at GlcN (Linhardt, 2003; Wang and Chi, 2018). Low-molecular-weight heparins (LMWHs) are derived from heparin and possess similar primary structures. Compared to heparin, the average molecular weights of LMWHs are usually between ~4,000 and ~8,000 Da, containing ~6–~12 disaccharide units (Weitz, 1997; Li et al., 2012). The decrease of average molecular weight improves LMWHs' bioavailability, including increasing *in vivo* half-life, enhancing pharmacology, changing activity profile, and reducing thrombin inhibitory activity. Owing to these improved properties, LMWHs have been widely used as clinical anticoagulants (Warkentin et al., 1995; Cohen et al., 1997; Linhardt

FIGURE 1 | Workflow of hepParser to decipher components of LMWHs.
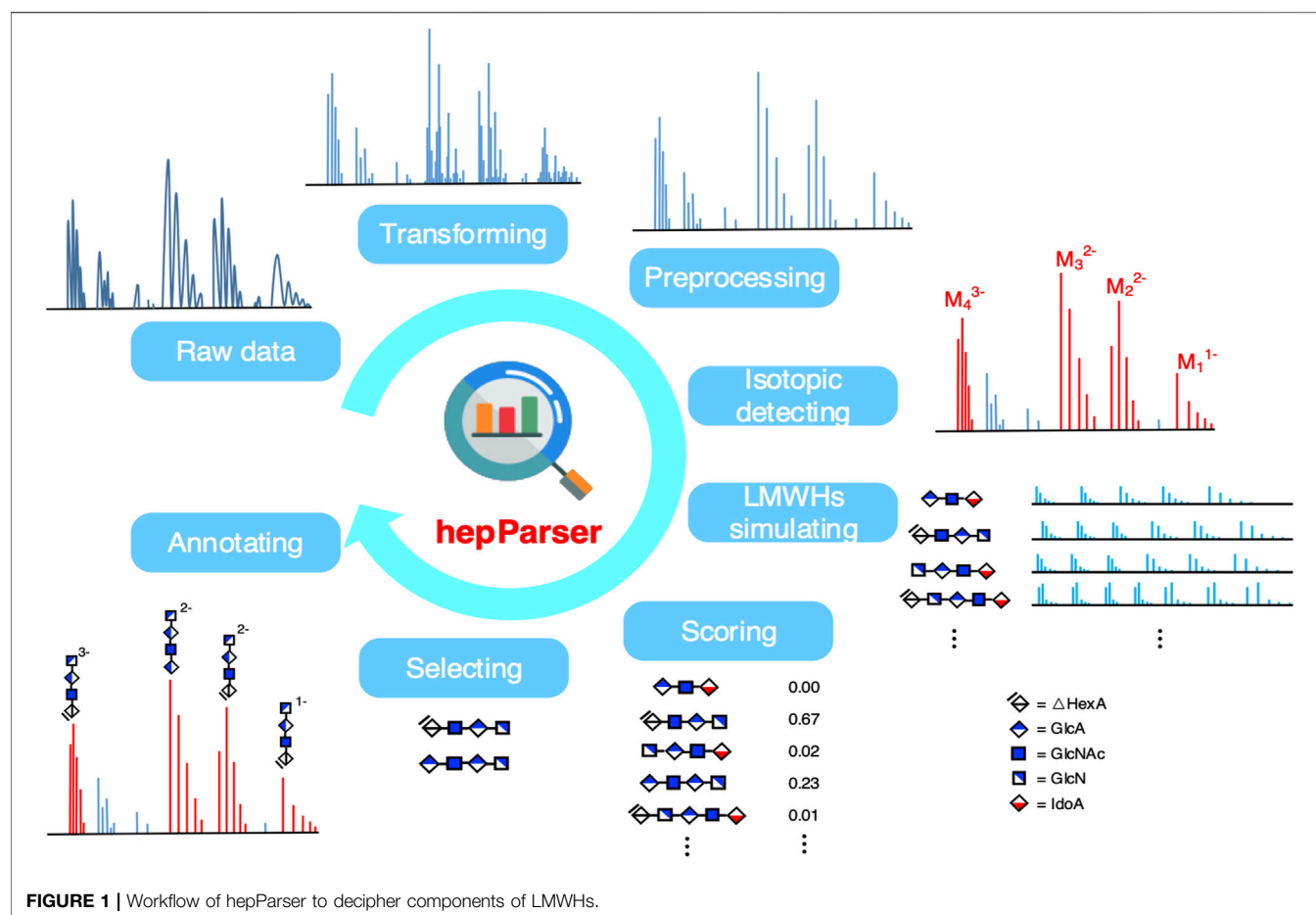
and Gunay, 1999; Norrby, 2006; Bhaskar et al., 2012). Biological functions of LMWHs are closely related to specific components and structural diversity, and therefore, deciphering components and identifying structures are critical to developing high-efficiency anticoagulants (Hemkerl, 1992). However, it is unfeasible to dissociate each component in LMWHs completely. Besides complex components, derivatives with labile sulfate loss are greatly analogous to each other and are indistinguishable (Jones et al., 2011; Kailemia et al., 2012; Duan and Jonathan Amster, 2018). The complexity and high similarity of components put forward tremendous challenges for analyzing and sequencing LMWHs.

A variety of methods have been applied for parsing LMWHs such as liquid chromatography (LC), capillary electrophoresis (CE), and size exclusion chromatography (SEC). However, these techniques provide little precise structural information on LMWHs (Pervin et al., 1995; Mao et al., 2002; Guo et al., 2003). Proton and carbon nuclear magnetic resonance (NMR) spectroscopy can present the most detailed information of the primary structure of LMWHs (Li et al., 2012). However, large amounts of samples (e.g., hundreds of micrograms) are required for NMR analysis, and high-throughput analysis and detailed

structural features cannot be achieved (Guerrini et al., 2007). Due to high sensitivity and rich structural information, mass spectrometry (MS), especially coupled to LC, has become the primary method in characterizing and sequencing LMWHs (Wolff et al., 2007; Kailemia et al., 2012; Kailemia et al., 2013). Nevertheless, peaks with multiple charges and noises due to experimental instruments or interfering impurity are usually present in mass spectra and significantly increase the difficulties in deciphering components of LMWHs.

Fortunately, some computational tools have been proposed to assist researchers in analyzing the components of LMWHs. GlycoWorkbench is a popular software program that can be used to interpret mass spectra, but automatic analysis of LMWHs' components cannot be performed (Ceroni et al., 2008; Slysz et al., 2010). Maxwell et al. developed a software program called GlycReSoft to identify and quantify heparin components based on mass spectra deconvoluted by DeconTools (Slysz et al., 2010; Maxwell et al., 2012). The software may be of low efficiency for large LC/MS datasets due to the dependence on DeconTools for deconvolution (Mechref et al., 2013). Hu et al. developed an algorithm named HS-SEQ for *de novo* sequencing of heparan sulfate samples and assigning positions of acetate and sulfate

FIGURE 2 | An example to explain the necessity and validity of peak merging. The left three windows show the original peaks before peak merging **(A)**, and the right three windows show the merged peaks after peak merging **(B)**. The blue dashed boxes represent the regions that are zoomed in.
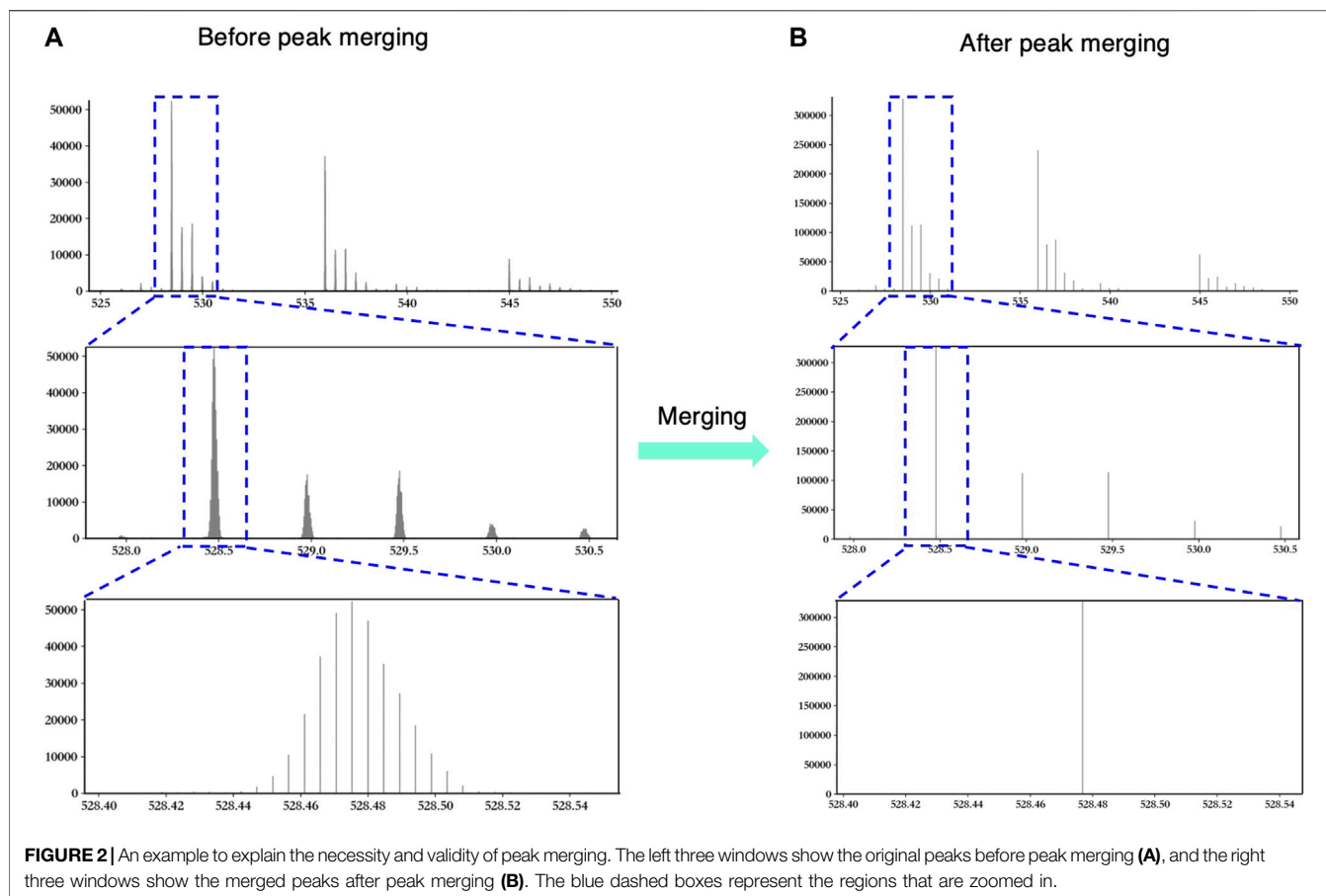
groups on the oligosaccharide chains (Hu et al., 2014). However, HS-SEQ requires mass spectra with high resolution, and conflicts of assignment may cause incorrect identification (Duan and Jonathan Amster, 2018). Chiu et al. developed GAG-ID to sequence heparin mixtures by parsing LC-MS/MS data, whereas those mixtures require complete chemical derivatization to tackle labile sulfate modifications (Chiu et al., 2015).

In this work, we developed a new algorithm named hepParser to decipher main components of LMWHs based on the LC/MS data. HepParser eliminates the interferences of noisy isotopic peak clusters and spectra shifts on profiling results using the designed peak merging and peak calibration algorithms and conducts automatic analysis of LMWHs' components with high speed and accuracy. Furthermore, with the assistance of a well-fitting model, theoretical isotopic distribution of a given mass can be produced by hepParser, and the unreliable isotopic peak clusters can be discarded based on the similarity between the experimental and theoretical isotopic distributions, which makes the profiling results more confident. HepParser has achieved an excellent performance on the tested general LMWH sample with 2–8 degrees of polymerization. We also compared the performance of hepParser with that of GlycReSoft on heparan sulfate samples. The results reported by hepParser were more comprehensive and reasonable.

MATERIALS AND METHODS

Experimental Methods

Heparin sodium (0.1%, 125 U/ml) extracted from porcine intestinal mucosa was obtained from Bioroyee (Beijing, China), and LMWHs were prepared through degradation using sodium nitrite and heparinase I. Sodium nitrite (2.95% of the Heparin) dissolved in 2.5 mol/l hydrochloric acid was added to heparin, which was then allowed to react at 35°C for 90 min. Sodium hydroxide was added to the solution to terminate the degradation (pH, 10), and sodium borohydride (1% of heparin) was injected to the solution for reduction overnight. Excess acetic acid was added to the above solution (pH, 4.0) and allowed to react for 15 min. The redundant sodium borohydride was therefore eliminated, and then 2 mol/L sodium hydroxide was added to neutralize the solution. Methanol was added to the reacted solution (67%), and the solution was stored at 4°C for 24 h. The solution was centrifuged, and the precipitates were primary LMWHs. The LMWHs were finally obtained after lyophilization using a 1,000 Da dialysis bag. Exhaustive digestion of the prepared LMWHs with heparinase I was performed at 25°C for 48 h. Briefly, 100 μl 0.2 IU/ml heparinase I dissolved in 10 mM monobasic potassium phosphate (pH = 7) was added to 50 μg LMWHs in the presence of 100 μl sodium/calcium acetate solution (pH = 7, containing 2 mM calcium acetate and 0.1 mg/ml bovine

FIGURE 3 | Profiling results of dp2 **(A)** and dp4 **(B)** oligosaccharides. The scan number represents the sampling point order in the mass spectrometry experiment. Each matched isotopic peak cluster in experimental spectra is marked in blue and annotated by the number and charge state of corresponding components. The score of each component is calculated by **Eq.1**, which indicates the number and quality of matching peaks for the component.

serum albumin). After incubation, the reacted solution was heated at 100°C for 2 min to inactivate the enzymes and further filtered on a 0.22 μm filter prior to LC/MS analysis.

LC/MS data were acquired on LMWH samples using Acquity Xevo G2-S Q-TOF UPLC/MS systems (Waters, Milford, MA). An Acquity UPLC BEH C18 column (2.1 mm × 100 mm, 1.7 μm particles) was used for chromatographic separations. The column temperature was maintained at 40 C throughout the separation, and a flow rate of 0.5 ml/min was used. 10 μl of the 0.2 mM digested LMWHs dissolved in water was injected for each separation. A binary solvent system was used for gradient elution. Solvent A was composed of 5% acetonitrile in water, and solvent B consisted of 80% acetonitrile in water. Both mobile phases contained 15 mM of pentylamine or hexylamine (PTA or HXA, ion-paring reagents) and 50 mM 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, buffering agent). MS analysis was performed on

a Waters Xevo G2-S quadrupole time-of-flight (Q-TOF) mass spectrometer equipped with an electrospray ionization (ESI) source. All the MS spectra were obtained in negative mode, and the mass range was 0–1,200 Da with a scanning rate of 0.5 s. The source temperature and the desolvation temperature were 120°C and 200°C.

Computational Methods

HepParser aims to uncover the components of LMWHs based on the MS spectra of the sample of interest. The pipeline of hepParser software is shown in **Figure 1**. In summary, hepParser would firstly perform three preprocessing steps for the given spectrum, followed by detecting the isotopic peak clusters and determining the charge of each cluster. Then, all possible components of LMWHs in a reasonable m/z range would be enumerated according to the composition rules of LMWH structures. Subsequently, the proposed

TABLE 1 | The exported detailed profiling result of dp4 oligosaccharides.

m/z	Charge	Isotopic peak	Components	Loss
303.6865	3	1	[1, 1, 2, 0, 5, 0, 0]	[2, 0, 0, 0]
325.3366	3	1	[1, 1, 2, 0, 5, 0, 0]	[0, 0, 1, 0]
325.3366	3	1	[1, 1, 2, 0, 6, 0, 0]	[1, 0, 1, 0]
330.3379	3	1	[1, 1, 2, 0, 5, 0, 0]	[1, 0, 0, 0]
330.3379	3	1	[1, 1, 2, 0, 6, 0, 0]	[2, 0, 0, 0]
356.9920	3	1	[1, 1, 2, 0, 5, 0, 0]	[0, 0, 0, 0]
356.9920	3	1	[1, 1, 2, 0, 6, 0, 0]	[1, 0, 0, 0]
408.5484	2	1	[1, 1, 2, 0, 2, 0, 0]	[0, 1, 0, 0]
410.0510	2	1	[0, 2, 2, 0, 3, 0, 0]	[0, 1, 1, 0]
416.0533	2	1	[1, 1, 2, 0, 2, 0, 0]	[0, 0, 0, 0]
416.5541	2	1	[0, 2, 2, 1, 3, 0, 0]	[0, 0, 1, 1]
425.0584	2	1	[0, 2, 2, 0, 3, 0, 0]	[1, 0, 0, 0]
448.5271	2	1	[1, 1, 2, 0, 5, 0, 0]	[1, 0, 1, 0]
456.0327	2	1	[1, 1, 2, 0, 5, 0, 0]	[2, 0, 0, 0]
456.5334	2	1	[0, 2, 2, 1, 3, 0, 0]	[0, 1, 0, 1]
465.0354	2	1	[0, 2, 2, 0, 3, 0, 0]	[0, 0, 0, 0]
488.5050	2	1	[1, 1, 2, 0, 5, 0, 0]	[0, 0, 1, 0]
488.5050	2	1	[1, 1, 2, 0, 6, 0, 0]	[1, 0, 1, 0]
490.0054	2	1	[0, 2, 2, 0, 6, 0, 0]	[0, 0, 2, 0]
496.0106	2	1	[1, 1, 2, 0, 5, 0, 0]	[1, 0, 0, 0]
496.0106	2	1	[1, 1, 2, 0, 6, 0, 0]	[2, 0, 0, 0]
496.5116	2	1	[0, 2, 2, 1, 5, 0, 0]	[0, 0, 1, 1]
497.5096	2	1	[0, 2, 2, 0, 6, 0, 0]	[1, 0, 1, 0]
505.0155	2	1	[0, 2, 2, 0, 6, 0, 0]	[2, 0, 0, 0]
505.0155	2	3	[0, 2, 2, 1, 5, 0, 0]	[1, 0, 0, 1]
528.4828	2	1	[1, 1, 2, 0, 5, 0, 0]	[0, 1, 0, 0]
528.4828	2	1	[1, 1, 2, 0, 6, 0, 0]	[0, 0, 1, 0]
529.9821	2	1	[0, 2, 2, 0, 6, 0, 0]	[0, 1, 1, 0]
535.9886	2	1	[1, 1, 2, 0, 5, 0, 0]	[0, 0, 0, 0]
535.9886	2	1	[1, 1, 2, 0, 6, 0, 0]	[1, 0, 0, 0]
536.4898	2	1	[0, 2, 2, 1, 5, 0, 0]	[0, 1, 0, 1]
536.4898	2	1	[0, 2, 2, 1, 6, 0, 0]	[0, 0, 1, 1]
537.4889	2	1	[0, 2, 2, 0, 6, 0, 0]	[0, 0, 1, 0]
544.9936	2	1	[0, 2, 2, 0, 6, 0, 0]	[1, 0, 0, 0]
544.9936	2	3	[0, 2, 2, 1, 5, 0, 0]	[0, 0, 0, 1]
544.9936	2	3	[0, 2, 2, 1, 6, 0, 0]	[1, 0, 0, 1]
547.4698	2	3	[1, 1, 2, 0, 6, 0, 0]	[0, 1, 0, 1]
555.9730	2	2	[0, 2, 2, 0, 6, 0, 0]	[0, 1, 0, 1]
575.9665	2	1	[1, 1, 2, 0, 6, 0, 0]	[0, 0, 0, 0]
576.4674	2	1	[0, 2, 2, 1, 6, 0, 0]	[0, 1, 0, 1]
577.4669	2	1	[0, 2, 2, 0, 6, 0, 0]	[0, 1, 0, 0]
584.9724	2	1	[0, 2, 2, 0, 6, 0, 0]	[0, 0, 0, 0]
584.9724	2	3	[0, 2, 2, 1, 6, 0, 0]	[0, 0, 0, 1]
1074.992	1	3	[1, 1, 2, 0, 5, 0, 0]	[0, 0, 0, 0]
1074.992	1	2	[0, 2, 2, 1, 5, 0, 0]	[0, 1, 0, 1]
1074.992	1	3	[1, 1, 2, 0, 6, 0, 0]	[1, 0, 0, 0]
1074.992	1	2	[0, 2, 2, 1, 6, 0, 0]	[0, 0, 1, 1]
1075.993	1	1	[0, 2, 2, 0, 6, 0, 0]	[0, 0, 1, 0]
1090.999	1	1	[0, 2, 2, 0, 6, 0, 0]	[1, 0, 0, 0]
1090.999	1	3	[0, 2, 2, 1, 5, 0, 0]	[0, 0, 0, 1]
1090.999	1	3	[0, 2, 2, 1, 6, 0, 0]	[1, 0, 0, 1]

score function would give a matching score for each component. At last, hepParser would greedily select components with high scores and annotate them on the original spectrum.

Data Preprocessing

To make the downstream LMWHs' profiling more precise and efficient, three preprocessing steps were carried out for the selected mass spectrum from total ion chromatograms (TICs), including peak merging, peak denoising, and peak calibration.

a) Peak Merging

In order to read and process the original MS data more conveniently, we first converted them to the "mzML" format by MSConvert (Chambers et al., 2012) with the default setting. However, this conversion may split intensive peaks into several low peaks, which would lead to the shift of peak center and the increase of meaningless matching.

To fix the over-segmentation problem, we divided all peaks into several groups according to the trend of intensity change of adjacent peaks. The intensity of peaks in each group first increased and then decreased when all peaks were sorted by m/z. Then, all peaks in the same group were merged as a new peak, the m/z value is the centroid of all these peaks' m/z, and the intensity is the sum of these peaks' intensity.

Figure 2 shows an example to explain the necessity and validity of peak merging vividly. In **Figure 2A**, the original MS data show the over-segmentation when we keep zooming in the peak. Obviously, those peaks in the third window in **Figure 2A** should be merged as one peak. As shown in **Figure 2B**, hepParser solved the over-segmentation problem after peak merging.

b) Peak Denoising

MS spectra usually contain a large quantity of noisy peaks which would interfere in the profiling results greatly. Using hepParser, two denoising methods were applied to exclude noises including instrumental noises. If a low intensity value is observed for several times (default 1,000) in one spectrum, all peaks with this intensity will be treated as the instrumental noise. Therefore, all these peaks will be filtered out. In addition, a peak will also be considered the noise when its absolute intensity is lower than the given threshold (500 in this study) or its relative intensity is lower than 0.001 (the relative intensity of the highest peaks was set as 1).

c) Peak Calibration

To correct the possible spectra's shift caused by a mass spectrometer and data preprocessing, five peaks with the highest intensity in this spectrum are selected to pre-match with all LMWHs' components. Differences in the exact relative molecular weight (MW) of each matched component and the peak's m/z are recorded. Then, the optimal shift value which minimizes the sum of all m/z differences is calculated and calibrated for the given spectrum.

Isotopic Peak Cluster Detection

The detection of isotopic peak clusters and the determination of charge state are of great significance for ESI-MS spectra analysis and have a direct influence on the accuracy of the subsequent matching process. HepParser first detected all possible isotopic clusters as the candidates and then estimated their possibilities by

FIGURE 4 | (A) Fitting results of theoretical isotopic peak distribution. **(B,C)** Comparison of peak number and time cost between unmerged spectra and merged spectra in dp2, dp4, dp6, and dp8 oligosaccharides.
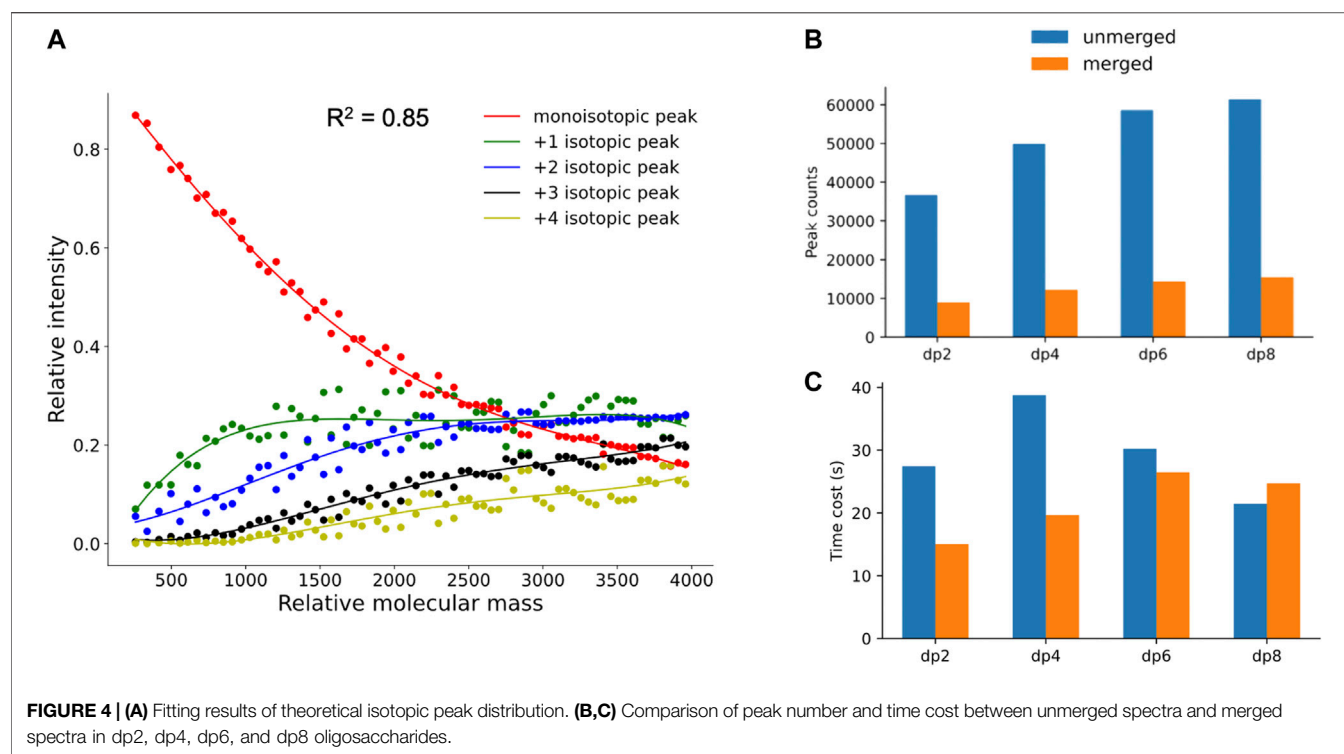
the similarity between candidates' intensity distributions and theoretical distributions.

a) Isotopic Cluster Candidate Extraction

Firstly, three parameters should be set before extraction, including the max possible charge state (default 5), the max peak number considered for each cluster (default 5), and the tolerance of peak matching (default 20 ppm). Then for each possible charge state, hepParser will search each peak in the spectrum to determine whether it is an isotopic peak cluster candidate by checking the differences in m/z between the peak and the surrounding peaks. If more than three peaks (including itself) satisfy the give tolerance, these peaks will be considered an isotopic peak cluster with the corresponding charge state. Afterward, hepParser extracted all possible isotopic peak clusters and recorded their charge states and intensity distributions.

b) Isotopic Peak Cluster Filtering

In order to obtain more reliable isotopic peak clusters, an intensity fitting model which can calculate the theoretical intensity distribution at the given m/z was trained to discard the cluster with an unreasonable intensity distribution. Firstly, numerous theoretical LWMHs' components whose relative molecular masses are distributed in a wide range were enumerated. Then, the theoretical isotopic cluster intensity distributions of possible charge states were estimated using the Brain algorithm (Dittwald et al., 2013). Afterward, we fitted the first to last isotopic peak intensity in every isotopic cluster by the

fourth-degree polynomial fitting algorithm, respectively. After calculating the theoretical cluster intensity distribution for each isotopic cluster candidate, the Jensen–Shannon divergence between the theoretical distribution and the experimental distribution was utilized to measure their qualities, and candidates with more than 0.9 similarity score (range 0–1) will be considered reliable isotopic peak clusters.

Low-Molecular-Weight Heparins' Theoretical Spectra Simulation

A comprehensive component database which contains all possible components of LMWHs was constructed by enumerating components satisfying the composition rules of LMWH structures. Each component is recorded as a tuple with seven elements, and the representation of each element is as follows: [$\triangle HexA, HexA, GlcN, Ac, SO3, Levoglucosan, Anhydromannitol$]. The simulated theoretical MS^1 spectra of each component contain all possible peaks with different charge states and isotopic peaks. In addition, considering that LMWH components may lose chemical groups (such as sulfate groups) during mass spectrometry experiments, we also simulate all possible derivatization peaks and corresponding isotopic peak clusters in theoretical MS^1 spectra of each component. The possible lost groups include $SO_3, NH, NHSO_3$, and COO. The user can control the maximum number of lost groups in one component (default 2).

Scoring

Each component is scored according to the matching of theoretical isotopic peak clusters and reliable experimental isotopic peak clusters. The score function is as follows:

FIGURE 5 | Profiling results of dp2 **(A)** and dp4 **(B)** oligosaccharides without peak merging and peak calibration. The regions marked by red rectangles contain isotopic peak clusters with low intensity which can be interpreted in **Figure 3**.
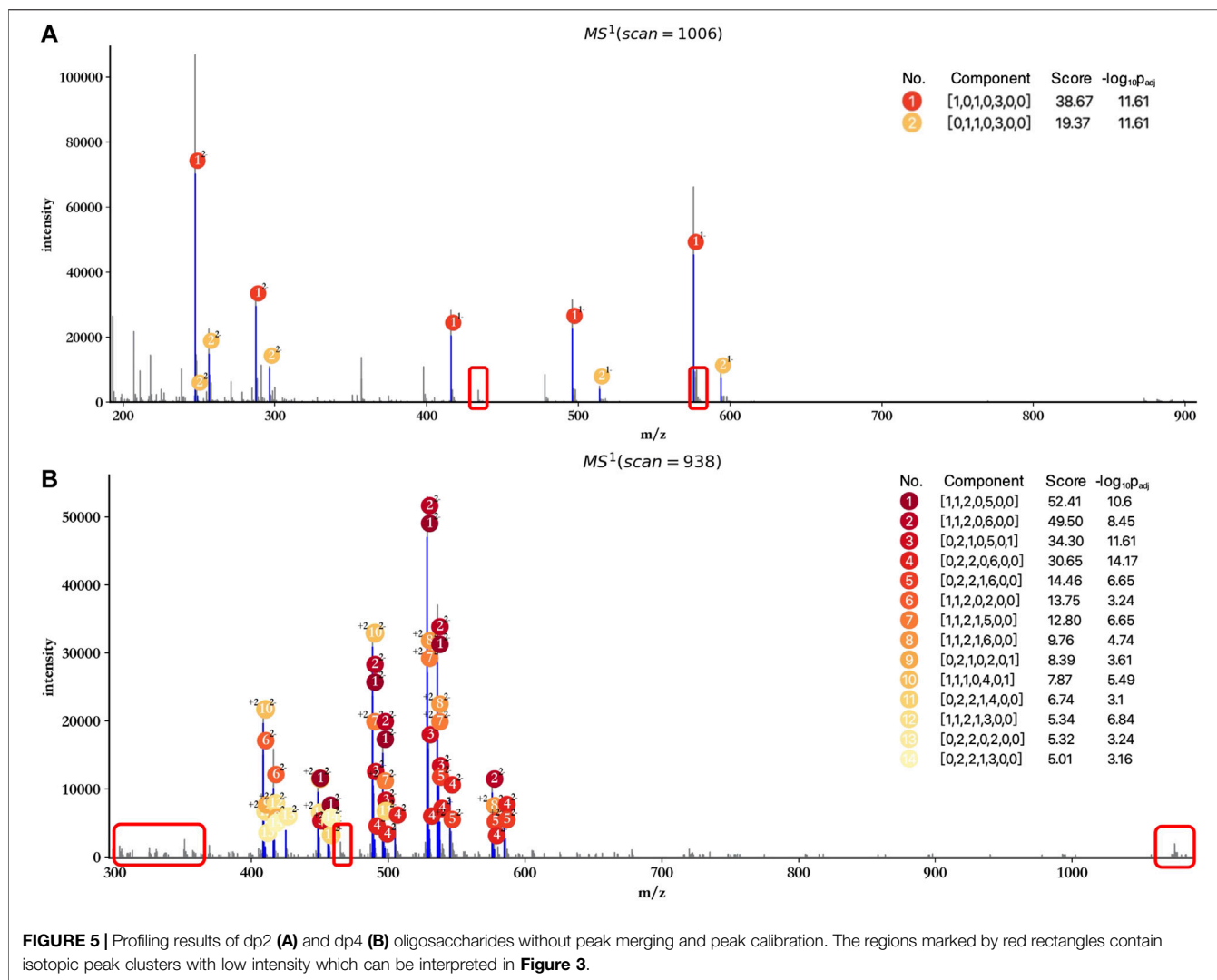
$$S_c = \sum_{p_e, p_t} \beta \ln\left(I_{p_e} + 1\right)\left(1 - JS\left(p_e, p_t\right)\right) \qquad (1)$$

,

where p_e and p_t are the matched experimental and theoretical isotopic peak clusters and I_{p_e} is the total relative intensity of p_e. The logarithmic I_{p_e} is aimed at measuring the effect of peak intensity for this score. $JS(p_e, p_t)$ represents the Jensen–Shannon divergence of p_e distribution and p_t distribution, which shows the similarity of p_e and p_t distributions. β is the derived weight coefficient to control the effect of derived isotopic peaks, which is equal to 0.9 for derived p_t or otherwise equal to 1. Obviously, the more the number of matched isotopic clusters, the more intensive the matched isotopic cluster, and the more similar the matched isotopic cluster, the higher the score for that component.

Component Selection

After scoring all enumerated LMWHs' components, hepParser will report several of them as the final results to annotate the original spectrum. The component selection should guarantee the following two constraints:

i. Most isotopic peaks in the original mass spectrum can be explained by the selected components.

ii. Each component selected needs to make sufficient and irreplaceable contributions to the interpretation of the original mass spectrum.

To determine the optimal selection for the given spectrum, a global score function which controls the number of selections and a next candidate searching function are designed as follows to guide hepParser:

$$S_g = \alpha^n \sum_{i=1}^{n} \sum_{p_e, p_t \in E_i - H} \beta \ln\left(I_{p_e} + 1\right)\left(1 - JS\left(p_e, p_t\right)\right), \qquad (2)$$

$$C_{next} = \arg\max_{i \in unselected} \left\{ \sum_{p_e, p_t \in E_i - H} \beta \ln\left(I_{p_e} + 1\right)\left(1 - JS\left(p_e, p_t\right)\right) \right\}, \qquad (3)$$
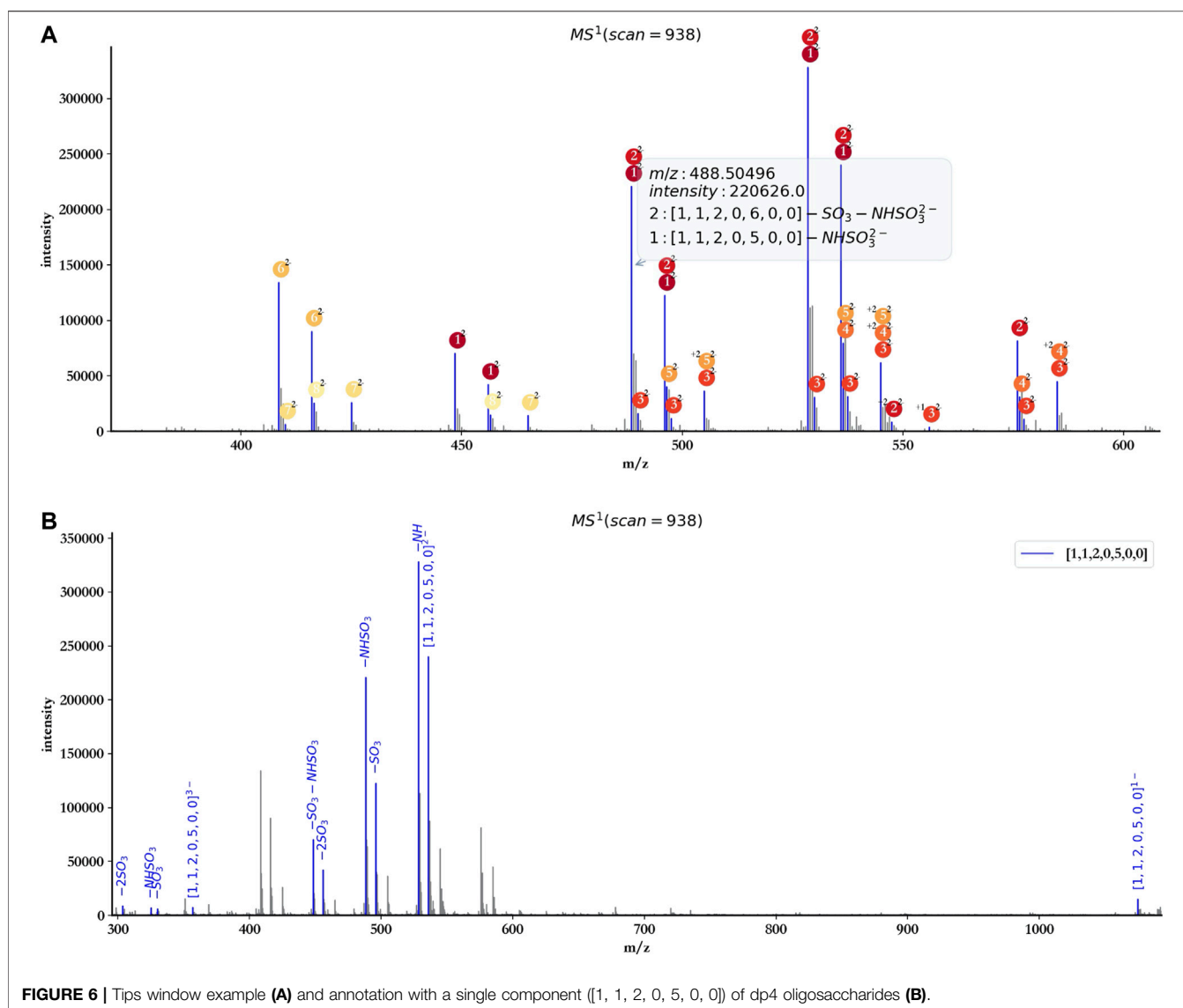
FIGURE 6 | Tips window example **(A)** and annotation with a single component ([1, 1, 2, 0, 5, 0, 0]) of dp4 oligosaccharides **(B)**.

where S_g represents the score of the current selection and n represents the number of candidates which have been selected. α is the penalty factor to control the candidates' number in the final selection (default 0.99). The specific score (noted as S_{sp}) for each i is the variant of S_C in **Eq. 2**, where E_i represents all the matched experimental isotopic peak clusters of the ith candidate and H represents all experimental isotopic peak clusters that have been explained by the first $i-1$ candidates. Therefore, only those isotopic peak clusters that cannot be matched by the first $i-1$ candidates will be calculated for the ith candidate. Meanwhile, C_{next} will greedily choose the candidate with the highest S_{sp} as the next one to be considered. Finally, the selection with the highest S_g will be considered the optimal component collection for the given spectrum.

Significance Testing

To further measure the reliability of the reported components, a test of significance was performed after component scoring and selection. The null hypothesis (H_0) is that the component is not in the sample, which means that all matching peaks of the component are accidental matches. Now, we need to calculate the probability of the above event's occurrence and obtain its p-value.

Firstly, we should calculate the probability (denoted as p_m) that one theoretical peak of the component is randomly matched by the experimental spectrum:

$$p_m = \frac{\sum_{i=1}^{n} tolerance\ of\ Tp_i}{total\ m/z\ range\ of\ E} \tag{4}$$

where "*total m/z range of E*" represents the distance from the leftmost peak to the rightmost peak of all peaks in the experimental spectrum, n represents the number of theoretical peaks of the component in the range of experimental spectra, and "*tolerance of Tp_i*" represents the size of the tolerance window of the ith theoretical peak, which can be calculated by its m/z and ppm.
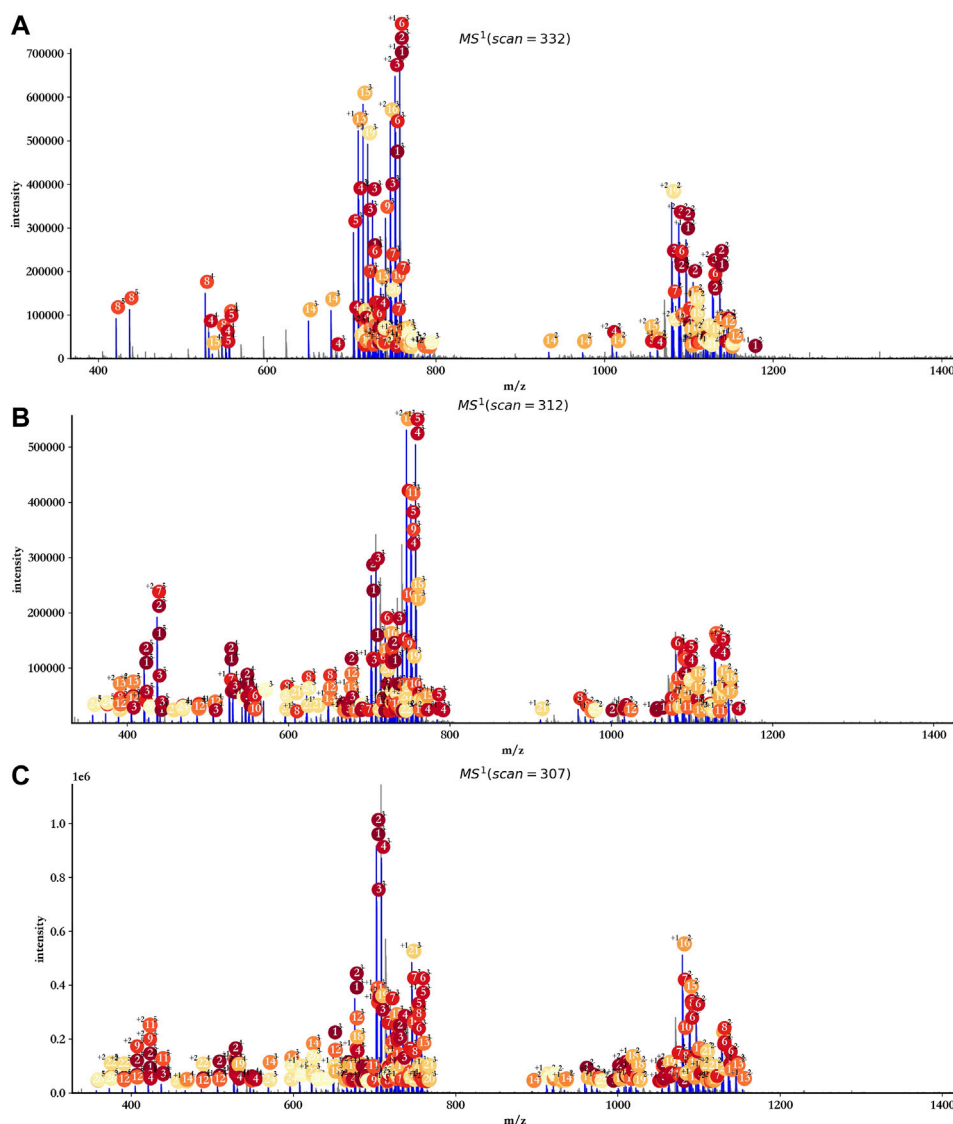
FIGURE 7 | Annotation results of triplicate HS samples using hepParser. **(A)** hn042408-01.mzML. **(B)** hn042408-03.mzML. **(C)** hn042408-06.mzML.

Suppose that the number of theoretical peaks which are randomly matched (denoted as X) follows a binomial distribution $B(n, p_m)$, the probability density function is as follows:

$$P(X = k) = \binom{n}{k} p_m^k (1 - p_m)^{n-k}. \tag{5}$$

Then, for each component with k matched peaks, the p-value can be calculated by

$$pvalue = F(X \geq k) = \sum_{i=k}^{n} P(X = i). \tag{6}$$

Because the hypothesis test with multiple comparisons may give a false-positive result, we applied the "Benjamini and Hochberg" (BH) method (Benjamini and Hochberg, 1995), one of the most commonly used methods to control the false

discovery rate (FDR), to adjust the obtained p-values. At last, we used -log10 fold change to transform the adjusted p-values (denoted as $-log_{10}p_{adj}$). According to our experience, the component with $-log_{10}p_{adj}>3$ is reliable.

Annotation
The last step for hepParser is annotating the spectrum by the components selected in step 5. Every matched isotopic peak in the experimental spectrum will be labeled by the corresponding component with the charge state and the loss derivatization information.

Open-Source Public Archive
HepParser is developed using Python 3.7, and the source code, sample data, and tutorial are available at the following GitHub website: https://github.com/Sunmile/hepParser.

FIGURE 8 | Comparison between hepParser and GlycReSoft results of triplicate HS samples. **(A)** Relationship of hepParser and GlycReSoft results. The 12 distinct components of hepParser are on the left (in blue), while the nine distinct components of GlycReSoft are on the right (in green). **(B)** Annotation results using the 12 distinct components of hepParser. **(C)** Annotation results using the nine distinct components of GlycReSoft (other seven components which matched no peak are not shown).

RESULTS AND DISCUSSION

Performance on Low-Molecular-Weight Heparin Sample

The capability of deciphering components of LMWHs with high confidence from mass spectra was the primary goal of the development of hepParser. To this aim, a series of test experiments were carried out on the LMWH sample, including profiling of HPLC fractions with dp2 and dp4.

The profiling results of dp2 and dp4 oligosaccharides are illustrated in **Figures 3A,B**. HepParser deciphered two main components in dp2 oligosaccharides which can explain most isotopic peak clusters with the consideration of losing different derivatizations and different charge states. Meanwhile, eight main components were determined in dp4 oligosaccharides and also can interpret most isotopic peak clusters, indicating the high accuracy of hepParser. In addition, less than 1 minute was required for hepParser to obtain the profiling results, which manifests the overwhelming advantage of hepParser in analysis speed compared to the manual analysis.

HepParser can export the profiling results to an editable table for downstream analysis. The exported details in the table include experimental matched m/z values, charge states, the orders of matched isotopic peaks in the clusters, the compositions of parsed LMWHs' components, and the lost groups of each matched component. Taking dp4 oligosaccharides, for instance, the exported details are listed and shown in **Table 1**.

The Fitting Curves of Theoretical Isotopic Distribution

As mentioned in *Materials and Methods*, to filter out the unreliable isotopic peak clusters, hepParser would train a fitting model with the polynomial fitting algorithm to calculate the theoretical isotopic distribution of a given mass. 70 compositions with different masses which are evenly distributed in the range of 200–4,000 Da were sampled from 6,167 possible compositions to calculate the theoretical isotopic distribution for fitting. As shown in **Figure 4A**, the fourth-degree polynomial fitting algorithm has a good performance in this fitting. The average goodness of fit (R^2) is 0.85 for the five fitting curves, which indicates that the fitting model has the ability of simulating theoretical isotopic distributions to estimate the confidence of each isotopic peak cluster candidate.

Validity for Peak Merging and Peak Calibration

To verify the necessity and validity of peak merging and peak calibration in data preprocessing, the comparisons between spectra with and without the peak merging step were conducted. As shown in **Figure 4B**, the peak numbers in spectra of oligosaccharides with different degrees of polymerization were all greatly reduced after peak merging. As illustrated in **Figure 4C**, less time was required for hepParser to decipher dp2, dp4, and dp6 oligosaccharides from the MS spectra

after peak merging. However, more time was required to decipher dp8 oligosaccharides from the MS spectra after peak merging. The main reason is that the relative amount of dp8 fraction was small, and multiple peaks with low intensity were produced in the MS spectra. It is likely that most isotopic peaks were filtered out in the MS spectra of dp8 oligosaccharides without peak merging.

In addition, attention has to be paid to the profiling performance without peak merging and peak calibration. As shown in **Figure 5**, the score and significance of most matched components decreased significantly compared to the results obtained after peak merging and peak calibration (**Figure 3**). Few monoisotopic peaks can be assigned to oligosaccharide components, and isotopic peak clusters with low intensity cannot be interpreted by any components (marked by red rectangles) when merging and calibration were not performed. Therefore, peak merging and peak calibration are crucial for LMWHs' profiling using hepParser.

Annotation With a Single Component

To facilitate the user to check the detailed information of each annotated peak, hepParser would show a tips window when the user hovered the cursor over the peak (**Figure 6A**).

The user also can annotate the spectrum with a single component using hepParser. As shown in **Figure 6B**, using the component [1, 1, 2, 0, 5, 0, 0] to annotate the spectrum of dp4 oligosaccharides, hepParser can label the matched isotopic peaks more clearly, including the information of possible lost groups, which will be useful for the downstream analysis.

Comparison With GlycReSoft

To further validate the effectiveness, we compared the results reported by hepParser with those provided by GlycReSoft on the triplicate HS sample. The version of GlycReSoft we used is 1.0, which is available at https://code.google.com/archive/p/glycresoft/downloads.

As the data provided by GlycReSoft were mainly from dp8 oligosaccharides of heparan sulfate, oligosaccharides with degree of polymerization from 7 to 9 were mainly considered by hepParser with the setting of mass accuracy at 20 ppm, and Levoglucosan and Anhydromannitol residues were ruled out. The main annotation results are shown in **Figure 7**. Almost all reliable isotopic peak clusters were interpreted by hepParser very well. The detailed matching results and reporting components can be found in **Supplementary Tables S1–S3**.

In view of the fact that GlycReSoft gives the component results of dp8 in the public sample data, we focused on the comparison of the dp8 oligosaccharides reported by hepParser and GlycReSoft. Before the comparison, we first transformed the results of hepParser to the format of GlycReSoft results. What we need to declare here is that one component of hepParser may be transformed to several components of GlycReSoft for the consideration of two possible groups lost in hepParser. For instance, [0, 4, 4, 0, 12, 0, 0] would be transformed to [0, 4, 4, 0, 12], [0, 4, 4, 0, 11], and [0, 4, 4, 0, 10]. As shown in **Figure 8A**,

hepParser deciphered 35 dp8 oligosaccharides, while GlycReSoft reported 32, in which 25 components were shared by them. The distinct components of hepParser and GlycReSoft are extracted and shown on the left (12 of hepParser) and the right (9 of GlycReSoft) in **Figure 8A**. The detailed components in this comparison can be found in **Supplementary Table S4**.

Then, we compare the ability of interpreting the experimental spectrum of two groups' distinct components and show it in **Figures 8B,C**. The 12 distinct components of hepParser can interpret amounts of peaks including some intensive peaks, and all components obtained high matching scores and high significance (**Figure 8B**), which indicated that they did exist in the sample data and the results of hepParser were reasonable. In contrast, there was no sufficient evidence to prove the existence of the nine distinct components of GlycReSoft. Only two of them can interpret five low intensity peaks (**Figure 8C**), which indicated the accuracy of hepParser to a certain extent.

CONCLUSION

An open-source software program named hepParser was developed and applied for profiling of LMWHs. The efficiency was significantly improved through the peak merging strategy, and components of LMWHs were automatically analyzed with high accuracy. The general LMWH was analyzed in this proof-of-concept study, and profiling of oligosaccharides with different degrees of polymerization was successfully performed with high speed. As known to us all, deciphering components of LMWHs based on LC/MS data often involves time-consuming manual efforts and professional prior knowledge. The developed hepParser in this study can deal with the produced data rapidly and provide components of LMWHs automatically, which should facilitate analysis and functional studies of LMWHs. Structure identification of main components of LMWHs based on multistage mass spectrometry is planned for the future.

REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bhaskar, U., Sterner, E., Hickey, A. M., Onishi, A., Zhang, F., Dordick, J. S., et al. (2012). Engineering of Routes to Heparin and Related Polysaccharides. *Appl. Microbiol. Biotechnol.* 93 (1), 1–16. doi:10.1007/s00253-011-3641-4

Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S. M. (2008). GlycoWorkbench: a Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *J. Proteome Res.* 7 (4), 1650–1659. doi:10.1021/pr7008252

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* 30 (10), 918–920. doi:10.1038/nbt.2377

Chiu, Y., Huang, R., Orlando, R., and Sharp, J. S. (2015). GAG-ID: Heparan Sulfate (HS) and Heparin Glycosaminoglycan High-Throughput Identification Software*. *Mol. Cell Proteom.* 14 (6), 1720–1730. doi:10.1074/mcp.m114.045856

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

SS and CH conceived the study. SS, HW and YW designed the approach concept and computational model. CH designed the liquid chromatography/mass spectrometry methodology. CH and ZG performed and analyzed the mass spectral data. HW, YW and MH implemented the approach. YW, DB and YL provided constructive comments and advice on the approach concept. CH and ZG carried out sample preparation and participated in MS data acquisition. HW, CH and SS wrote the manuscript. CZ provided constructive comments and advice to improve the manuscript. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.723149/full#supplementary-material

Cohen, M., Demers, C., Gurfinkel, E. P., Turpie, A. G. G., Fromell, G. J., Goodman, S., et al. (1997). A Comparison of Low-Molecular-Weight Heparin with Unfractionated Heparin for Unstable Coronary Artery Disease. *N. Engl. J. Med.* 337 (7), 447–452. doi:10.1056/nejm199708143370702

Dittwald, P., Claesen, J., Burzykowski, T., Valkenborg, D., and Gambin, A. (2013). BRAIN: A Universal Tool for High-Throughput Calculations of the Isotopic Distribution for Mass Spectrometry. *Anal. Chem.* 85 (4), 1991–1994. doi:10.1021/ac303439m

Duan, J., and Jonathan Amster, I. (2018). An Automated, High-Throughput Method for Interpreting the Tandem Mass Spectra of Glycosaminoglycans. *J. Am. Soc. Mass. Spectrom.* 29 (9), 1802–1811. doi:10.1007/s13361-018-1969-z

Guerrini, M., Guglieri, S., Naggi, A., Sasisekharan, R., and Torri, G. (2007). Low Molecular Weight Heparins: Structural Differentiation by Bidimensional Nuclear Magnetic Resonance Spectroscopy. *Semin. Thromb. Hemost.* 33, 478–487. doi:10.1055/s-2007-982078

Guo, X., Condra, M., Kimura, K., Berth, G., Dautzenberg, H., and Dubin, P. L. (2003). Determination of Molecular Weight of Heparin by Size Exclusion Chromatography with Universal Calibration. *Anal. Biochem.* 312 (1), 33–39. doi:10.1016/s0003-2697(02)00428-1

Hemkerl, H. C. (1992). On the Relationship between Molecular Mass and Anticoagulant Activity in a Low Molecular Weight Heparin (Enoxaparin). *Thromb. Haemost.* 67 (5), 556–562. doi:10.1055/s-0038-1648493

Hu, H., Huang, Y., Mao, Y., Yu, X., Xu, Y., Liu, J., et al. (2014). A Computational Framework for Heparan Sulfate Sequencing Using High-Resolution Tandem Mass Spectra. *Mol. Cell Proteom.* 13 (9), 2490–2502. doi:10.1074/mcp.m114.039560

Jones, C. J., Beni, S., Limtiaco, J. F. K., Langeslay, D. J., and Larive, C. K. (2011). Heparin Characterization: Challenges and Solutions. *Annu. Rev. Anal. Chem.* 4, 439–465. doi:10.1146/annurev-anchem-061010-113911

Kailemia, M. J., Li, L., Ly, M., Linhardt, R. J., and Amster, I. J. (2012). Complete Mass Spectral Characterization of a Synthetic Ultralow-Molecular-Weight Heparin Using Collision-Induced Dissociation. *Anal. Chem.* 84 (13), 5475–5478. doi:10.1021/ac3015824

Kailemia, M. J., Li, L., Xu, Y., Liu, J., Linhardt, R. J., and Amster, I. J. (2013). Structurally Informative Tandem Mass Spectrometry of Highly Sulfated Natural and Chemoenzymatically Synthesized Heparin and Heparan Sulfate Glycosaminoglycans. *Mol. Cell Proteomics* 12 (4), 979–990. doi:10.1074/mcp.m112.026880

Li, L., Zhang, F., Zaia, J., and Linhardt, R. J. (2012). Top-down Approach for the Direct Characterization of Low Molecular Weight Heparins Using LC-FT-MS. *Anal. Chem.* 84 (20), 8822–8829. doi:10.1021/ac302232c

Linhardt, R. J., and Gunay, N. S. (1999). Production and Chemical Processing of Low Molecular Weight Heparins. *Semin. Thromb. Hemost.* 25 (3), 5–16.

Linhardt, R. J. (2003). 2003 Claude S. Hudson Award Address in Carbohydrate Chemistry. Heparin: Structure and Activity. *J. Med. Chem.* 46 (13), 2551–2564. doi:10.1021/jm030176m

Mao, W., Thanawiroon, C., and Linhardt, R. J. (2002). Capillary Electrophoresis for the Analysis of Glycosaminoglycans and Glycosaminoglycan-Derived Oligosaccharides. *Biomed. Chromatogr.* 16 (2), 77–94. doi:10.1002/bmc.153

Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., et al. (2012). GlycReSoft: a Software Package for Automated Recognition of Glycans from LC/MS Data. *PloS one* 7 (9), e45474. doi:10.1371/journal.pone.0045474

Mechref, Y., Hu, Y., Desantos-Garcia, J. L., Hussein, A., and Tang, H. (2013). Quantitative Glycomics Strategies. *Mol. Cell Proteom.* 12 (4), 874–884. doi:10.1074/mcp.r112.026310

Norrby, K. (2006). Low-molecular-weight Heparins and Angiogenesis. *APMIS* 114 (2), 79–102. doi:10.1111/j.1600-0463.2006.apm_235.x

Pervin, A., Gallo, C., Jandik, K. A., Han, X.-J., and Linhardt, R. J. (1995). Preparation and Structural Characterization of Large Heparin-Derived Oligosaccharides. *Glycobiology* 5 (1), 83–95. doi:10.1093/glycob/5.1.83

Slysz, G. W., Baker, E. S., Shah, A. R., Jaitly, N., Anderson, G. A., and Smith, R. D. (2010). "The DeconTools Framework: an Application Programming Interface Enabling Flexibility in Accurate Mass and Time Tag Workflows for Proteomics and Metabolomics," in Proceeding of the 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake, UT, May 23–27, 2010.

Wang, Z., and Chi, L. (2018). Recent Advances in Mass Spectrometry Analysis of Low Molecular Weight Heparins. *Chin. Chem. Lett.* 29 (1), 11–18. doi:10.1016/j.cclet.2017.08.050

Warkentin, T. E., Levine, M. N., Hirsh, J., Horsewood, P., Roberts, R. S., Gent, M., et al. (1995). Heparin-induced Thrombocytopenia in Patients Treated with Low-Molecular-Weight Heparin or Unfractionated Heparin. *N. Engl. J. Med.* 332 (20), 1330–1336. doi:10.1056/nejm199505183322003

Weitz, J. I. (1997). Low-molecular-weight Heparins. *N. Engl. J. Med.* 337 (10), 688–698. doi:10.1056/nejm199709043371007

Wolff, J. J., Amster, I. J., Chi, L., and Linhardt, R. J. (2007). Electron Detachment Dissociation of Glycosaminoglycan Tetrasaccharides. *J. Am. Soc. Mass. Spectrom.* 18 (2), 234–244. doi:10.1016/j.jasms.2006.09.020

Comparison of Different Labeling Techniques for the LC-MS Profiling of Human Milk Oligosaccharides

Yinzhi Lang[1†], Yongzhen Zhang[1], Chen Wang[1], Limei Huang[1], Xiaoxiao Liu[1], Ni Song[1], Guoyun Li[1,2]* and Guangli Yu[1,2]*

[1]Key Laboratory of Marine Drugs, Ministry of Education, Shandong Provincial Key Laboratory of Glycoscience and Glycotechnology, School of Medicine and Pharmacy, Ocean University of China, Qingdao, China, [2]Laboratory for Marine Drugs and Bioproducts, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Human milk oligosaccharides (HMOs) exhibit various biological activities for infants, such as serving as prebiotics, blocking pathogens, and aiding in brain development. HMOs are a complex mixture of hetero-oligosaccharides that are generally highly branched, containing multiple structural isomers and no intrinsic chromophores, presenting a challenge to both their resolution and quantitative detection. While liquid chromatography-mass spectrometry (LC-MS) has become the primary strategy for analysis of various compounds, the very polar and chromophore-free properties of native glycans hinder their separation in LC and ionization in MS. Various labeling approaches have been developed to achieve separation of glycans with higher resolution and greater sensitivity of detection. Here, we compared five commonly used labeling techniques [by 2-aminobenzamide, 2-aminopyridine, 2-aminobenzoic acid (2-AA), 2,6-diaminopyridine, and 1-phenyl-3-methyl-5-pyrazolone] for analyzing HMOs specifically under hydrophilic-interaction chromatography-mass spectrometry (HILIC-MS) conditions. The 2-AA labeling showed the most consistent deprotonated molecular ions, the enhanced sensitivity with the least structural selectivity, and the sequencing-informative tandem MS fragmentation spectra for the widest range of HMOs; therefore, this labeling technique was selected for further optimization under the porous graphitized carbon chromatography-mass spectrometry (PGC-MS) conditions. The combination strategy of 2-AA labeling and PGC-MS techniques provided online decontamination (removal of excess 2-AA, salts, and lactose) and resolute detection of many HMOs, enabling us to characterize the profiles of complicated HMO mixtures comprehensively in a simple protocol.

Keywords: human milk oligosaccharides, labeling, 2-aminobenzoic acid, HILIC-MS, PGC-MS

INTRODUCTION

Human milk oligosaccharides (HMOs) have been assigned to a variety of important biological functions for infants (Newburg, 2013; Bode, 2015; Wicinski et al., 2020), such as prevention of pathogens binding to epithelial cell surfaces (Hester et al., 2013; Gonia et al., 2015; Triantis et al., 2018; Ha et al., 2020), functioning as prebiotics (Boehm et al., 2005; Marcobal and Sonnenburg, 2012; Underwood et al., 2015; Walsh et al., 2020), and enhancing brain development (Oliveros et al., 2018;

Docq et al., 2020; Wu et al., 2020). Characterizing the structure(s) and concentration(s) of the isolated specific HMO(s) is essential to elucidate its (or their) molecular mechanisms involved in *in vitro* studies. For understanding the difference of HMOs in different populations and the changes of HMOs at different lactations, component analysis of a whole HMO mixture is necessary. However, HMOs are hetero-oligosaccharides composed of different glycosidic-linked glucose, galactose, *N*-acetyl-glucosamine, fucose, and *N*-acetyl-neuraminic acid residues (Kunz et al., 2000; Bode, 2006), where the five monosaccharides can be linked in various ways through at least 12 α- and β-glycosidic linkages, resulting in a structurally complex array of linear, branched, and isomeric structures (Bode, 2012; Smilowitz et al., 2014). The structural diversity and complexity present a challenge to their resolute separation and detection.

Over the years, many analytical techniques have been applied to analyze HMOs as reviewed previously (Mantovani et al., 2016). Nuclear magnetic resonance spectroscopy (NMR) and off-line mass spectrometry (MS) have previously been the major methods for HMO structural elucidations (Chai et al., 2005). High-pH anion-exchange chromatography (HPAEC) coupled with pulsed amperometric detection (PAD) required the preliminary sample preparation to remove any other biomolecules (Thurl et al., 1997; Coppa et al., 1999). Liquid chromatography (LC) and capillary electrophoresis (CE) coupled with UV adsorption or laser-induced fluorescence (LIF) detection have previously been the major methods for HMO (relatively) quantification analysis by tagging the native oligosaccharides with chromophores and fluorophores (Ruhaak et al., 2010). Even though these methods permit higher UV sensitivity or fluorescence detection, they do not provide specific structural information, and this is a key point due to the highly complex nature of milk glycan mixtures. Furthermore, these methods are severely limited by the small number of HMO standards commercially available. As a consequence, the combination of analytical separation (e.g., LC) and MS has been extensively and successfully used for oligosaccharide compositional profiling analysis in various milk samples (Tao et al., 2008; Tao et al., 2009; Wu et al., 2010; Wu et al., 2011). Since the native glycan analysis was ambiguous due to the separation of anomers under LC conditions, Carlito B. Lebrilla and his colleagues employed a mild reduction reaction to transform the native oligosaccharides to their alditol forms to simplify the chromatograms and determined many oligosaccharides in bovine milk and human milk samples (Tao et al., 2008; Tao et al., 2009; Wu et al., 2010; Wu et al., 2011). This method allows identification and relatively quantification based on the MS signal; however, the very polar hydrophobic groups or the lack of hydrophobic groups (e.g., chromophores, fluorophores, etc.) is always associated with a weak volatility during electrospray ionization (ESI), especially for some large oligosaccharides, which may thus lead to insufficient ionization of such oligosaccharide alditols in MS. Therefore, future research is needed to explore suitable labeling approaches for HMOs to specifically enhance the analytical performance of LC-MS techniques (Ruhaak et al., 2010).

As mentioned above, a few labeling approaches have previously been developed for optical detection of HMOs when coupled with CE or LC analytical separations, for example, labeling the HMOs with 2-aminoacridone (2-AMAC), 1-phenyl-3-methyl-5-pyrazolone (PMP), 2-aminopyridine (2-AP), 2-aminobenzoic acid (2-AA), 2-aminobenzamide (2-AB), 8-aminopyrene-1,3,6-trisulfonate (APTS), etc. (Song et al., 2002; Domann et al., 2007; Asakuma et al., 2008; Leo et al., 2010; Mariño et al., 2011). The most common labeling reaction employed is reductive amination, which can attach a fluorescence label containing amino groups to the anomeric center of a native glycan's reducing terminus. It is usually performed in a dimethyl sulfoxide solution containing acetic acid (or citric acid), for example, 2-AB, 2-AA, 2,6-diaminopyridine (DAP), AEAB, 2-AMAC, 2-AP, APTS, etc. (Anumula, 1994; Bigge et al., 1995; Morelle et al., 2005; Xia et al., 2005; Anumula, 2006; Song et al., 2009; Galeotti et al., 2012; Kishimoto et al., 2020; Kinoshita et al., 2021). 2-AB and 2-AP have been widely applied in hydrophilic-interaction chromatography and mass spectrometry (HILIC-MS) profiling, and databases for structural assignments based on standardized elution positions have been developed (Morelle et al., 2005; Takegawa et al., 2005; Royle et al., 2008; Kozak et al., 2015). DAP, with two aromatic amine groups, has been developed as a versatile tag for the combination of structural characterization and biomolecular interaction analysis in glycan array fabrication studies (Xia et al., 2005). Anumula developed an alternative mild reaction for 2-AA specifically via acetate-borate-buffered methanol solution (Anumula, 2006). 2-AA bears one negative charge specifically, showing versatility in CE separations as well as matrix-assisted laser desorption/ionization (MALDI) analysis. 2-AA derivatization was reported to allow for simultaneous MALDI analysis of neutral and acidic N-glycans (Anumula and Dhume, 1998). However, whether such benefit remains for its application to ESI-MS analysis of glycans has rarely been discussed. Another widely employed derivatization reaction, Michael addition for conjugation of PMP (without the primary amine group), was performed under alkaline conditions (Strydom, 1994; You et al., 2008; Wang et al., 2018). PMP labeling shows high sensitivity by UV absorption and has been expected to reduce the risk of losing sialic acids (Saba et al., 1999; Wang et al., 2013). These labeling approaches have also been widely applied to *N*-glycan analysis by improving the optical detection and furthermore have been reported to be capable of increasing the sensitivity of MS detection (Anumula, 1994; Bigge et al., 1995; Morelle et al., 2005; Xia et al., 2005; Anumula, 2006; Song et al., 2009; Galeotti et al., 2012; Kishimoto et al., 2020; Kinoshita et al., 2021).

To allow compatibility with the LC-MS system, the reaction mixtures usually need to be purified from excess salts and fluorescent tags by solid-phase extraction (SPE) (Redmond and Packer, 1999; Zhang et al., 2014) or liquid-liquid extraction (Yuen et al., 2002). While this process increases glycan purity and facilitates in-depth qualification analysis by following the LC-MS technique, it may introduce a selective loss of glycans for samples that are a mix of structurally diverse glycans (Blank et al., 2011). For the LC separation, HILIC has been widely utilized to separate the reductive amination-derivatized
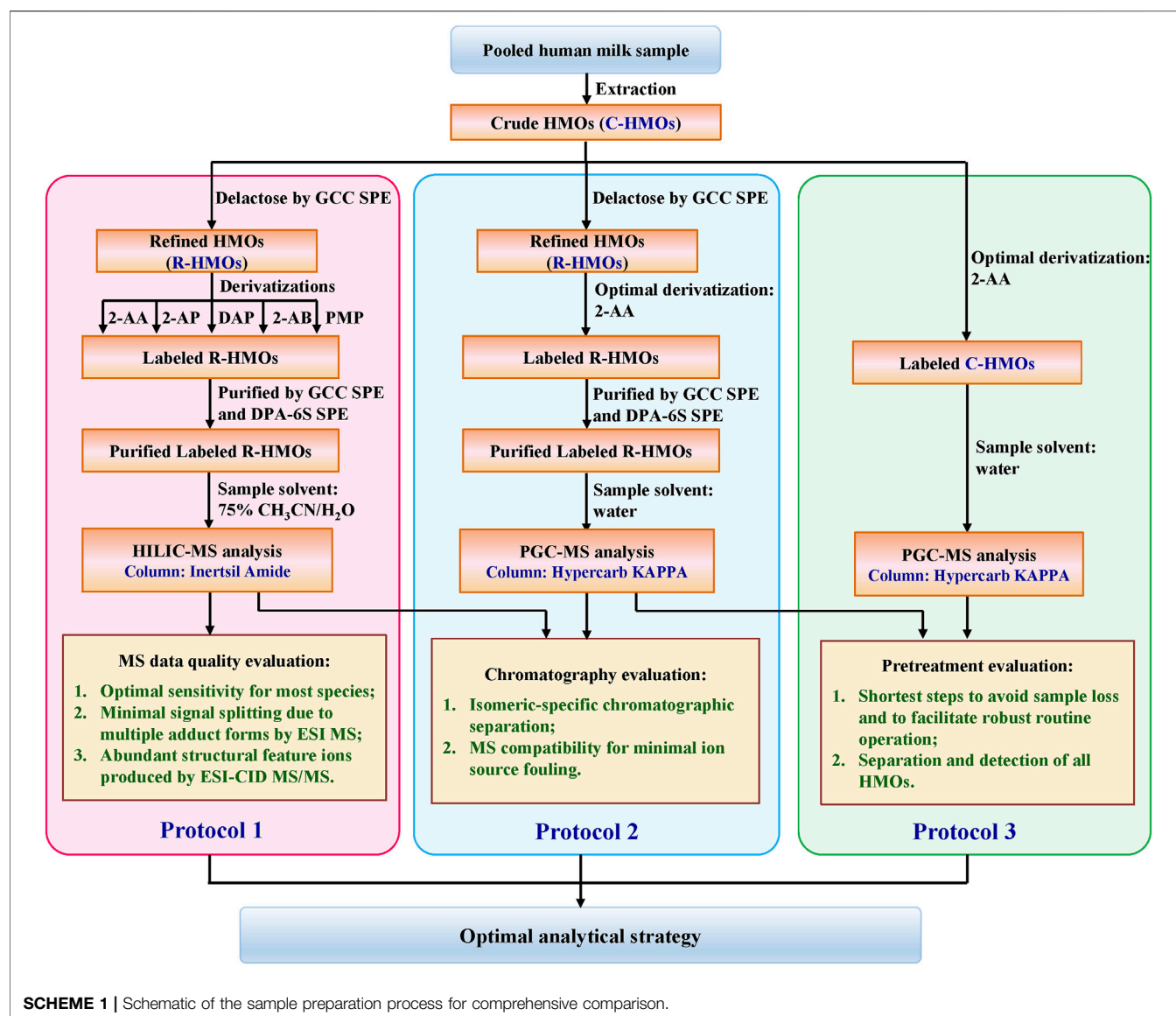
SCHEME 1 | Schematic of the sample preparation process for comprehensive comparison.

glycans (Ahn et al., 2010; Lauber et al., 2015; Cesla et al., 2016; Kinoshita et al., 2021), and reversed-phase chromatography (RPC) has been employed to separate the PMP-derivatized glycans (Melmer et al., 2011; Zauner et al., 2012). Surprisingly, porous graphitic carbon chromatography (PGC) has rarely been combined with chromophore-derivatization methods for glycan analysis, although it has been widely applied to LC-MS analysis of HMO reduced alditols (Tao et al., 2008; Tao et al., 2009; Wu et al., 2010; Wu et al., 2011). For the MS detection, David J. Harvey systematically studied the effect of different derivatives of six purchased N-glycans on electrospray ionization (ESI) sensitivity and collision-induced dissociation (CID) fragmentation using an LC-offline Q-TOF mass spectrometer (Harvey, 2000). The in-line LC-MS mode provided powerful resolution and quantitative detection for compounds; however, the components of injected analytes and

mobile phases need to be considered as vital factors that can carry the interfered sodium adduct ions which would result in the varying detection sensitivity and identification ambiguity.

Several reviews systematically evaluated and compared the influence of different N-glycan derivatives on the MS performance at both off-line or in-line modes (Suzuki et al., 1996; Harvey, 2000; Saba et al., 2001; Gao et al., 2003; Lattova et al., 2005; Pabst et al., 2009; De Leoz et al., 2020), and many conclusive results were obtained, providing useful guidance on selection of suitable analytical strategies for N-glycan analysis. For analysis of HMOs, a number of analytical techniques have been developed, and several reviews provided comparative evaluation on the quantification analysis of HMOs (Ninonuevo et al., 2006; Bao and Newburg, 2008; Mantovani et al., 2016; Grabarics et al., 2017; Tonon et al., 2019; van Leeuwen, 2019; Auer et al., 2021). Regarding the compositional profiling analysis of HMO subjects,

however, comparative studies that focused on LC-MS-based HMO labeling strategies are limited (Mariño et al., 2011). This study aimed to compare how five different widely used labeling techniques (2-AA, 2-AB, 2-AP, DAP, and PMP), two sample pretreatment procedures (SPE and none), and two LC-MS techniques (HILIC-MS and PGC-MS) can influence the MS performance and the compositional profiling analysis results of a HMO mixture. This fundamental study will support future research to discover HMO biomarkers relevant to infant-protective functions and to elucidate the maternal expression dynamics of HMOs associated with the lactation cycle.

MATERIALS AND METHODS

Chemicals and Reagents

Mature milk samples from 13 healthy lactating women volunteers were pooled as one sample (Qingdao, Shandong, China). This study was carried out in accordance with the recommendations of the Scientific Ethics Special Committee guidelines of Ocean University of China. The labels (2-AB, 2-AP, 2-AA, DAP, and PMP) and reducing agents (sodium cyanoborohydride, NaCNBH₃, and sodium borohydride, NaBH₄) were purchased from Sigma-Aldrich (Vienna, Austria). Acetonitrile, ammonium formate, ammonium acetate, ammonium bicarbonate, sodium acetate, formic acid, acetic acid, trifluoroacetic acid, dimethyl sulfoxide, and ammonia were obtained from Merck (Darmstadt, Germany). Water was prepared using a Milli-Q® system (Millipore, MA, United States). The self-packed SPE tubes: the bottom frit was loaded to the bottom of the cartridge using a push rod, filling the sorbent, and the top frit was loaded to the top of the cartridge using a push rod. Sorbents: graphitized carbon cartridges (GCCs, 300 mg of bed weight, a 7 ml cartridge volume) were purchased from Grace Davison (IL, United States), and polyamide S6 cartridges (Discovery® DPA-6S, 300 mg, a 7 ml cartridge volume) were purchased from Sigma-Aldrich (MO, United States). Prior to separation, cartridges were activated according to the manufacturer's protocol.

Extraction of HMOs

A crude mixture of HMOs was prepared from the pooled human milk sample as described previously (Ward, 2009). Briefly, 2 ml of milk was centrifuged at 4,500 g at 4°C for 30 min, and the majority of the fatty layer (upper layer) was discarded. 4 ml of ethanol was then added to the defatted milk (bottom layer), vortexed, and kept at 4°C for overnight. The insoluble protein-rich precipitate was removed by centrifugation at 4,500 g at 4°C for 10 min. The oligosaccharide-rich fraction (top layer) was dried *in vacuo* and referred to as C-HMOs.

A refined mixture of HMOs was isolated from the residual peptides and the high amount of lactose by SPE using GCC (Ward, 2009; Blank et al., 2011). Briefly, the C-HMOs were reconstituted in water, loaded on the GCC, and then washed with water containing 0.1% trifluoroacetic acid to remove lactose and salts. HMOs were eluted with 40% acetonitrile/water (v/v) (containing 0.1% trifluoroacetic acid). The pooled eluent was lyophilized and referred to as R-HMOs.

Derivatization of HMOs

Labeling by reductive amination. 2-AB, 2-AP, and DAP derivatives were essentially prepared by the method described by J. C. Bigge et al. (1995). Dried C-HMOs or R-HMOs (1 mg) were dissolved in anhydrous 70% dimethyl sulfoxide: 30% acetic acid (100 μl), followed by NaCNBH₃ and the labeling reagent (2-AB, 2-AP, or DAP) being added to give final concentrations of 63 mg/ml and 50 mg/ml, respectively. The mixture was heated for 2 h at 65°C. Derivatization of HMOs with 2-AA was carried out as described previously (Anumula, 2014). A solution of 4% sodium acetate (w/v) and 2% boric acid (w/v) in methanol was prepared first. The derivatization reagent was freshly prepared by dissolving 63 mg/ml of NaCNBH₃ and 44 mg/ml of 2-AA in 1.0 ml of the above methanol-sodium acetate-borate solution. The HMOs (1 mg in 20 μl of water) were mixed with 100 μl of the above 2-AA solution for 1 h at 80°C.

Labeling by Michael addition. C-HMOs or R-HMOs (1 mg) were dissolved in 0.3 M NaOH in water (50 μl), mixed with a methanolic solution of PMP (0.5 M, 50 μl), and incubated for 1 h at 70°C (Strydom, 1994). After cooling, the mixture was neutralized by adding 50 μl of HCl solution (0.3 M).

Derivatization by reduction. C-HMOs or R-HMOs (1 mg) were reduced to their alditol forms using 1.0 M NaBH₄ in water (100 μl) and incubated for 1.5 h at 65°C (Lattova et al., 2005). This was designed as a reference method without introducing chromophores.

For derivatizations of C-HMOs, reactions were terminated with 1 ml of water, lyophilized, and used directly for LC-MS analysis.

For derivatizations of R-HMOs, the reaction mixtures were purified by one or two SPE processes prior to LC-MS analysis. The reference R-HMO alditols were isolated from excess salts by GCC SPE, as described for removal of lactose above. For 2-AB-, 2-AP-, DAP-, 2-AA-, and PMP-derivatized HMOs, we employed a tandem SPE method: DPA-6S SPE for removing excess hydrophobic labels and GCC SPE for desalting. The labeling reaction mixture was diluted with 1 ml of 95% acetonitrile/water (v/v), loaded on the DPA-6S cartridge, and washed with the same solvent, and then, labeled HMOs (probably with residual salts) were eluted with 20% acetonitrile/water (v/v). This eluent was then loaded to GCC SPE to remove salts. Briefly, the eluent was dried and reconstituted in 1 ml of water, loaded on the GCC, and then washed with water containing 0.1% trifluoroacetic acid to remove salts. HMOs were eluted with 40% acetonitrile/water (v/v) (containing 0.1% trifluoroacetic acid). Samples were lyophilized before LC-MS analysis.

LC-MS Analysis of HMO Derivatives

All experiments were performed using a 1,260 series capillary LC system (Agilent Technologies, Inc.) coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific). Conditions for ionic strength, pH, and the buffer system were optimized to achieve resolute elution behavior for LC and sensitive ionization efficiency for MS.

The applied ES voltage for both negative and positive modes was 3.0 kV, with a capillary temperature of 275°C. The capillary voltage was set at −41 V with the tube lens voltage of −120 V in the negative mode and +18 V with the tube lens voltage of +95 V
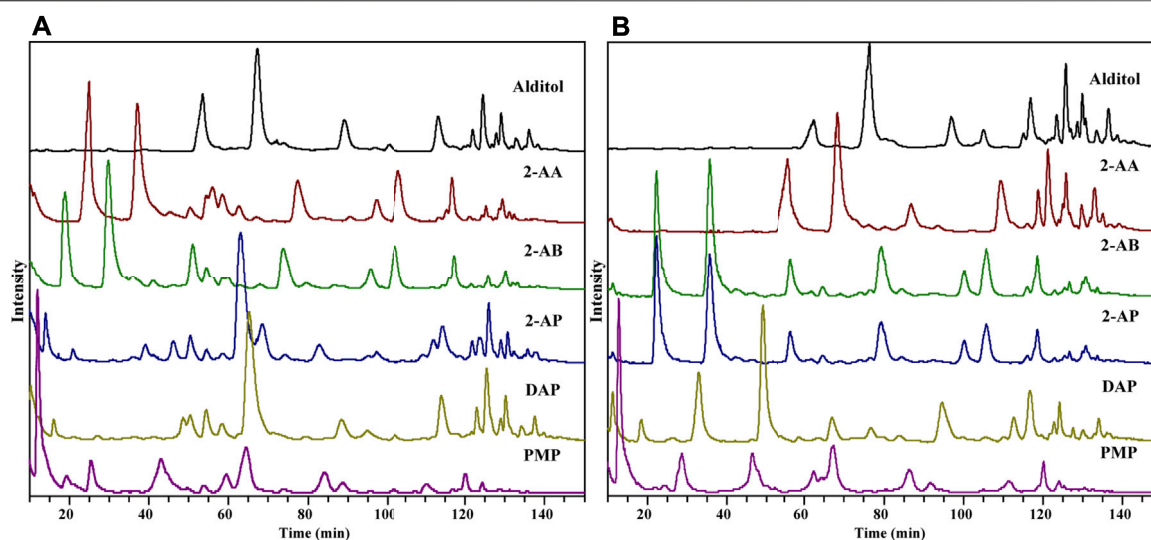
FIGURE 1 | TIC of HMO derivatives by HILIC-MS in the positive-ion **(A)** and negative-ion **(B)** modes.

in the positive-ion mode. The instrument was operated in the Fourier transform mode with m/z ranging from 200 to 3,000 Da. The full MS scan (at a resolution of 60,000) was followed by a data-dependent MS/MS scan of the five most abundant ions in the ion trap. For MS/MS in the ion trap, the normalized collision energy was set to 25 arbitrary units with an automated gain control target value of 1×10^4.

An Inertsil® Amide column (150 × 0.5 mm, 3 μm, GL Sciences) was used for the HILIC separation. The buffer system for HILIC-MS was optimized using 10 mM ammonium formate to reduce glycan peaks' overlapping and minimize ionization suppression. Solvent A consisted of 10 mM ammonium formate in 40% acetonitrile/water (v/v), and solvent B was 90% acetonitrile/ water (v/v) containing 10 mM ammonium formate. Both solvents were buffered to pH 4.0 with formic acid for the positive mode or buffered to pH 7.0 with ammonium for the negative mode. A rather flat gradient was delivered at 15 μl/min after a loading time of 25 min at 40°C: 10–17% A (25–45 min), 17–23%A (45–85 min), 23–30% A (85–105 min), and 30–80% A (105–145 min). Samples were reconstituted in 50 μl of 75% acetonitrile/water. The injection volume was 0.1 μl for all samples.

PGC separation was performed on a Hypercarb™ KAPPA column (100 × 0.5 mm, 3 μm, Thermo Scientific) at a flow rate of 8 μl/min. PGC-MS was operated in the negative-ion mode, employing ammonium bicarbonate as the optimal mobile phase buffer. For each sample, 0.1 μl (dissolved in 50 μl of water) was loaded onto the analytical column and eluted with a binary solvent consisting of A (10 mM ammonium bicarbonate, pH 8.0) and B (acetonitrile) at 8 μl/min at 35°C. An optimized glycan gradient elution was used for the separation of glycan mixtures as follows: 2–10% B (0–20 min), 10%–16%B (20–110 min), and 16–30% B (110–150 min).

Data Analysis

Raw LC/MS data were auto-processed into lists of neutral masses and abundances using the DeconTools software (Jaitly et al.,

2009). GlycResoft combines the raw neutral mass peaks into compounds, correcting mass spectrometric adducts (Maxwell et al., 2012). The program scores the data, generates a list of candidate glycan compositions, and matches these against the compound list. Putative structures can be assigned based on known human milk oligosaccharide patterns: glycan compositions containing hexose (Hex), N-acetylhexosamine (HexNAc), fucose (Fuc), N-acetyneuraminic acid (NeuAc), the label group (2-AB, 2-AP, DAP, 2-AA, PMP, or the alditol form), and dehydration ($-H_2O$) were considered. All glycan assignments were made within a specified tolerance level (≤5 ppm). Some of the data from the tandem MS analysis of the corresponding glycans were further processed using GlycoWorkbench to ensure correlation with database assignments. All the quantitative data were normalized to the total identified oligosaccharide peak area (in the format of percentage, %).

RESULTS AND DISCUSSION

The structural diversity and complexity of HMOs make the industry synthesis and production of various HMO standards very challenging. The limited available HMOs and their high cost made it difficult for us to start the assays by collecting a reasonable mixture of HMO standards in known relative proportions. To seek out an optimal strategy with broad applicability, a stock of HMOs pooled from 13 random healthy lactating mothers was instead used as our object since the person-to-person variability in glycosyltransferase expression affects the HMOs' composition (Kobata, 1992; McGuire et al., 2017; Thurl et al., 2017). The overall workflow for comparing different combination strategies for HMO profiling is summarized in **Scheme 1**. Three protocols were developed sequentially to optimize HMO analysis. Protocol 1: lactose-free R-HMOs were derivatized by different labeling techniques,

purified by SPE approaches, and analyzed by the HILIC-MS approach; Protocol 2: lactose-free R-HMOs were derivatized by the optimal labeling technique proposed, purified by SPE approaches, and analyzed by the PGC-MS approach; Protocol 3: raw C-HMOs with lactose were derivatized by the optimal labeling technique and analyzed by the PGC-MS approach directly. Chromatographic and MS data obtained from all the protocols are described below.

Comparing Labeling Techniques in Terms of MS Data Quality

Six derivatization reactions were selected for HMO analysis in this comparative study (**Supplementary Figure S1**). 2-AB, 2-AP, DAP, 2-AA, or PMP labeling reactions were performed under strong acidic, weak acidic, or alkaline conditions as described in the experimental section. The reductive reaction of native glycans to glycan alditols was conducted under neutral conditions, acting as a reference method. With the design incorporating different reaction conditions in Protocol 1, LC-MS-based profiling analysis of HMOs utilizing different derivatizations could be systematically evaluated. Different LC separation and MS setting conditions will influence the ionization efficiency, the types of adducts formed, and the fragmentation patterns for differently derivatized oligosaccharides. To allow a fair comparison of different derivatized HMOs, a general HILIC was selected as our initial chromatography approach because it was usually performed well with all reductively tagged derivatives (Ahn et al., 2010; Lauber et al., 2015; Cesla et al., 2016). Also, to prevent MS signal contamination by salts, labeling reagents and lactose, a prelabeling GCC SPE, followed by a post-labeling DPA-6S SPE and GCC SPE, was employed to purify the analytes. The total ion chromatography (TIC) of HMOs derivatized by 2-AA, 2-AB, 2-AP, DAP, and PMP is shown in **Figure 1**, with reduced HMO alditols as a reference. To compare the compatibility of each derivatization with the MS detection, we looked into their ionization behaviors in ESI-MS, their fragmentation patterns in CID-MS/MS, their compositional profiling results and their relative sensitivity based on the MS signal strength.

Ionization Behaviors of Different Derivatives

HMOs can be classified into four groups according to their residual modification: fucosylated neutral HMOs, nonfucosylated neutral HMOs, sialylated acidic HMOs, and the fucosylated and sialylated acidic HMOs. To avoid biased comparison, four glycan compositions were selected as representatives of the four HMO classes. The representative glycan compositions (denoted numerically in this text as Fuc-Hex-HexNAc-Neu5Ac; nomenclature is described in **Supplementary Table S1**) were as follows: 1-2-0-0 (FL series, for fucosylated neutral HMOs), 0-3-1-0 (LNT series, for nonfucosylated neutral HMOs), 0-3-1-1 (LST series, for sialylated acidic HMOs), and 1-4-2-1 (MFMSLNH series, for fucosylated and sialylated acidic HMOs).

In the positive ESI-MS mode, all five labeled derivatives formed consistent abundant protonated ions ($[M + H]^+$ or $[M+2H]^{2+}$) for four representative glycans (**Supplementary Figure S2**; **Supplementary Table S2**), while the reference

alditol derivatives produced heterogeneous sodium adduct ions ($[M + H]^+$, $[M+2H]^{2+}$, $[M + Na]^+$, and $[M + NH_4]^+$) and exhibited inconsistent ionization patterns between fucosylated HMOs (**Supplementary Figures S2A,D**) and nonfucosylated HMOs (**Supplementary Figures S2B,C**). In the negative mode, 2-AB, 2-AP, and DAP derivatives unexpectedly produced heterogeneous and inconsistent molecular ions for four glycans ($[M–H]^-$, $[M–2H]^{2-}$, $[M + HCOO]^-$, $[M + HCOO–H]^{2-}$, $[M+2HCOO + Na]^-$, $[M+2HCOO–H + Na]^{2-}$, $[M + HCOONH_4–H]^-$, $[M + HCOO–H + Na]^-$, and $[M + Na–2H]^-$) (**Supplementary Figure S2**) as well as the reference alditols. On the contrary, 2-AA and PMP derivatives still shared a consistent ionization pattern by prominent deprotonated ions ($[M–H]^-$ or $[M–2H]^{2-}$) for all representative glycans. The less the sodium adducts formed for each HMO, the more promising the labeling technique is since this can help solve the notorious problem of unfair detection of hetero-oligosaccharides caused by inconsistent molecular ion signal splitting.

Fragmentation Patterns of Different Derivatives

$2'$-FL of the FL series (1-2-0-0), LNT of the LNT series (0-3-1-0), LST-b of the LST series (0-3-1-1), and F-LST-a of the MFMSLNT series (1-3-1-1) were further selected as representatives for comparison analysis of their fragmentation efficiency. Their MS/MS spectra were characterized (**Supplementary Figure S3**; **Supplementary Table S3**) using the systematic nomenclature for carbohydrate fragmentation (Domon and Costello, 1988), assisted with the fragmentation rules reported previously (Lang et al., 2014; Liu et al., 2015; Lang et al., 2018).

All six positive ESI-CID MS/MS spectra of $2'$-FL derivatives were characterized by exclusive Y-type cleavage at every glycosidic bond (**Supplementary Figure S3A**). Four LNT derivatives (2-AP, 2-AB, DAP, and PMP) characterized by abundant Y ions; one LNT-2-AA derivative produced comparable abundance of Y ions and B_2 ions, and the reference LNT-alditol was fragmented into prominent B ions and weak Y ions instead (**Supplementary Figure S3B**). For sialylated acidic HMO, LST-b, three derivatives (2-AP, DAP, and PMP) produced only trace amounts of Y ions and the reference alditol did not produce any Y ions, while 2-AA and 2-AB derivatives were featured by abundant diagnostic glycosidic cleavage ions with a balanced intensity distribution, for example, Y ions, B ions, specific D ions, and characteristic Y/Y-type ions (**Supplementary Figure S3C**). For the longer acidic F-LST-a, unfortunately, the positive-ion mode did not provide any informative CID-MS/MS fragmentation spectra for all six derivatives, probably due to the big molecular size and the negatively charged sialic acid group.

Under negative-ion conditions (**Supplementary Figure S3**), the fragmentation varied substantially for four representative glycans for three derivatives (2-AB, 2-AP, and DAP) and the reference alditols, such as intense cross-ring ions ($^{1,3}A_2$ ions) for $2'$-FL, extensive glycosidic cleavages (Y_2, Z_3, and B_2 ions) for LNT, or minor fragment ions ($Y_{3\beta}$, $^{0,2}X_{3\beta}$ ions) for acidic LST-b and longer acidic F-LST-a. PMP derivatives gave the least satisfactory identification spectra for all four glycans because the loss of PMP from $[M–H]^-$ and Z_1 ions competitively inhibited the

FIGURE 2 | Relative quantifications of all HMO peaks for 2-AB, 2-AP, DAP, alditol, 2-AA, and PMP derivatives in the positive (+) and negative (−) modes depicted as a heatmap. In the heatmaps, the region in black indicates no detectable glycans, and glycans are marked with four colors based on the classification of structure features.

FIGURE 3 | Comparison of ESI-MS ionic abundances observed for 2-AB, 2-AP, DAP, alditol, 2-AA, and PMP derivatives in the positive (+) and negative (−) modes. Data acquired from common identified glycan peaks comprising four or more than four monosaccharides **(A)** and trisaccharides **(B)**. For each type of derivative, three reactions were set up, and three measurements were taken on each product. Percent errors were on the order of ±5%.

production of other diagnostic sugar cleavage ions. Surprisingly, 2-AA derivatives produced sequentially fragmented Y ions (e.g., Y_1, Y_2, Y_3, etc.) with high abundance for all four representatives. This consistent fragmentation information for a mixture of hetero-oligosaccharides could work as a reference for future research on optimizing the multiple reaction monitoring-mass spectrometry (MRM-MS) technique for a wide-range detection of HMOs in biological mixtures which usually have significant structure and size distribution (Fong et al., 2011; Mank et al., 2019).

Compositional Analysis of HMOs by Different Labeling Techniques

Following data acquisition on HILIC-MS, peaks were assigned based on the monosaccharide composition (Fuc-Hex-HexNAc-Neu5Ac, x-x-x-x) (**Supplementary Figure S4**). Retention of the

four HMO classes can be easily observed: fucosylated neutral (blue), nonfucosylated neutral (green), sialylated acidic (pink), and fucosylated and sialylated acidic (red). The overlaid extracted glycan chromatograms illustrate the extensive and complex glycan components. Total HMO intensities for each derivative were normalized to 100% to allow compositional profiling analysis. A clear view of the composition distribution was easily discerned when the numeric data of relative quantities of the glycan structures were further normalized by logarithmic (base 2) and transformed in a heatmap (**Figure 2**).

The total number of glycan compositions (peaks of isomers were summed as one glycan composition) identified for all derivatives was 68. As expected, for all six types of derivatives, the negative ESI-MS mode rendered 12–32 more HMOs (monosaccharide compositions) detectable than the positive-ion

mode did. More minor sialylated acidic HMOs (e.g., 0-4-2-**2**, 0-5-3-**2**, 0-6-4-**1**, 0-7-5-**1**, 0-8-6-**1**, 1-4-2-**2**, 2-4-2-**2**, 3-5-3-**1**, 4-5-3-**1**, 1-5-3-**2**, 2-5-3-**2**, 3-5-3-**2**, 1-6-4-**1**, 3-6-4-**1**, 2-6-4-**2**, and 2-7-5-**1**) and minor fucosylated neutral HMOs with large molecular weights (e.g., **4**-4-2-**0**, **3**-5-3-**0**, **4**-5-3-**0**, **5**-5-3-**0**, **3**-6-4-**0**, **4**-6-4-**0**, **5**-6-4-**0**, **1**-7-5-**0**, **2**-7-5-**0**, **3**-7-5-**0**, and **4**-7-5-**0**) were detected under negative HILIC-MS conditions. In addition, the reference alditol strategy (reduced but not labeled) enabled the most glycans (62 monosaccharide compositions) being detected; the labeling strategies of PMP, 2-AA, and DAP allowed the moderate numbers (51–53 monosaccharide compositions) of glycans detected, while the labeling strategies of 2-AB and 2-AP allowed less than 45 numbers of glycans (monosaccharide compositions) detected. Therefore, for qualification analysis, especially the in-depth structural characterization of a mixture of HMOs, reducing the oligosaccharides to their alditols and analyzing in the negative-ion MS mode would benefit the analysis best by covering the most structures.

Previously published work from the Lebrilla group using the chemical reduction and PGC-TOF-MS characterized the fine structures of 45 neutral glycans (isomers) (18 monosaccharide compositions) and 30 sialylated acidic glycans (isomers) (14 monosaccharide compositions), establishing very fruitful libraries for HMO structures (Wu et al., 2010; Wu et al., 2011). With the milk oligosaccharide standards available, the absolute quantification methods for measuring the predominant HMOs were further developed by them using MRM (Hong et al., 2014). As discussed above in *Ionization Behaviors of Different Derivatives* and *Fragmentation Patterns of Different Derivatives*, the reference alditols produced heterogeneous adduct molecular ions for four representative glycans in the negative-ion mode, which may lead to the unfair detection of hetero-oligosaccharides for ESI-MS1-based profiling analysis, and the fragmentation patterns varying substantially between glycans, which theoretically might increase the sensitivity difference between different glycans when applied for the MRM-based absolute quantification analysis. To figure out whether the former factor will affect the profiling analysis of HMOs, we compared the sensitivity of six types of derivatives based on MS1 signal intensity.

Relative Sensitivity Evaluation for Different Labeling Techniques

To achieve the sensitivity comparison analysis of six derivatives, the signal intensities of common identified glycans for each HMO class were summed for each derivative (**Figure 3**). Since several papers have reported the specific loss of 3-FL during the clean-up step of GCC-SPE (Blank et al., 2011; Xu et al., 2017; van Leeuwen, 2019), we split the comparison into two sessions: HMOs with more than four monosaccharides (**Figure 3A**) and HMOs with three monosaccharides (**Figure 3B**).

For comparison of the HMOs with more than four monosaccharides (**Figure 3A**), the order of detection sensitivity for 2-AB, 2-AP, DAP, and alditol derivatives varied substantially for four HMO classes in both ion modes, which agreed with our hypothesis that the formation of heterogeneous sodium adduct ions would split the signal of each peak and thus lead to a variation of sensitivity between different glycans.

Therefore, we expected enhanced sensitivity by 2-AA or PMP derivatizations, which promises consistent ionization behaviors for different HMO classes. The intensities of 2-AA and PMP derivatives did rank the second highest or the highest for two neutral HMO classes in both ion modes. For the sialylated HMO class in positive, the signal of PMP derivatives, however, ranked the lowest, while the signal of 2-AA derivatives was comparable to that of 2-AP derivatives, which was not satisfactory. For two acidic HMO classes in the negative-ion mode, despite the low intensities of PMP derivatives, the intensities of 2-AA derivatives ranked the second highest for sialylated HMOs and even the highest for fucosylated and sialylated HMOs. The overall enhanced sensitivity by 2-AA derivatization for all four HMO classes made 2-AA more promising. This observation of sensitivity difference between different structural featured HMOs indirectly proved that the varied multiple adduct sodium forms will affect the glycan profiling results.

Another comparison on the small trisaccharides (**Figure 3B**) however revealed that 2-AA derivatization had an additional specific loss of sialylated lactose (SL) series (0-2-0-1) compared with other labeling methods. The SL series are the most abundant components of acidic HMOs and have become a hot spot in HMO biology research (Weiss and Hennet, 2012; Kurakevich et al., 2013; Moon et al., 2016; Hobbs et al., 2021).

For preventing the loss of 3-FL, Xu et al. skipped the SPE step and directly injected the reduced HMO alditols containing an overload of lactose into LC-MS; however, 2′-FL was co-eluted with the abundant lactose, and thus, its ion was suppressed by lactose (Xu et al., 2017). Subsequently, Gu et al. developed an approach to quantitate HMOs including 3-FL by combining three analytical methods, HPAEC-PAD (for determining 3-FL), PGC-LC-MS (for other HMOs), and one-dimensional ^{1}H-NMR (for showing relative levels of different structural elements), following a GCC-SPE process (Gu et al., 2021a). This approach provided more accurate information on the relevant HMOs, although the operation process was labor-intensive, thus posing challenges to the development of high-throughput protocols (Gu et al., 2021a). Here, we also expect to optimize the 2-AA-associated analytical protocols further to avoid the selective loss of small HMOs since 2-AA labeling had shown quite beneficial observations for most of the medium or large HMOs.

Comparing Three Combination Strategies in Terms of Pretreatment and LC-MS Methods

To achieve a robust glycan profiling analysis, we seek to develop a compatible LC-MS system which can separate and determine the HMOs with reasonable resolution as well as suppress the interference from excess salt, labels, and lactose. As is known, reversed-phase chromatography (RPC) is considered a mature technique that can separate the ionic salts, less polar analytes, and hydrophobic molecules reasonably based on their polarity level difference. For our derivatized HMOs, the hydrophobic label group does prefer to bind to the RPC stationary phase, whereas the extremely polar glycan moiety tends to be washed out with extremely weak retention. Since the fluorescent label hydrophobic moiety is much smaller than the glycan
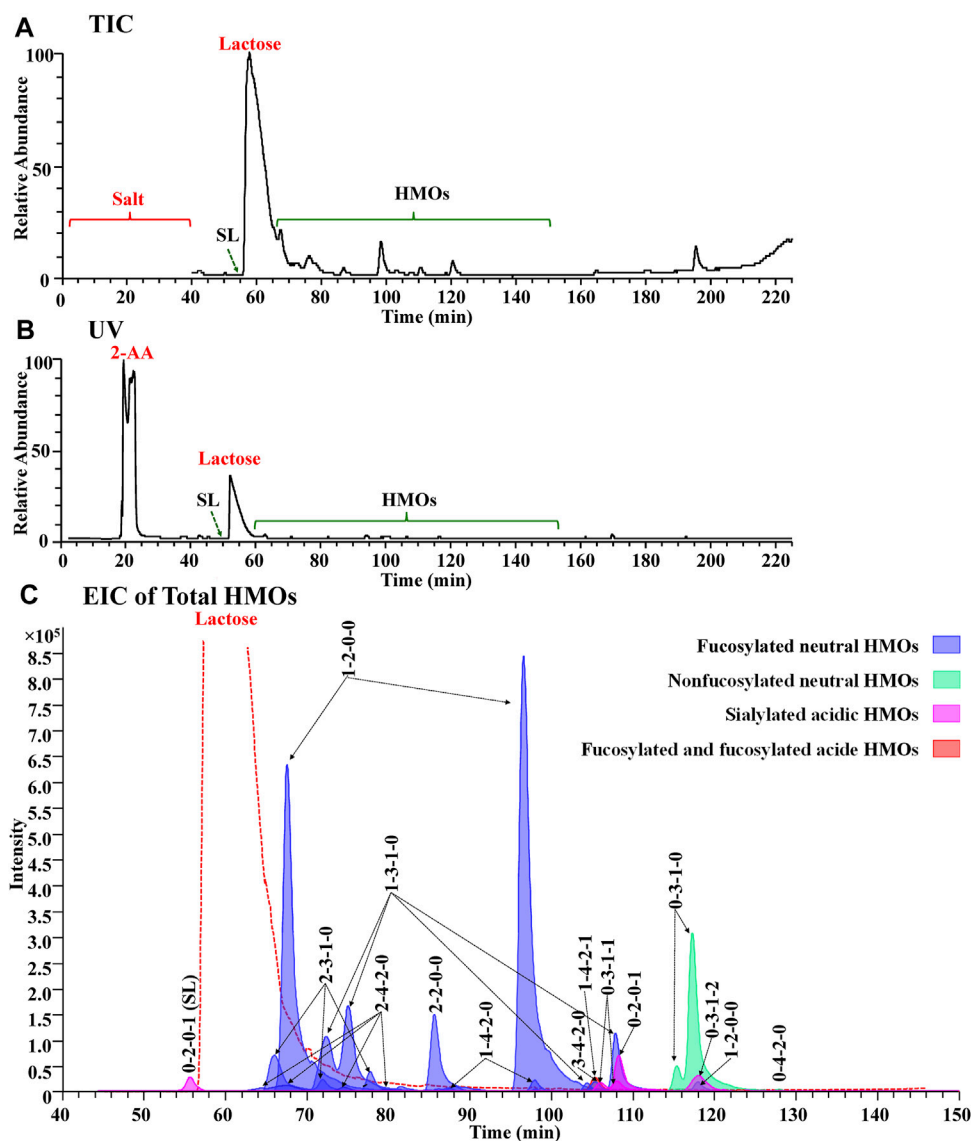
FIGURE 4 | Full MS compatibility of the PGC method in the Protocol 3 strategy was demonstrated by the agreement of TIC **(A)**, UV **(B)**, and EIC of total HMOs **(C)**. In **(C)**, each glycan class was designated with a different color: fucosylated neutral (blue), nonfucosylated neutral (green), sialylated acidic (pink), and fucosylated and sialylated acidic (red).

hydrophilic moiety, the overall retention of 2-AA-labeled HMOs is very weak on the conventional RPC. Alternatively, PGC, which has been called the ultimate reversed-phase material, shows stronger adsorption of polar analytes. PGC has been employed successively for glycomic analysis of reduced milk oligosaccharide alditols (Wu et al., 2010; Wu et al., 2011). Inspired by Xu's SPE skipping process and Wu's PGC-based glycomic analysis work, we optimized our 2-AA labeling-based HMO profiling analysis by designing another two analytical protocols → SPE-labeling-SPE-PGC/MS (referred to as Protocol 2) and labeling-PGC/MS (referred to as Protocol 3), in addition to SPE-labeling-SPE-HILIC/MS described in sections above (referred to as Protocol 1) (**Scheme 1**). Protocol 2 used the same sample loaded to Protocol 1, meaning that it was carried out with the same SPE route as Protocol 1; this aims to seek out whether an enhanced LC separation capacity or a similar level of

MS data quality could be acquired with PGC-MS, compared with HILIC-MS. Protocol 3 cut off all the cleanup SPE steps that were utilized in Protocol 1 and Protocol 2; this aims to seek out whether the selective loss of specific HMOs could be prevented with the simple rapid sample preparation process.

Evaluation of the Chromatographic Separation and Mass Spectrometric Data Quality

In the preliminary part of this work, a series of experimental PGC-MS conditions were optimized to achieve a satisfactory LC separation and noncontamination compatibility with MS. Using a buffer system of 10 mM ammonia bicarbonate for both Protocol 2 and Protocol 3, we obtained a surprisingly good chromatogram from Protocol 3, with online removal of salts and labels from total milk carbohydrates, very limited overlapping between labeled lactose and

FIGURE 5 | Overlaid EIC showing the elution profile of 2-AA-labeled HMOs via Protocol 1 **(A)**, Protocol 2 **(B)**, and Protocol 3 **(C)** in the negative mode. Each glycan class was designated with a different color: fucosylated neutral (blue), nonfucosylated neutral (green), sialylated acidic (pink), and fucosylated and sialylated acidic (red). Figures on the right side of the panels show low-abundance peaks (compositions of major peaks, including x-2-0-0, x-3-1-0, x-4-2-0, 0-3-0-0, and 0-3-1-0, were excluded, and x = any number of fucose residues).

labeled HMOs, and resolute separation of most labeled HMOs (**Figure 4**). The ionic salts and the hydrophobic 2-AA labels were co-washed out 10 minutes away from HMOs, allowing us to use the diverter valve built in an LTQ-Orbitrap XL mass spectrometer to automatically switch the contaminating eluents to the waste (**Figures 4A,B**). The retention of the overload 2-AA-labeled lactose between that of 2-AA-labeled SL (0-2-0-1) and that of the other 2-AA labeled HMOs (**Figure 4C**) prevented the signal suppression from lactose to

the HMOs, making both 2-AA labeling and the SPE-skipping process unique for the PGC-MS method.

For the evaluation of chromatographic separation of 2-AA-labeled HMOs, we compared the chromatograms obtained from Protocol 1, Protocol 2, and Protocol 3 (**Figure 5**) and observed that the retention behaviors of 2-AA-labeled HMOs differed on two chromatography modes. The HILIC method (Protocol 1) was somewhat referred to as the "size separation" method because the
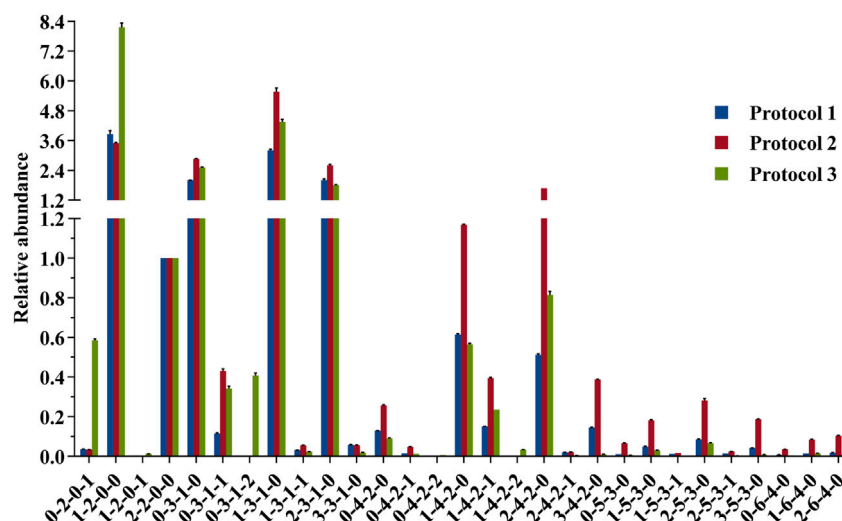
FIGURE 6 | Comparison of the relative abundances of 2-AA-labeled HMOs derived from 3 protocol strategies in the negative mode. Data were acquired from common major glycan peaks. LDFT (2-2-0-0) was selected as the internal standard element for a fair comparison. For each strategy, three measurements were taken on each product. Percent errors were on the order of ±5%.

retention correlated to the size of HMOs (**Figure 5A**). In contrast, the PGC method (Protocol 2 and Protocol 3) offered remarkable chromatographic separation of most glycans (**Figures 5B,C**) based on a combination mechanism of size, charge, and conformation properties. **Supplementary Table S4** lists the counts of visible extracted ion chromatogram (EIC) peaks for each monosaccharide composition obtained from the three protocols. Using the same sample preparation process, Protocol 1 provided 32 visible EIC glycan peaks (for 23 monosaccharide compositions), while Protocol 2 increased the visible EIC separation to 74 glycan peaks (for 23 monosaccharide compositions), indicating the powerful separation of isomers by PGC. Using the same PGC-MS method as Protocol 2, Protocol 3 had 74 EIC glycan peaks visibly determined based on 29 monosaccharide compositions. Protocol 3 had six more monosaccharide compositions extractable (e.g., 1-2-0-1, 0-3-1-2, 0-4-2-2, 1-4-2-2, 2-5-3-1, and 1-6-4-0), suggesting that the simple rapid sample preparation avoided selective loss of these glycans. The reason why Protocol 3 had more monosaccharide compositions extracted did not have more visible EIC peaks (counting the isomers) than Protocol 2 might rely on the fact that the amount of HMOs injected in Protocol 3 is about 10-fold less because of the high ratio of lactose and some minor isomeric peaks were too low to be visibly observed. Increasing the injection amount of the analytes in further future studies will improve the number of glycans determined. Between Protocols 2 and 3, there was little retention time shifted or little chromatographic resolution decreased, indicating that the injected interferences associated with Protocol 3 did not affect the retention capacity of 2-AA-labeled HMOs on the PGC column.

As described in *Compositional Analysis of HMOs by Different Labeling Techniques*, Wu et al. characterized 45 neutral glycans (isomers) (18 monosaccharide compositions) (Wu et al., 2010) and 30 sialylated acidic glycans (isomers) (14 monosaccharide compositions) (Wu et al., 2011) in 2 separate famous works. The

elegancy of the Protocol 3 approach is that it allowed a comparable number of glycan peaks (74) determined without fractioning the HMOs into neutral and acid parts. With further detailed characterization of ESI-CID-MS/MS spectra, an alternative library of 2-AA-labeled HMOs could be established in addition to the existing libraries of reduced HMO alditols. Considering this study focused on the compositional profiling analysis of HMOs based on monosaccharide constitution, the full identification of each HMO isomer for the Protocol 3 strategy is not described in this context. With each visible EIC peak representing an HMO structure, it has the potential to allow 74 HMOs semi-quantified or absolute-quantified when standards are available (Austin et al., 2016; Austin and Bénet, 2018; Austin et al., 2019; Samuel et al., 2019; Gu et al., 2021b). Further validation work by the MRM approach is warranted for this point (Hong et al., 2014; Totten et al., 2014; Xu et al., 2017).

Moreover, the ionization behaviors for four representative HMOs in PGC-MS (Protocol 2 and Protocol 3) were consistent (**Supplementary Figure S5**) to produce predominant deprotonated molecular ions in the negative-ion mode, providing the same MS data quality as HILIC-MS (Protocol 1). A comprehensive glycan profile can therefore be obtained with fair detection sensitivity of each structural element. Using the logarithm embedded in GlycResoft, we can analyze the relative abundance of each monosaccharide compositions in a high-throughput way, with much better sensitivity and accuracy than the manual integration of each EIC peak area.

Evaluation of the Sample Loss With or Without Pretreatment

The glycan profile for human milk using each protocol was characterized based on monosaccharide composition. To achieve a fair comparison between the three protocols, we need to select a reasonable internal glycan to normalize the relative abundances. This internal glycan for normalization analysis should not be involved in
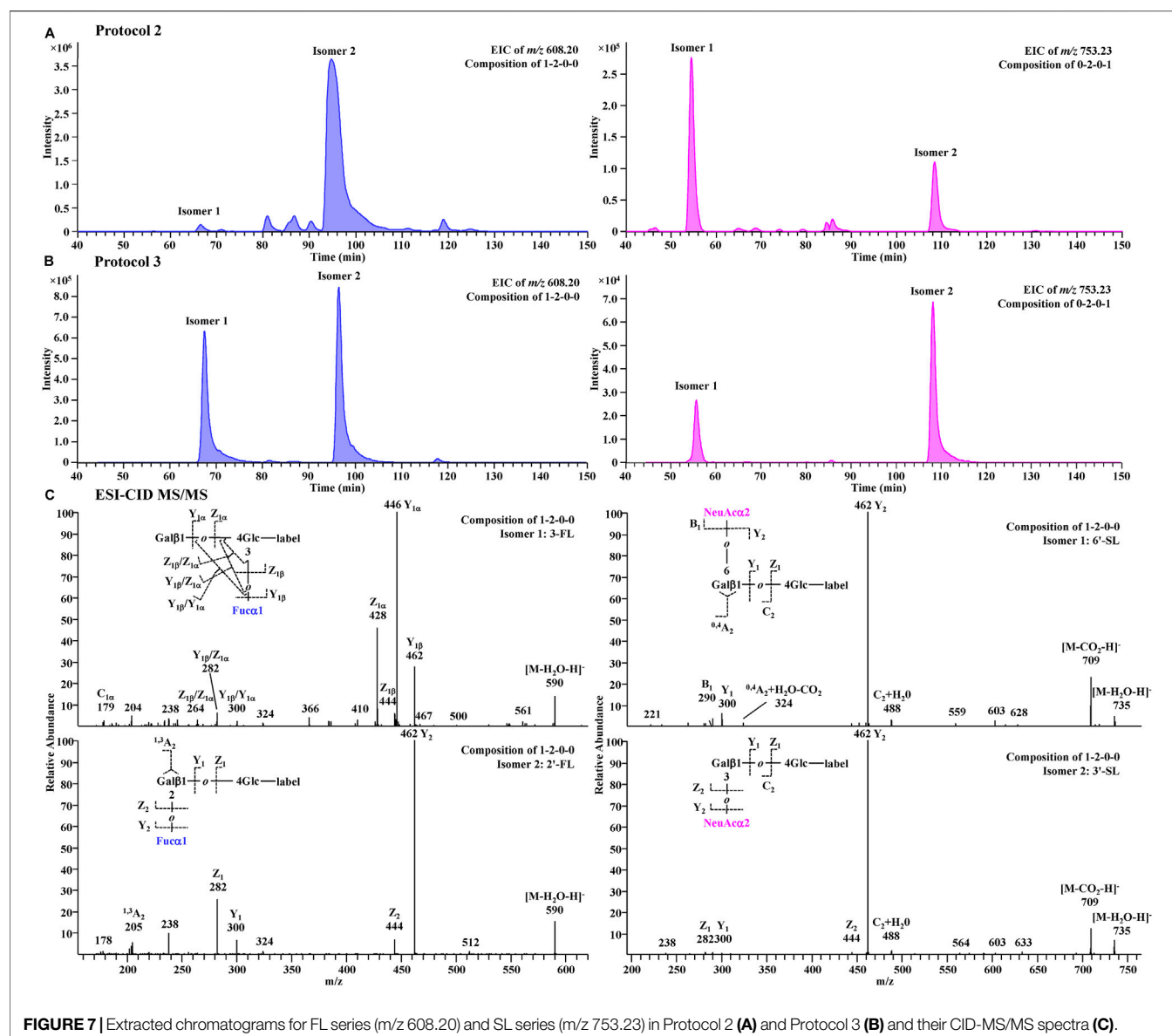
FIGURE 7 | Extracted chromatograms for FL series (m/z 608.20) and SL series (m/z 753.23) in Protocol 2 **(A)** and Protocol 3 **(B)** and their CID-MS/MS spectra **(C)**.

selective loss of small or large HMOs associated with SPE, liable to loss of sialic acid associated with the ESI source, or the signal ambiguated by isomeric separation. LDFT, a neutral difucosyllactose tetrasaccharide with only one structural configuration [Fucα1-2Galβ1-4(Fucα1-3) Glc-label], was therefore selected as the optimal internal standard for the comparative profiling analysis (**Figure 6**).

Comparing the glycan profiles acquired from HILIC-MS and PGC-MS using the same sample, Protocol 2 gave higher abundance ratios of larger glycans (e.g., x-3-1-x, x-4-2-x x-5-3-x, and x-6-4-x series) and lower abundance ratios of small glycans than Protocol 1 did. A possible explanation for this result may rely on the fact that the HILIC system needs a starting eluent of a high acetonitrile content (75%), requesting a similar sample solution before injection, thus resulting in possible precipitation of the large glycans (Ruhaak et al., 2010), while PGC needs an aqueous solvent as the starting eluent, making water as the sample solution, thus allowing all polar glycans

loaded to the system. Another explanation for the ratio distribution of small HMOs and large HMOs may be the low recovery of the internal standard LDFT on the PGC column specifically, but it is hard to believe that the recovery of other larger neutral HMOs was improved while this medium tetrasaccharide was not. The elution position of LDFT in HILIC was about 86 min and in PGC was about 85 min, with almost no interferences happening at this elution position. As published previously by Austin et al., using HILIC and 75% acetonitrile as the sample loading solution, they obtained recoveries of spiked standards between 85 and 120% for all structures tested, including some quite large glycans (Austin and Bénet, 2018), which is also true. Since their calibration study was carried out on one system, the solubility issue might contribute to the systematic error only. This ratio distribution difference of small and large HMOs between two LC-MS systems indicates that the compositional profiling analysis of HMOs can

be influenced by the two different LC-MS systems, but this does not mean that bias would happen for the standard-calibration-based quantification study via one LC-MS system. Nevertheless, further tests will be needed for the solubility/precipitation-associated loss of HMOs in 75% acetonitrile.

A further comparison of glycan profiles acquired from the same PGC-MS with or without SPE approaches indicated that Protocol 3 provided higher abundance ratios for most small glycans, especially the FL series and SL series, than Protocol 2. For example, glycan species 1-2-0-0 (FL series) showed an ~2-fold higher abundance ratio in Protocol 3 profiles than in Protocol 2, and glycan species 0-2-0-1 (SL series) showed an ~17-fold higher abundance ratio in Protocol 3 than in Protocol 2. To identify which glycan isomers were selectively lost for FL and SL series during the SPE process, we extracted their ion chromatograms (EIC) from Protocol 2 and Protocol 3 (**Figures 7A,B**). Detailed structures of the isomeric peaks regarding linkage positions were characterized (**Figure 7C**) according to the tandem mass spectra fragmentation principles reported previously (Wheeler and Harvey, 2000; Chai et al., 2001; Chai et al., 2002). 3-FL [Galβ1-4(Fucα1-3)Glc-label], one of the two FL isomers, was determined by the glycosidic cleavages, $C_{1\alpha}$ (m/z 179), $Y_{1\alpha}$ (m/z 446), and $Z_{1\alpha}$ (m/z 428). For the other isomer, 2′-FL (Fucα1-2Gal β1-4Glc-label), the cross-ring cleavage, $^{1,3}A_2$ ion at m/z 205, clearly indicated that the glycosidic bond of Fuc to Gal was (1→2). The isomer of SL series, 6′-SL (NeuAcα1-6Galβ1-4Glc-label), was also distinguished from 3′-SL (NeuAcα1-3Galβ1-4Glc-label) by the cross-ring cleavage, $^{0,4}A_2+H_2O-CO_2$ ion (m/z 324). Normalized by LDFT, we compared their relative EIC peak area abundances between Protocol 2 and Protocol 3 (**Supplementary Figure S6**). While 2′-FL showed comparable abundances between two protocols, the relative abundance of 3-FL was ~60-fold higher in Protocol 3 than in Protocol 2. For SL isomers, Protocol 3 provided ~6 and ~40-fold higher relative abundances for 6′-SL and 3′-SL (>5 fold) than Protocol 2. The issue of selective sample loss of trisaccharides, especially 3-FL and 3′-SL, associated with the SPE approaches was therefore confirmed for both Protocol 1 and Protocol 2. In addition to the selective loss of 3-FL reported previously, the loss of 3′-SL in these SPE steps should be paid attention to. Protocol 3 prevents such sample loss extensively. In addition, for the disialylated acidic glycan species (e.g., 0-3-1-2, 0-4-2-2, and 1-4-2-2), Protocol 3 rendered them much more detectable than the other two protocols, reducing the loss of two negative charged glycan series. These observations proved Protocol 3 an adequate and reliable profiling analysis.

CONCLUSION

This study evaluated the influence of glycan labeling, sample preparation, and LC-MS methods on the compositional profiling analysis of HMOs. With an identical MS data quality, a robust chromatographic separation, and a rapid simple sample preparation, an analytical protocol of 2-AA labeling, followed by a direct PGC-MS analysis, provided an enhanced profiling analysis of HMOs. Benefited by the simple sample preparation process, 96-well plates can be employed to enable the high-throughput analysis

in future studies. With the sensitivity improved nonselectively, this analytical protocol is further expected to increase the number of HMOs quantifiable in a natural biological mixture by integrating with MRM-MS (Yuan et al., 2005; Unterieser and Mischnick, 2011). Full in-depth sequence characterization and isobaric glycan distinguishability are further needed to update the glycomic library based on this analytical strategy. This research will provide more tools and methods to bring the science of human milk oligosaccharides forward.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Scientific Ethics Special Committee of Ocean University of China. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YL, GL, and GY conceived conception and study design, contributed to data interpretation and discussion, and revised and edited the manuscript. YL, YZ, CW, LH, and XL conducted the sample preparation and revised the manuscript. YL and NS conducted the measurements and data evaluation.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.691299/full#supplementary-material

REFERENCES

Ahn, J., Bones, J., Yu, Y. Q., Rudd, P. M., and Gilar, M. (2010). Separation of 2-aminobenzamide Labeled Glycans Using Hydrophilic Interaction Chromatography Columns Packed with 1.7μm Sorbent. *J. Chromatogr. B* 878 (3-4), 403–408. doi:10.1016/j.jchromb.2009.12.013

Anumula, K. R. (2006). Advances in Fluorescence Derivatization Methods for High-Performance Liquid Chromatographic Analysis of Glycoprotein Carbohydrates. *Anal. Biochem.* 350 (1), 1–23. doi:10.1016/j.ab.2005.09.037

Anumula, K. R., and Dhume, S. T. (1998). High Resolution and High Sensitivity Methods for Oligosaccharide Mapping and Characterization by normal Phase High Performance Liquid Chromatography Following Derivatization with Highly Fluorescent Anthranilic Acid. *Glycobiology* 8 (7), 685–694. doi:10.1093/glycob/8.7.685

Anumula, K. R. (1994). Quantitative Determination of Monosaccharides in Glycoproteins by High-Performance Liquid Chromatography with Highly Sensitive Fluorescence Detection. *Anal. Biochem.* 220 (2), 275–283. doi:10.1006/abio.1994.1338

Anumula, K. R. (2014). Single Tag for Total Carbohydrate Analysis. *Anal. Biochem.* 457, 31–37. doi:10.1016/j.ab.2014.04.019

Asakuma, S., Urashima, T., Akahori, M., Obayashi, H., Nakamura, T., Kimura, K., et al. (2008). Variation of Major Neutral Oligosaccharides Levels in Human Colostrum. *Eur. J. Clin. Nutr.* 62 (4), 488–494. doi:10.1038/sj.ejcn.1602738

Auer, F., Jarvas, G., and Guttman, A. (2021). Recent Advances in the Analysis of Human Milk Oligosaccharides by Liquid Phase Separation Methods. *J. Chromatogr. B* 1162, 122497. doi:10.1016/j.jchromb.2020.122497

Austin, S., and Bénet, T. (2018). Quantitative Determination of Non-lactose Milk Oligosaccharides. *Analytica Chim. Acta* 1010, 86–96. doi:10.1016/j.aca.2017.12.036

Austin, S., De Castro, C. A., Sprenger, N., Binia, A., Affolter, M., Garcia-Rodenas, C. L., et al. (2019). Human Milk Oligosaccharides in the Milk of Mothers Delivering Term versus Preterm Infants. *Nutrients* 11 (6), 1282. doi:10.3390/nu11061282

Austin, S., De Castro, C., Bénet, T., Hou, Y., Sun, H., Thakkar, S., et al. (2016). Temporal Change of the Content of 10 Oligosaccharides in the Milk of Chinese Urban Mothers. *Nutrients* 8 (6), 346. doi:10.3390/nu8060346

Bao, Y., and Newburg, D. S. (2008). Capillary Electrophoresis of Acidic Oligosaccharides from Human Milk. *Electrophoresis* 29 (12), 2508–2515. doi:10.1002/elps.200700873

Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M., and Parekh, R. B. (1995). Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-amino Benzamide and Anthranilic Acid. *Anal. Biochem.* 230 (2), 229–238. doi:10.1006/abio.1995.1468

Blank, D., Gebhardt, S., Maass, K., Lochnit, G., Dotz, V., Blank, J., et al. (2011). High-throughput Mass finger Printing and Lewis Blood Group Assignment of Human Milk Oligosaccharides. *Anal. Bioanal. Chem.* 401 (8), 2495–2510. doi:10.1007/s00216-011-5349-9

Bode, L. (2012). Human Milk Oligosaccharides: Every Baby Needs a Sugar Mama. *Glycobiology* 22 (9), 1147–1162. doi:10.1093/glycob/cws074

Bode, L. (2006). Recent Advances on Structure, Metabolism, and Function of Human Milk Oligosaccharides. *J. Nutr.* 136 (8), 2127–2130. doi:10.1093/jn/136.8.2127

Bode, L. (2015). The Functional Biology of Human Milk Oligosaccharides. *Early Hum. Dev.* 91 (11), 619–622. doi:10.1016/j.earlhumdev.2015.09.001

Boehm, G., Stahl, B., Jelinek, J., Knol, J., Miniello, V., and Moro, G. E. (2005). Prebiotic Carbohydrates in Human Milk and Formulas. *Acta Paediatr. Suppl.* 94 (449), 18–21. doi:10.1111/j.1651-2227.2005.tb02149.x

Česla, P., Vaňková, N., Křenková, J., and Fischer, J. (2016). Comparison of Isocratic Retention Models for Hydrophilic Interaction Liquid Chromatographic Separation of Native and Fluorescently Labeled Oligosaccharides. *J. Chromatogr. A* 1438, 179–188. doi:10.1016/j.chroma.2016.02.032

Chai, W., Lawson, A. M., and Piskarev, V. (2002). Branching Pattern and Sequence Analysis of Underivatized Oligosaccharides by Combined MS/MS of Singly and Doubly Charged Molecular Ions in Negative-Ion Electrospray Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* 13 (6), 670–679. doi:10.1016/S1044-0305(02)00363-X

Chai, W., Piskarev, V. E., Zhang, Y., Lawson, A. M., and Kogelberg, H. (2005). Structural Determination of Novel Lacto-N-Decaose and its Monofucosylated Analogue from Human Milk by Electrospray Tandem Mass Spectrometry and 1H NMR Spectroscopy. *Arch. Biochem. Biophys.* 434 (1), 116–127. doi:10.1016/j.abb.2004.09.035

Chai, W., Piskarev, V., and Lawson, A. M. (2001). Negative-ion Electrospray Mass Spectrometry of Neutral Underivatized Oligosaccharides. *Anal. Chem.* 73 (3), 651–657. doi:10.1021/ac0010126

Coppa, G., Pierani, P., Zampini, L., Carloni, I., Carlucci, A., and Gabrielli, O. (1999). Oligosaccharides in Human Milk during Different Phases of Lactation. *Acta Paediatr. Suppl.* 88 (430), 89–94. doi:10.1111/j.1651-2227.1999.tb01307.x

De Leoz, M. L. A., Duewer, D. L., Fung, A., Liu, L., Yau, H. K., Potter, O., et al. (2020). NIST Interlaboratory Study on Glycosylation Analysis of Monoclonal Antibodies: Comparison of Results from Diverse Analytical Methods. *Mol. Cel Proteomics* 19 (1), 11–30. doi:10.1074/mcp.RA119.001677

Docq, S., Spoelder, M., Wang, W., and Homberg, J. R. (2020). The Protective and Long-Lasting Effects of Human Milk Oligosaccharides on Cognition in Mammals. *Nutrients* 12 (11), 3572. doi:10.3390/nu12113572

Domann, P. J., Pardos-Pardos, A. C., Fernandes, D. L., Spencer, D. I. R., Radcliffe, C. M., Royle, L., et al. (2007). Separation-based Glycoprofiling Approaches Using Fluorescent Labels. *Proteomics* 7 (1), 70–76. doi:10.1002/pmic.200700640

Domon, B., and Costello, C. E. (1988). A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates. *Glycoconjugate J.* 5(4), 397–409. doi:10.1007/Bf01049915

Fong, B., Ma, K., and McJarrow, P. (2011). Quantification of Bovine Milk Oligosaccharides Using Liquid Chromatography-Selected Reaction Monitoring-Mass Spectrometry. *J. Agric. Food Chem.* 59 (18), 9788–9795. doi:10.1021/jf202035m

Galeotti, F., Coppa, G. V., Zampini, L., Maccari, F., Galeazzi, T., Padella, L., et al. (2012). On-line High-Performance Liquid Chromatography-Fluorescence Detection-Electrospray Ionization-Mass Spectrometry Profiling of Human Milk Oligosaccharides Derivatized with 2-aminoacridone. *Anal. Biochem.* 430 (1), 97–104. doi:10.1016/j.ab.2012.07.027

Gao, X., Yang, J., Huang, F., Wu, X., Li, L., and Sun, C. (2003). Progresses of Derivatization Techniques for Analyses of Carbohydrates. *Anal. Lett.* 36 (7), 1281–1310. doi:10.1081/AL-120021087

Gonia, S., Tuepker, M., Heisel, T., Autran, C., Bode, L., and Gale, C. A. (2015). Human Milk Oligosaccharides Inhibit Candida Albicans Invasion of Human Premature Intestinal Epithelial Cells. *J. Nutr.* 145 (9), 1992–1998. doi:10.3945/jn.115.214940

Grabarics, M., Csernák, O., Balogh, R., and Béni, S. (2017). Analytical Characterization of Human Milk Oligosaccharides - Potential Applications in Pharmaceutical Analysis. *J. Pharm. Biomed. Anal.* 146, 168–178. doi:10.1016/j.jpba.2017.08.039

Gu, F., Kate, G. A. t., Arts, I. C. W., Penders, J., Thijs, C., Lindner, C., et al. (2021a). Combining HPAEC-PAD, PGC-LC-MS, and 1D 1H NMR to Investigate Metabolic Fates of Human Milk Oligosaccharides in 1-Month-Old Infants: a Pilot Study. *J. Agric. Food Chem.* 69 (23), 6495–6509. doi:10.1021/acs.jafc.0c07446

Gu, F., Wang, S., Beijers, R., de Weerth, C., and Schols, H. A. (2021b). Structure-Specific and Individual-dependent Metabolization of Human Milk Oligosaccharides in Infants: A Longitudinal Birth Cohort Study. *J. Agric. Food Chem.* 69 (22), 6186–6199. doi:10.1021/acs.jafc.0c07484

Ha, S.-H., Kwak, C.-H., Park, J.-Y., Abekura, F., Lee, Y.-C., Kim, J.-s., et al. (2020). 3′-sialyllactose Targets Cell Surface Protein, SIGLEC-3, and Induces Megakaryocyte Differentiation and Apoptosis by Lipid Raft-dependent Endocytosis. *Glycoconj J.* 37 (2), 187–200. doi:10.1007/s10719-019-09902-1

Harvey, D. J. (2000). Electrospray Mass Spectrometry and Fragmentation of N-Linked Carbohydrates Derivatized at the Reducing Terminus. *J. Am. Soc. Mass. Spectrom.* 11(10), 900–915. doi:10.1016/S1044-0305(00)00156-2

Hester, S. N., Chen, X., Li, M., Monaco, M. H., Comstock, S. S., Kuhlenschmidt, T. B., et al. (2013). Human Milk Oligosaccharides Inhibit Rotavirus Infectivity *In Vitro* and in Acutely Infected Piglets. *Br. J. Nutr.* 110 (7), 1233–1242. doi:10.1017/S0007114513000391

Hobbs, M., Jahan, M., Ghorashi, S. A., and Wang, B. (2021). Current Perspective of Sialylated Milk Oligosaccharides in Mammalian Milk: Implications for Brain and Gut Health of Newborns. *Foods* 10 (2), 473. doi:10.3390/foods10020473

Hong, Q., Ruhaak, L. R., Totten, S. M., Smilowitz, J. T., German, J. B., and Lebrilla, C. B. (2014). Label-free Absolute Quantitation of Oligosaccharides Using Multiple Reaction Monitoring. *Anal. Chem.* 86 (5), 2640–2647. doi:10.1021/ac404006z

Jaitly, N., Mayampurath, A., Littlefield, K., Adkins, J. N., Anderson, G. A., and Smith, R. D. (2009). Decon2LS: An Open-Source Software Package for Automated Processing and Visualization of High Resolution Mass Spectrometry Data. *BMC Bioinformatics* 10 (1), 87. doi:10.1186/1471-2105-10-87

Kinoshita, M., Nakatani, Y., Yamada, K., Yamamoto, S., and Suzuki, S. (2021). A Rapid and Facile Preparation of APTS-Labeled N-Glycans by Combination of Ion Pair-Assisted Extraction and HILIC-SPE for Routine Glycan Analysis. *J. Pharm. Biomed. Anal.* 195, 113875. doi:10.1016/j.jpba.2020.113875

Kishimoto, Y., Okada, F., Maesako, T., Yamamoto, S., Kinoshita, M., Hayakawa, T., et al. (2020). Analysis of 2-aminopyridine Labeled Glycans by Dual-Mode Online Solid Phase Extraction for Hydrophilic Interaction and Reversed-phase Liquid Chromatography. *J. Chromatogr. A* 1625, 461194. doi:10.1016/j.chroma.2020.461194

Kobata, A. (1992). Structures and Functions of the Sugar Chains of Glycoproteins. *Eur. J. Biochem.* 209 (2), 483–501. doi:10.1111/j.1432-1033.1992.tb17313.x

Kozak, R. P., Tortosa, C. B., Fernandes, D. L., and Spencer, D. I. R. (2015). Comparison of Procainamide and 2-aminobenzamide Labeling for Profiling and Identification of Glycans by Liquid Chromatography with Fluorescence Detection Coupled to Electrospray Ionization-Mass Spectrometry. *Anal. Biochem.* 486, 38–40. doi:10.1016/j.ab.2015.06.006

Kunz, C., Rudloff, S., Baier, W., Klein, N., and Strobel, S. (2000). OLIGOSACCHARIDES INHUMANMILK: Structural, Functional, and Metabolic Aspects. *Annu. Rev. Nutr.* 20 (1), 699–722. doi:10.1146/annurev.nutr.20.1.699

Kurakevich, E., Hennet, T., Hausmann, M., Rogler, G., and Borsig, L. (2013). Milk Oligosaccharide Sialyl(2,3)lactose Activates Intestinal CD11c+ Cells through TLR4. *Proc. Natl. Acad. Sci.* 110 (43), 17444–17449. doi:10.1073/pnas.1306322110

Lang, Y., Zhao, X., Liu, L., and Yu, G. (2014). Applications of Mass Spectrometry to Structural Analysis of marine Oligosaccharides. *Mar. Drugs* 12 (7), 4005–4030. doi:10.3390/md12074005

Lang, Y. Z., Liu, S. L., Wang, C., Zhang, X., Lu, Y. J., Cai, C., et al. (2018). Separation and Structural Sequence Analysis of Sialylated HMOs via Tandem Mass Spectrometry. *Chem. J. Chin. Universities-Chinese* 39 (4), 645–652. doi:10.7503/cjcu20170593

Lattová, E., Snovida, S., Perreault, H., and Krokhin, O. (2005). Influence of the Labeling Group on Ionization and Fragmentation of Carbohydrates in Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* 16 (5), 683–696. doi:10.1016/j.jasms.2005.01.021

Lauber, M. A., Yu, Y.-Q., Brousmiche, D. W., Hua, Z., Koza, S. M., Magnelli, P., et al. (2015). Rapid Preparation of Released N-Glycans for HILIC Analysis Using a Labeling Reagent that Facilitates Sensitive Fluorescence and ESI-MS Detection. *Anal. Chem.* 87 (10), 5401–5409. doi:10.1021/acs.analchem.5b00758

Leo, F., Asakuma, S., Fukuda, K., Senda, A., and Urashima, T. (2010). Determination of Sialyl and Neutral Oligosaccharide Levels in Transition and Mature Milks of Samoan Women, Using Anthranilic Derivatization Followed by Reverse Phase High Performance Liquid Chromatography. *Biosci. Biotechnol. Biochem.* 74 (2), 298–303. doi:10.1271/bbb.90614

Liu, S. L., Lang, Y. Z., Zhu, H., Yan, L. N., Lu, Y. J., Zhao, X. L., et al. (2015). Isolation and Structural Identification of Neutral Human Milk Oligosaccharides. *Chem. J. Chin. Universities-Chinese* 36 (6), 1087–1093. doi:10.7503/cjcu20150118

Mank, M., Welsch, P., Heck, A. J. R., and Stahl, B. (2019). Label-free Targeted LC-ESI-MS2 Analysis of Human Milk Oligosaccharides (HMOS) and Related Human Milk Groups with Enhanced Structural Selectivity. *Anal. Bioanal. Chem.* 411 (1), 231–250. doi:10.1007/s00216-018-1434-7

Mantovani, V., Galeotti, F., Maccari, F., and Volpi, N. (2016). Recent Advances on Separation and Characterization of Human Milk Oligosaccharides. *Electrophoresis* 37 (11), 1514–1524. doi:10.1002/elps.201500477

Marcobal, A., and Sonnenburg, J. L. (2012). Human Milk Oligosaccharide Consumption by Intestinal Microbiota. *Clin. Microbiol. Infect.* 18 (4), 12–15. doi:10.1111/j.1469-0691.2012.03863.x

Mariño, K., Lane, J. A., Abrahams, J. L., Struwe, W. B., Harvey, D. J., Marotta, M., et al. (2011). Method for Milk Oligosaccharide Profiling by 2-aminobenzamide Labeling and Hydrophilic Interaction Chromatography. *Glycobiology* 21 (10), 1317–1330. doi:10.1093/glycob/cwr067

Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., et al. (2012). GlycReSoft: a Software Package for Automated Recognition of Glycans from LC/MS Data. *PLoS One* 7 (9), e45474. doi:10.1371/journal.pone.0045474

McGuire, M. K., Meehan, C. L., McGuire, M. A., Williams, J. E., Foster, J., Sellen, D. W., et al. (2017). What's normal? Oligosaccharide Concentrations and Profiles in Milk Produced by Healthy Women Vary Geographically. *Am. J. Clin. Nutr.* 105 (5), 1086–1100. doi:10.3945/ajcn.116.139980

Melmer, M., Stangler, T., Premstaller, A., and Lindner, W. (2011). Comparison of Hydrophilic-Interaction, Reversed-phase and Porous Graphitic Carbon Chromatography for Glycan Analysis. *J. Chromatogr. A* 1218 (1), 118–123. doi:10.1016/j.chroma.2010.10.122

Moon, J. S., Joo, W., Ling, L., Choi, H. S., and Han, N. S. (2016). *In Vitro* digestion and Fermentation of Sialyllactoses by Infant Gut Microflora. *J. Funct. Foods* 21, 497–506. doi:10.1016/j.jff.2015.12.002

Morelle, W., Page, A., and Michalski, J.-C. (2005). Electrospray Ionization Ion Trap Mass Spectrometry for Structural Characterization of Oligosaccharides Derivatized with 2-aminobenzamide. *Rapid Commun. Mass. Spectrom.* 19 (9), 1145–1158. doi:10.1002/rcm.1900

Newburg, D. S. (2013). Glycobiology of Human Milk. *Biochem. Mosc.* 78 (7), 771–785. doi:10.1134/S0006297913070092

Ninonuevo, M. R., Park, Y., Yin, H., Zhang, J., Ward, R. E., Clowers, B. H., et al. (2006). A Strategy for Annotating the Human Milk Glycome. *J. Agric. Food Chem.* 54 (20), 7471–7480. doi:10.1021/jf0615810

Oliveros, E., Vázquez, E., Barranco, A., Ramírez, M., Gruart, A., Delgado-García, J., et al. (2018). Sialic Acid and Sialylated Oligosaccharide Supplementation during Lactation Improves Learning and Memory in Rats. *Nutrients* 10 (10), 1519. doi:10.3390/nu10101519

Pabst, M., Kolarich, D., Pöltl, G., Dalik, T., Lubec, G., Hofinger, A., et al. (2009). Comparison of Fluorescent Labels for Oligosaccharides and Introduction of a New Postlabeling Purification Method. *Anal. Biochem.* 384 (2), 263–273. doi:10.1016/j.ab.2008.09.041

Redmond, J. W., and Packer, N. H. (1999). The Use of Solid-phase Extraction with Graphitised Carbon for the Fractionation and Purification of Sugars. *Carbohydr. Res.* 319(1-4), 74–79. doi:10.1016/S0008-6215(99)00130-5

Royle, L., Campbell, M. P., Radcliffe, C. M., White, D. M., Harvey, D. J., Abrahams, J. L., et al. (2008). HPLC-based Analysis of Serum N-Glycans on a 96-well Plate Platform with Dedicated Database Software. *Anal. Biochem.* 376 (1), 1–12. doi:10.1016/j.ab.2007.12.012

Ruhaak, L. R., Zauner, G., Huhn, C., Bruggink, C., Deelder, A. M., and Wuhrer, M. (2010). Glycan Labeling Strategies and Their Use in Identification and Quantification. *Anal. Bioanal. Chem.* 397 (8), 3457–3481. doi:10.1007/s00216-010-3532-z

Saba, J. A., Shen, X., Jamieson, J. C., and Perreault, H. (1999). Effect of 1-Phenyl-3-Methyl-5-Pyrazolone Labeling on the Fragmentation Behavior of Asialo and Sialylated N-Linked Glycans under Electrospray Ionization Conditions. *Rapid Commun. Mass. Spectrom.* 13 (8), 704–711. doi:10.1002/(sici)1097-0231(19990430)13:8<704::aid-rcm543>3.0.co;2-v

Saba, J. A., Shen, X., Jamieson, J. C., and Perreault, H. l. n. (2001). Investigation of Different Combinations of Derivatization, Separation Methods and Electrospray Ionization Mass Spectrometry for Standard Oligosaccharides and Glycans from Ovalbumin. *J. Mass. Spectrom.* 36 (5), 563–574. doi:10.1002/jms.158

Samuel, T. M., Binia, A., de Castro, C. A., Thakkar, S. K., Billeaud, C., Agosti, M., et al. (2019). Impact of Maternal Characteristics on Human Milk Oligosaccharide Composition over the First 4 Months of Lactation in a Cohort of Healthy European Mothers. *Sci. Rep.* 9 (1), 11767. doi:10.1038/s41598-019-48337-4

Smilowitz, J. T., Lebrilla, C. B., Mills, D. A., German, J. B., and Freeman, S. L. (2014). Breast Milk Oligosaccharides: Structure-Function Relationships in the Neonate. *Annu. Rev. Nutr.* 34 (1), 143–169. doi:10.1146/annurev-nutr-071813-105721

Song, J.-F., Weng, M.-Q., Wu, S.-M., and Xia, Q.-C. (2002). Analysis of Neutral Saccharides in Human Milk Derivatized with 2-aminoacridone by Capillary Electrophoresis with Laser-Induced Fluorescence Detection. *Anal. Biochem.* 304 (1), 126–129. doi:10.1006/abio.2001.5589

Song, X., Xia, B., Stowell, S. R., Lasanajak, Y., Smith, D. F., and Cummings, R. D. (2009). Novel Fluorescent Glycan Microarray Strategy Reveals Ligands for Galectins. *Chem. Biol.* 16 (1), 36–47. doi:10.1016/j.chembiol.2008.11.004

Strydom, D. J. (1994). Chromatographic Separation of 1-Phenyl-3-Methyl-5-Pyrazolone-Derivatized Neutral, Acidic and Basic Aldoses. *J. Chromatogr. A* 678(1), 17–23. doi:10.1016/0021-9673(94)87069-1

Suzuki, S., Kakehi, K., and Honda, S. (1996). Comparison of the Sensitivities of Various Derivatives of Oligosaccharides in LC/MS with Fast Atom Bombardment and Electrospray Ionization Interfaces. *Anal. Chem.* 68 (13), 2073–2083. doi:10.1021/ac951144c

Takegawa, Y., Deguchi, K., Ito, S., Yoshioka, S., Nakagawa, H., and Nishimura, S.-I. (2005). Simultaneous Analysis of 2-Aminopyridine-Derivatized Neutral and Sialylated Oligosaccharides from Human Serum in the Negative-Ion Mode by Sonic spray Ionization Ion Trap Mass Spectrometry. *Anal. Chem.* 77 (7), 2097–2106. doi:10.1021/ac048499t

Tao, N., DePeters, E. J., Freeman, S., German, J. B., Grimm, R., and Lebrilla, C. B. (2008). Bovine Milk Glycome. *J. Dairy Sci.* 91 (10), 3768–3778. doi:10.3168/jds.2008-1305

Tao, N., DePeters, E. J., German, J. B., Grimm, R., and Lebrilla, C. B. (2009). Variations in Bovine Milk Oligosaccharides during Early and Middle Lactation Stages Analyzed by High-Performance Liquid Chromatography-Chip/mass Spectrometry. *J. Dairy Sci.* 92 (7), 2991–3001. doi:10.3168/jds.2008-1642

Thurl, S., Henker, J., Siegel, M., Tovar, K., and Sawatzki, G. (1997). Detection of Four Human Milk Groups with Respect to Lewis Blood Group Dependent Oligosaccharides. *Glycoconj J.* 14 (7), 795–799. doi:10.1023/a:1018529703106

Thurl, S., Munzert, M., Boehm, G., Matthews, C., and Stahl, B. (2017). Systematic Review of the Concentrations of Oligosaccharides in Human Milk. *Nutr. Rev.* 75 (11), 920–933. doi:10.1093/nutrit/nux044

Tonon, K. M., Miranda, A., Abrão, A. C. F. V., de Morais, M. B., and Morais, T. B. (2019). Validation and Application of a Method for the Simultaneous Absolute Quantification of 16 Neutral and Acidic Human Milk Oligosaccharides by Graphitized Carbon Liquid Chromatography - Electrospray Ionization - Mass Spectrometry. *Food Chem.* 274, 691–697. doi:10.1016/j.foodchem.2018.09.036

Totten, S. M., Wu, L. D., Parker, E. A., Davis, J. C. C., Hua, S., Stroble, C., et al. (2014). Rapid-throughput Glycomics Applied to Human Milk Oligosaccharide Profiling for Large Human Studies. *Anal. Bioanal. Chem.* 406 (30), 7925–7935. doi:10.1007/s00216-014-8261-2

Triantis, V., Bode, L., and van Neerven, R. J. J. (2018). Immunological Effects of Human Milk Oligosaccharides. *Front. Pediatr.* 6, 190. doi:10.3389/fped.2018.00190

Underwood, M. A., Gaerlan, S., De Leoz, M. L. A., Dimapasoc, L., Kalanetra, K. M., Lemay, D. G., et al. (2015). Human Milk Oligosaccharides in Premature Infants: Absorption, Excretion, and Influence on the Intestinal Microbiota. *Pediatr. Res.* 78 (6), 670–677. doi:10.1038/pr.2015.162

Unterieser, I., and Mischnick, P. (2011). Labeling of Oligosaccharides for Quantitative Mass Spectrometry. *Carbohydr. Res.* 346 (1), 68–75. doi:10.1016/j.carres.2010.11.001

van Leeuwen, S. S. (2019). Challenges and Pitfalls in Human Milk Oligosaccharide Analysis. *Nutrients* 11 (11), 2684. doi:10.3390/nu11112684

Walsh, C., Lane, J. A., van Sinderen, D., and Hickey, R. M. (2020). Human Milk Oligosaccharides: Shaping the Infant Gut Microbiota and Supporting Health. *J. Funct. Foods* 72, 104074. doi:10.1016/j.jff.2020.104074

Wang, C., Yuan, J., Wang, Z., and Huang, L. (2013). Separation of One-Pot Procedure Released O-Glycans as 1-Phenyl-3-Methyl-5-Pyrazolone Derivatives by Hydrophilic Interaction and Reversed-phase Liquid Chromatography Followed by Identification Using Electrospray Mass Spectrometry and Tandem Mass Spectrometry. *J. Chromatogr. A* 1274, 107–117. doi:10.1016/j.chroma.2012.12.005

Wang, W., Chen, F., Wang, Y., Wang, L., Fu, H., Zheng, F., et al. (2018). Optimization of Reactions between Reducing Sugars and 1-Phenyl-3-Methyl-5-Pyrazolone (PMP) by Response Surface Methodology. *Food Chem.* 254, 158–164. doi:10.1016/j.foodchem.2018.02.001

Ward, R. E. (2009). Isolation of Milk Oligosaccharides Using Solid-phase Extraction. *Open Glycoscience* 5 (2), 9–15. doi:10.2174/1875398100902010009

Weiss, G. A., and Hennet, T. (2012). The Role of Milk Sialyllactose in Intestinal Bacterial Colonization. *Adv. Nutr.* 3 (3), 483S–488S. doi:10.3945/an.111.001651

Wheeler, S. F., and Harvey, D. J. (2000). Negative Ion Mass Spectrometry of Sialylated Carbohydrates: Discrimination of N-Acetylneuraminic Acid Linkages by MALDI-TOF and ESI-TOF Mass Spectrometry. *Anal. Chem.* 72 (20), 5027–5039. doi:10.1021/ac000436x

Wiciński, M., Sawicka, E., Gębalski, J., Kubiak, K., and Malinowski, B. (2020). Human Milk Oligosaccharides: Health Benefits, Potential Applications in Infant Formulas, and Pharmacology. *Nutrients* 12 (1), 266. doi:10.3390/nu12010266

Wu, K.-J., Chen, Y.-H., Bae, E.-K., Song, Y., Min, W., and Yu, S.-J. (2020). Human Milk Oligosaccharide 2′-Fucosyllactose Reduces Neurodegeneration in Stroke Brain. *Transl. Stroke Res.* 11 (5), 1001–1011. doi:10.1007/s12975-019-00774-z

Wu, S., Grimm, R., German, J. B., and Lebrilla, C. B. (2011). Annotation and Structural Analysis of Sialylated Human Milk Oligosaccharides. *J. Proteome Res.* 10 (2), 856–868. doi:10.1021/pr101006u

Wu, S., Tao, N., German, J. B., Grimm, R., and Lebrilla, C. B. (2010). Development of an Annotated Library of Neutral Human Milk Oligosaccharides. *J. Proteome Res.* 9 (8), 4138–4151. doi:10.1021/pr100362f

Xia, B., Kawar, Z. S., Ju, T., Alvarez, R. A., Sachdev, G. P., and Cummings, R. D. (2005). Versatile Fluorescent Derivatization of Glycans for Glycomic Analysis. *Nat. Methods* 2 (11), 845–850. doi:10.1038/nmeth808

Xu, G., Davis, J. C., Goonatilleke, E., Smilowitz, J. T., German, J. B., and Lebrilla, C. B. (2017). Absolute Quantitation of Human Milk Oligosaccharides Reveals Phenotypic Variations during Lactation. *J. Nutr.* 147 (1), 117–124. doi:10.3945/jn.116.238279

You, J., Sheng, X., Ding, C., Sun, Z., Suo, Y., Wang, H., et al. (2008). Detection of Carbohydrates Using New Labeling Reagent 1-(2-Naphthyl)-3-Methyl-5-Pyrazolone by Capillary Zone Electrophoresis with Absorbance (UV). *Analytica Chim. Acta* 609 (1), 66–75. doi:10.1016/j.aca.2007.12.022

Yuan, J., Hashii, N., Kawasaki, N., Itoh, S., Kawanishi, T., and Hayakawa, T. (2005). Isotope Tag Method for Quantitative Analysis of Carbohydrates by Liquid Chromatography-Mass Spectrometry. *J. Chromatogr. A* 1067 (1-2), 145–152. doi:10.1016/j.chroma.2004.11.070

Yuen, C.-T., Gee, C. K., and Jones, C. (2002). High-performance Liquid Chromatographic Profiling of Fluorescent Labelled N-Glycans on Glycoproteins. *Biomed. Chromatogr.* 16 (4), 247–254. doi:10.1002/bmc.154

Zauner, G., Koeleman, C. A. M., Deelder, A. M., and Wuhrer, M. (2012). Mass Spectrometric O-Glycan Analysis after Combined O-Glycan Release by Beta-Elimination and 1-Phenyl-3-Methyl-5-Pyrazolone Labeling. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1820 (9), 1420–1428. doi:10.1016/j.bbagen.2011.07.004

Zhang, Q., Li, H., Feng, X., Liu, B.-F., and Liu, X. (2014). Purification of Derivatized Oligosaccharides by Solid Phase Extraction for Glycomic Analysis. *PLoS One* 9 (4), e94232. doi:10.1371/journal.pone.0094232

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

A Linkage-specific Sialic Acid Labeling Strategy Reveals Different Site-specific Glycosylation Patterns in SARS-CoV-2 Spike Protein Produced in CHO and HEK Cell Substrates

Qiong Wang[1], Yan Wang[2], Shuang Yang[1,3], Changyi Lin[4], Lateef Aliyu[5], Yiqun Chen[5], Lisa Parsons[1], Yuan Tian[1], Hongpeng Jia[6], Andrew Pekosz[7], Michael J. Betenbaugh[5] and John F. Cipollo[1]*

[1]Laboratory of Bacterial Polysaccharides, Division of Bacterial, Parasitic and Allergenic Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Baltimore, MD, United States, [2]Mass Spectrometry Facility, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, United States, [3]Center for Clinical Mass Spectrometry, School of Pharmaceutical Sciences, Soochow University, Jiangsu, China, [4]Facility for Biotechnology Resources, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, United States, [5]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, United States, [6]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [7]Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus utilizes the extensively glycosylated spike (S) protein protruding from the viral envelope to bind to angiotensin-converting enzyme-related carboxypeptidase (ACE2) as its primary receptor to mediate host-cell entry. Currently, the main recombinant S protein production hosts are Chinese hamster ovary (CHO) and human embryonic kidney (HEK) cells. In this study, a recombinant S protein truncated at the transmembrane domain and engineered to express a C-terminal trimerization motif was transiently produced in CHO and HEK cell suspensions. To further evaluate the sialic acid linkages presenting on S protein, a two-step amidation process, employing dimethylamine and ammonium hydroxide reactions in a solid support system, was developed to differentially modify the sialic acid linkages on the glycans and glycopeptides from the S protein. The process also adds a charge to Asp and Glu which aids in ionization. We used MALDI-TOF and LC-MS/MS with electron-transfer/higher-energy collision dissociation (EThcD) fragmentation to determine global and site-specific N-linked glycosylation patterns. We identified 21 and 19 out of the 22 predicted N-glycosites of the SARS-CoV-2 S proteins produced in CHO and HEK, respectively. It was found that the N-glycosite at 1,158 position (N1158) and at 122, 282 and 1,158 positions (N122, N282 and N1158) were absent on S from CHO and HEK cells, respectively. The structural mapping of glycans of recombinant human S proteins reveals that CHO-Spike exhibits more complex and higher sialylation (α2,3-linked) content while HEK-Spike exhibits more high-mannose and a small amount of α2,3- and α2,6-linked sialic acids. The N74 site represents the most abundant glycosite on both spike proteins. The relatively higher amount of high-mannose abundant sites (N17, N234, N343, N616, N709, N717, N801, and N1134) on HEK-Spike suggests

that glycan-shielding may differ among the two constructs. HEK-Spike can also provide different host immune system interaction profiles based on known immune system active lectins. Collectively, these data underscore the importance of characterizing the site-specific glycosylation of recombinant human spike proteins from HEK and CHO cells in order to better understand the impact of the production host on this complex and important protein used in research, diagnostics and vaccines.
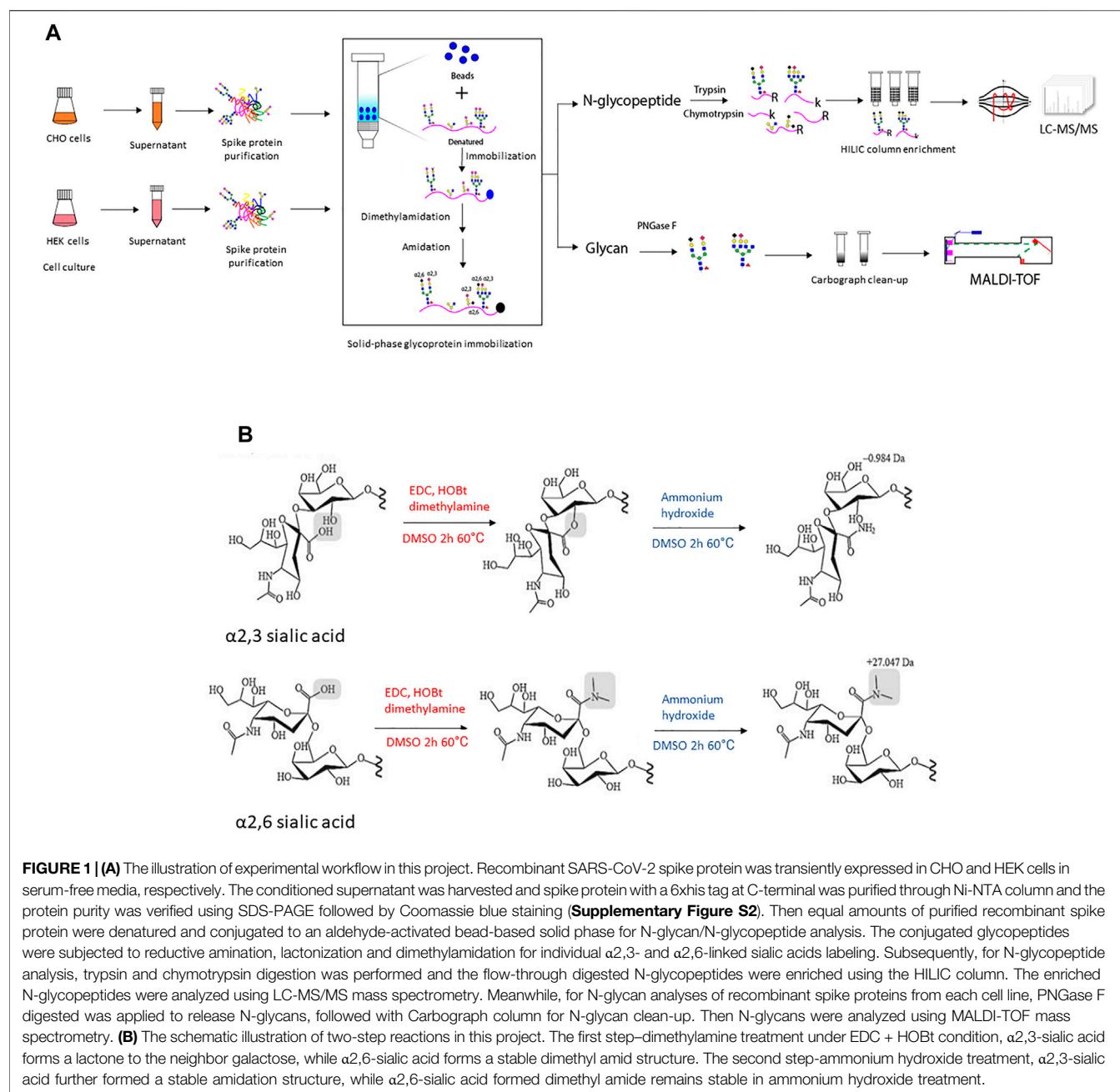
INTRODUCTION

The ongoing outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remains a major global pandemic affecting the lives of billions of people worldwide (Yuen et al., 2020). The pathogen of SARS-CoV-2 coronavirus causes fever and cough, shortness of breath, and hospitalization for pneumonia (Guo et al., 2020; Wiersinga et al., 2020). In view of its mortality, human-to-human and cross-species transmission ability, and virus mutation rate, there is a continuing need for effective diagnosis, efficient therapeutic treatment, and preventive means, including vaccines against this devastating disease (Wiersinga et al., 2020). As an enveloped virus, SARS-CoV-2 utilizes an extensively glycosylated spike (S) protein protruding from the viral envelope to bind to angiotensin-converting enzyme-related carboxypeptidase (ACE2) as its primary receptor (Bangaru et al., 2020; Wrapp et al., 2020). Current diagnostics, vaccine development, and neutralizing antibodies are all focused on S glycoprotein, which is the main humoral immune response target (Bangaru et al., 2020).

As a trimeric class I fusion protein, SARS-CoV-2 S has 1,273 amino acids (aa), consisting of a 13-aa signal peptide at the N-terminus, S1 subunit (14–685 aa) and membrane associated S2 subunit (686–1,273 aa) (Grant et al., 2020; Ghorbani et al., 2021). The S1 domain is responsible for binding to cell receptors, including the N-terminal domain (14–305 aa) and the receptor-binding domain (RBD, 319–541 aa); while the S2 domain mediates fusion between the virus and the cell membranes, which is compromised of the fusion peptide (FP) (788–806 aa), heptapeptide repeat sequence 1 (HR1) (912–984 aa), HR2 (1,163–1,213 aa), TM domain (1,213–1,237 aa), and cytoplasmic domain (1,237–1,273 aa) (Huang et al., 2020). The SARS-CoV-2 S binds to its receptor human ACE2 through the RBD region in its S1 subunit to initiate viral infection. Subsequently, human proteases proteolytically cleave the S protein at the S1/S2 and the S2' sites, which allows the virus to enter for membrane fusion (Belouzard et al., 2009; Glowacka et al., 2011; Hoffmann et al., 2020). The glycan profile of envelope glycoproteins can affect the pathobiological activities and characteristics of many viruses, such as glycoprotein folding, trafficking and stability, virus release, functional activity, and immune evasion including shielding of the peptide backbone by glycans (Watanabe et al., 2019). Similar to the SARS outbreak in 2003, the SARS-CoV-2 spike utilizes glycan shielding to thwart the host immune response. Here the surface of the virus envelope is coated with host-derived glycans (Casalino et al., 2020), which hides the protein surface to avoid detection by the body fluids and cellular components of the immune system (Sommerstein et al., 2015; Pritchard et al., 2015; Watanabe et al., 2020a), thereby promoting pathogen immune evasion. Moreover, when the viral glycoprotein evolve to mask immunogenic epitopes with particularly dense host-derived glycan arrays, such as high abundance of high-mannose N-glycans, neutralizing antibodies may not always recognize the underlying viral protein surface (Grant et al., 2020). Recently, a research group discovered that SARS-COV-2 spike glycoprotein glycans shield approximately 40% of the protein surface, although the glycans only account for 17% of the total molecular weight of the S trimer when expressed from human embryonic kidney cells (Grant et al., 2020; Yang et al., 2020). Molecular dynamic simulation of the interaction of spike and ACE2 proteins revealed the role of spike glycans in sterically masking polypeptide epitopes and directly participates in spike-ACE2 interactions (Zhao et al., 2020). Thus, understanding the glycosylation of recombinant S protein is very important for studying virus biology and immune response, and also help to provide information for the application of recombinant spike glycoprotein in diagnosis and vaccines.

As one of the most complex post-translational modifications found on secreted proteins and membrane-bound proteins, glycosylation can affect numerous physiological and pathological cell functions. Compared with other post-translational processes events, glycosylation is unique in its structural heterogeneity. The glycan pattern can vary greatly among individual glycosites of a given protein, different organisms, and different production hosts. Therefore, the characterization of glycans has presented a special analytical challenge due to the extraordinary heterogeneity and complexity caused by the number of glycan processing enzymes residing in the ER and Golgi apparatus. However, the advent of powerful mass spectrometry tools and related analysis methods has revolutionized our ability to identify and characterize protein glycosylation (Nwosu et al., 2011; Han and Costello, 2013; Sun et al., 2016). After the outbreak of the pandemic, multiple research groups reported glycomic and glycoproteomic analysis of the SARS-CoV-2 spike protein (De Leoz et al., 2020; Shajahan et al., 2020; Watanabe et al., 2020b; Sanda et al., 2021). However, these studies have not yet compared the production of spike proteins in multiple mammalian production hosts or characterized their sialic acid linkages.

FIGURE 1 | (A) The illustration of experimental workflow in this project. Recombinant SARS-CoV-2 spike protein was transiently expressed in CHO and HEK cells in serum-free media, respectively. The conditioned supernatant was harvested and spike protein with a 6xhis tag at C-terminal was purified through Ni-NTA column and the protein purity was verified using SDS-PAGE followed by Coomassie blue staining (**Supplementary Figure S2**). Then equal amounts of purified recombinant spike protein were denatured and conjugated to an aldehyde-activated bead-based solid phase for N-glycan/N-glycopeptide analysis. The conjugated glycopeptides were subjected to reductive amination, lactonization and dimethylamidation for individual α2,3- and α2,6-linked sialic acids labeling. Subsequently, for N-glycopeptide analysis, trypsin and chymotrypsin digestion was performed and the flow-through digested N-glycopeptides were enriched using the HILIC column. The enriched N-glycopeptides were analyzed using LC-MS/MS mass spectrometry. Meanwhile, for N-glycan analyses of recombinant spike proteins from each cell line, PNGase F digested was applied to release N-glycans, followed with Carbograph column for N-glycan clean-up. Then N-glycans were analyzed using MALDI-TOF mass spectrometry. **(B)** The schematic illustration of two-step reactions in this project. The first step–dimethylamine treatment under EDC + HOBt condition, α2,3-sialic acid forms a lactone to the neighbor galactose, while α2,6-sialic acid forms a stable dimethyl amid structure. The second step-ammonium hydroxide treatment, α2,3-sialic acid further formed a stable amidation structure, while α2,6-sialic acid formed dimethyl amide remains stable in ammonium hydroxide treatment.

Given that the two most important potential commercial recombinant protein production hosts are Chinese hamster ovary (CHO) and human embryonic kidney (HEK) cells, comparing the glycosylation profiles of recombinant spikes produced by these two organisms is useful for understanding what host-specific differences in glycosylation are and the potential relevance of these differences. Indeed, the differences in glycosylation between hamster ovary and human kidney cells may result in distinct glycan profiles of the same recombinant protein. For example, a comparison of 12 glycoproteins expressed from HEK and CHO suspension cell cultures revealed distinctive glycan structures, and CHO cells tend to express higher levels of sialic acid (Croset et al., 2012). In addition, CHO cells present exclusively α2,3-sialylation and trace amounts of Neu5Gc (Wang et al., 2015; Yehuda and Padler-Karavani, 2020). In contrast, HEK cells can simultaneously exhibit α2,3- or α2,6-linked sialylation. Sialic acids (Neu5Ac) are typically located at the terminus of oligosaccharides on glycoproteins, and these residues are mainly linked to the galactose residue by α2,3- or α2,6-linkages in humans (Varki et al., 2015; Zhou et al., 2020). In fact, differences in specific linkages can affect the function of viral protein such as the entry protein hemagglutinin of influenza A virus with an α2,6-linked sialic acid binding preference (Leung et al., 2012; Yang et al., 2017).

Therefore, in this study, recombinant full-size spike protein was transiently produced in suspension propagated CHO and HEK hosts. The resulting glycan and glycopeptide profiles were characterized in order to explore the different capabilities and characteristics of recombinant spike glycosylation. To further evaluate the presence of sialic acid linkages on the S glycans, a solid-phase method was used for two-step derivatization with dimethylamine and ammonium hydroxide to differentially modify the sialic acid linkages on glycans and glycopeptides from spike proteins secreted from the two hosts. **Figure 1** shows the schematic workflow of the glycan and glycoproteomic analytical process. We provided global and site-specific analysis of N-linked glycosylation on soluble full-size SARS-CoV-2 spike using MALDI-TOF and LC-MS/MS with electron-transfer/higher-energy collision dissociation (EThcD) fragmentation. It revealed extensive heterogeneity ranging from high-mannose type to complex type glycosylation profiles, and has detailed sialic acid linkage information. The structural mapping of glycans of recombinant human spike proteins shows that CHO-Spike is more complex and contains higher sialylation (α2,3-linked), while HEK-Spike has more high-mannose and a small amount of α2,3- and α2,6-linked sialylation. Some high-mannose abundant sites on HEK-Spike may indicate a comparatively different glycan shielding presentation and viral evasion ability compared with CHO-Spike, which interacts with high mannose specific lectins such as DC-SIGN and lung surfactant SP-D (Van Breedam et al., 2014). Overall, these data underscore the importance of characterizing glycosylation of recombinant human spike protein from HEK and CHO cells in order to better understand the impact of the production host on S protein used in research, diagnostics and vaccines.

MATERIALS AND METHODS

Cell Culture, Transfection and Protein Purification

The mammalian expression vector pCAGGS plasmid containing the stable soluble spike protein sequence was acquired from Dr. Andrew Pekosz of Bloomberg School of Public Health at Johns Hopkins University through the generosity of Dr. Florian Krammer of the Icahn School of Medicine at Mt. Sinai (Amanat et al., 2020; Stadlbauer et al., 2020). Substitutions at the furin cleavage sites and lysine (K) at 986 position and valine (V) at 987 position were replaced with prolines (P) in order to stabilize the quaternary structure for mammalian glycosylation processing (Watanabe et al., 2020b). This recombinant S protein was truncated at the transmembrane domain and engineered to express a C-terminal trimerization motif. A sequence comparison with the SARS-CoV-2 Wuhan strain is shown in **Supplementary Figure S1**. The CHOZN GS−/− cell line was acquired from MilliporeSigma. Suspension HEK 293 cells were acquired from the American Type Culture Collection (ATCC). For CHO cells, cell culture was maintained in EX-CELL CD CHO Fusion medium (MilliporeSigma) supplemented with 6 mM glutamine in 250 ml shake flasks with a working volume of 100 ml at 37°C and 8% CO_2. Transfection and transient expression of spike protein were performed using reagents and media from CHOgro® Expression System (Mirus Bio) following the corresponding protocol. The protein was harvested from 300 ml spent media 4 days after transfection. The protein was purified *via* the Ni-NTA purification approach (Spriestersbach et al., 2015). For HEK cells, cell culture was maintained in FreeStyle™ F17 Expression Medium (Thermo Fisher) supplemented with 6 mM glutamine in 250 ml shake flasks with a working volume of 100 ml at 37°C and 8% CO_2. Transfection was performed *via* a polyethyleneimine (PEI)-based approach (Longo et al., 2013). Specifically, 1 mg/ml of PEI in water was prepared and added to Opti-Pro SFM (Thermo Fisher) containing plasmid DNA in a 3.5:1 PEI to DNA ratio (w/w) to form the DNA-PEI complex. Spent media was collected 4 days after transfection and the protein was purified *via* the Ni-NTA purification approach. The purified spike protein with a 6xHis tag at the C terminal was evaluated using SDS-PAGE with Coomassie blue staining to check any impurities, as shown in **Supplementary Figure S2**. The protein concentration was measured by bicinchoninic acid assay (BCA assay).
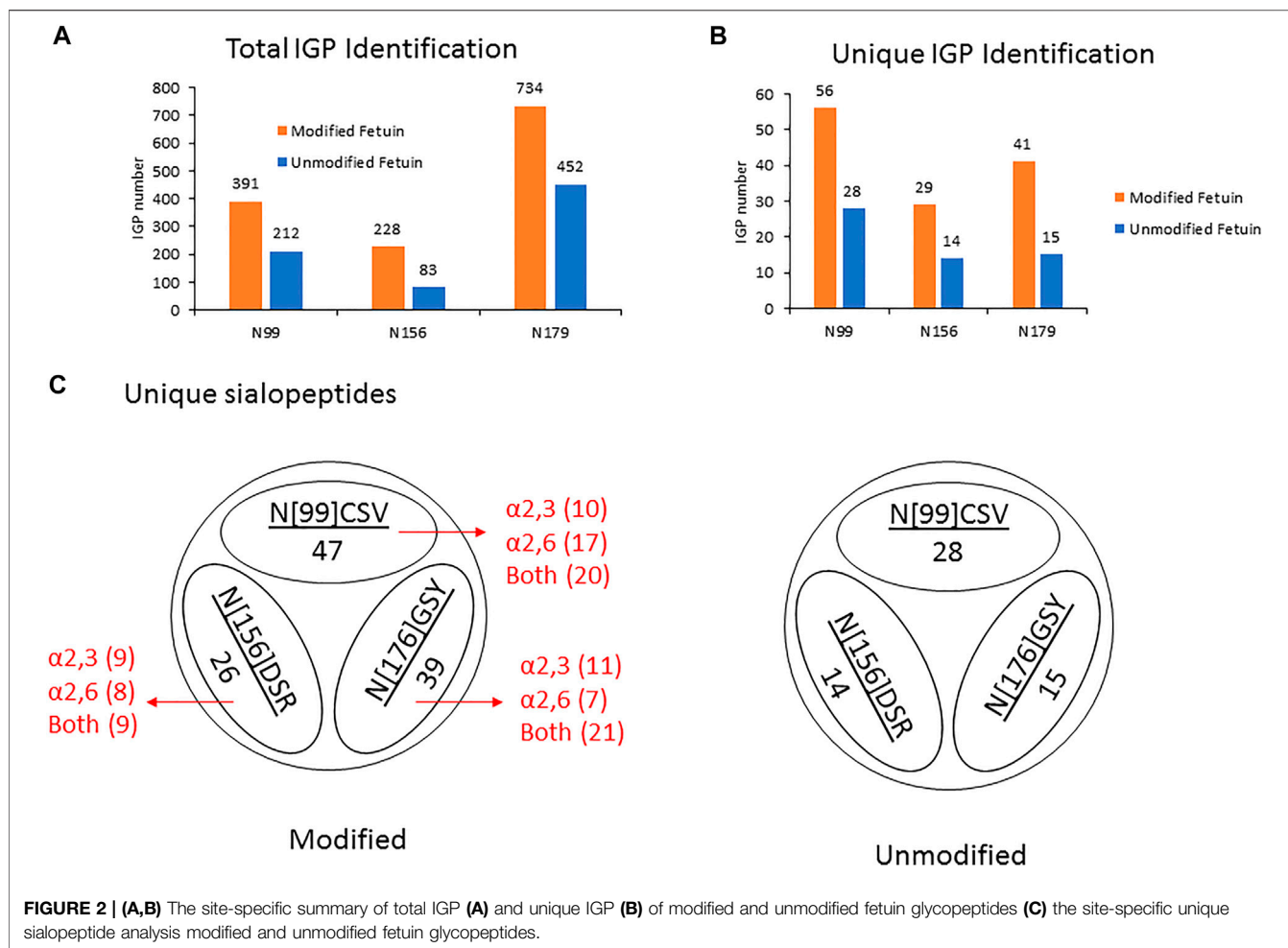
Solid-phase Protein Conjugates and Modifications

Purified S protein, 400 µg, was diluted in 500 µL solution containing 1x binding buffer and 1x denaturing buffer (New England BioLabs, B1704S) and denatured at 95°C for 5 min. 1x binding buffer was prepared by dissolving 294 mg sodium citrate and 53 mg sodium carbonate in 10 ml HPLC water. All chemicals were purchased from Sigma-Aldrich unless otherwise noted. Meanwhile, Aminolink resin (Fisher Scientific) was washed by 500 µL 1x binding buffer twice. Samples were then denatured, added to the pre-conditioned resin in a snap-cap spin column (SCSC, Fisher Scientific), and incubated overnight at room temperature. The 50 mM $NaCNBH_3$ reduction solution was added to the sample for another 4 h incubation. The resin was then washed with 1x PBS twice and further incubated with 50 mM $NaCNBH_3$ in PBS for another 4 h to block any active sites.

For sialic acid modification, samples were first treated with 200 µl 0.25 mol/L dimethylamine, 0.25 mol/L 1-ethyl-3-(3-dimethylamino) propyl)carbodiimide (EDC) and 0.25 mol/L 1-hydroxybenzotriazole (HOBt) in DMSO solution at 60°C for 2 h. EDC and HBOt were used as a carboxylic acid activator and a catalyst individually. Next, an equal volume of ammonium hydroxide (pH10) was added into the samples with another 2 h incubation at 60°C. Samples were then washed sequentially with 10% formic acid, 10% acetonitrile, 1 M NaCl and H_2O, each for three times. Finally, the sialic acid-modified protein conjugates were subjected to glycan and glycopeptide analysis individually.

N-Glycan Analysis

After solid-phase protein conjugates and modifications, 3ul of PNGase F (500,000 U/ml, New England BioLabs, Ipswich, MA) in NEB Glycobuffer 2 was added to the bead mixture and incubated overnight at 37°C. The extracted glycans were subjected to Carbocolumn N-glycan clean-up and stored at 4°C for MALDI-ToF analysis. The purified glycan was analyzed using a Bruker AutoFlex Speed MALDI-ToF/ToF spectrometer in the

FIGURE 2 | (A,B) The site-specific summary of total IGP **(A)** and unique IGP **(B)** of modified and unmodified fetuin glycopeptides **(C)** the site-specific unique sialopeptide analysis modified and unmodified fetuin glycopeptides.

reflective-positive ion mode. The MALDI-ToF MS parameters were set as following: mass range 800–6,000 Da, laser 70% and 8,000 summed shots per sample. Search for predicted glycan structures-based compositions was performed using GlycoWorkBench software. The α2,3-linked sialic acids formed a stable amidation structure with a mass shift of −0.984 Da, while the α2,6-linked sialic acids formed a stable dimethyl amidation structure in ammonium hydroxide with a mass shift of +27.047 Da.

On-Bead N-Glycopeptide Digestion and HILIC Enrichment

After protein conjugation and modification, the resin linked protein was incubated in 12 mM dithiothreitol (DTT) in 1M NH_4HCO_3 and 8 M urea at 37°C for 1 h. Iodoacetamide was then added to a final concentration of 16 mM (1 h at room temperature in the dark) to alkylate the protein. After alkylation, samples were washed by 1M NaCl, HPLC water and 25 mM NH_4HCO_3 sequentially (twice). For fetuin sample, sequence-grade trypsin digestion (protein: enzyme = 50:1, w/w) in 50 mM NH_4HCO_3 was performed at 37°C overnight. For spike samples, sequence-grade trypsin and chymotrypsin (as a cocktail)

digestion (protein: enzyme = 50:1, w/w) in 50 mM NH_4HCO_3 was performed at 37 °C overnight. The sample digest was then eluted in 80% ACN and subjected to HILIC N-glycopeptide enrichment. HILIC SPE chromatography was prepared as follows: add empty SPE (solid-phase extraction) frits to Grace Alltech extract-clean empty reservoir (1.5 ml; Fisher Scientific), load 500 ul TSKgel Amide-80 slurry in 50% ethanol (Sigma-Aldrich), and cap resin using empty SPE frits.

HILIC column enrichment was performed as follows: pre-condition HILIC column using 0.1% TFA and 60% ACN/0.1% TFA (1 ml, three times), load samples in 80% ACN/0.1% TFA (reload flow-through once), wash column by 80% ACN/0.1% TFA (1ml, twice), elute samples by 60% ACN/0.1% TFA, 40% ACN/0.1% TFA, and 0.1% TFA, and pool samples. The eluates were dried by rotary evaporation in a Speed-Vac instrument (Thermo Fisher) and re-suspended in 0.2% FA.

Nano LC-MS/MS Analysis

The dried peptide sample was resuspended in 0.2% formic acid (FA) and measured peptide concentration using A280 Nanodrop. Then, 1 μg of resuspended glycopeptide in 4 μL FA buffer were analyzed by LC-MS/MS using a Thermo Orbitrap Fusion Lumos
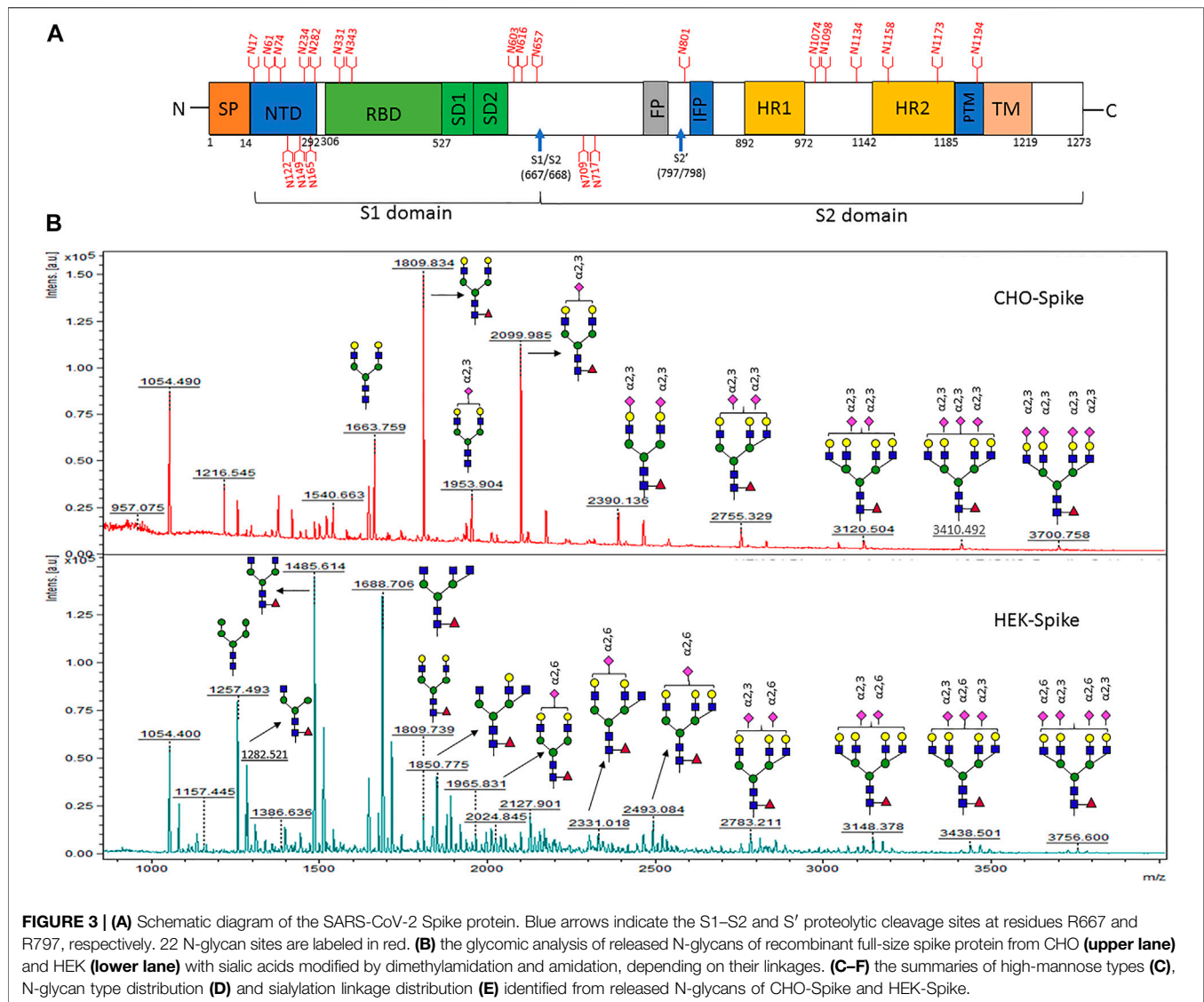
FIGURE 3 | (A) Schematic diagram of the SARS-CoV-2 Spike protein. Blue arrows indicate the S1–S2 and S′ proteolytic cleavage sites at residues R667 and R797, respectively. 22 N-glycan sites are labeled in red. (B) the glycomic analysis of released N-glycans of recombinant full-size spike protein from CHO (upper lane) and HEK (lower lane) with sialic acids modified by dimethylamidation and amidation, depending on their linkages. (C–F) the summaries of high-mannose types (C), N-glycan type distribution (D) and sialylation linkage distribution (E) identified from released N-glycans of CHO-Spike and HEK-Spike.

mass spectrometer (Thermo Fisher Scientific) for each run. The analytical method was adapted from a previous publication (Wang et al., 2021). Briefly, glycopeptides were first loaded and de-salted with a trap column (Thermo Fisher PepMap™, C18, 3 μm, 100 Å, 75 μm × 2 cm) at 5 μL/min with 100% Solvent A (0.1% formic acid in HPLC water) for 5 min. Then, glycopeptides were separated by an Accalaim PepMap™ 100 nano column (3 μm, 100 Å, 75 μm × 250 mm) using a linear gradient of 2.5–37.5% solvent B (80% ACN, 0.1% formic acid) over 85 min, with a wash at 90% B for 5 min. Data-dependent analysis (DDA) was carried out with a duty cycle of 2 s. Precursor masses were detected in the Orbitrap at resolution (R) = 120,000 (at m/z 200) with internal calibration (Easy IC). Stepped HCD spectra (HCD energy at 15, 25, and 35%) were acquired for precursors with charges between two and eight and intensities over 5.0×10^4 at R = 30,000. Dynamic exclusion was set at 20 s. If at least one of the three common glycan oxonium fragment ions (m/z 138.0545, 204.0867, and 366.1396 Da) was observed within

15 ppm mass accuracy, EThcD acquisition were performed in the orbitrap at R = 30,000. The electron-transfer dissociation (ETD) reagent target was 2.0×10^5, with supplemental collision energy at 15%. The ETD reaction time was dependent on the precursor charge state: 125 ms (ETD reaction time) for charge 2, 100 ms for 3, 75 ms for 4, and 50 ms for ≥5 (Wang et al., 2021).

Data Processing and Bioinformatics Analysis

Peptide identification was performed using Byonic version 4.0 and glycopeptide quantification and characterization was performed by Byologic version 4.0 (Protein Metrics Inc., San Carlos, CA). The Byonic software parameters are listed in Supplemental Table S1. The precursor mass tolerance was 10 ppm, and the fragment mass tolerance was 15 ppm. The manual score cutoff was 50, the PEP2D score was less than
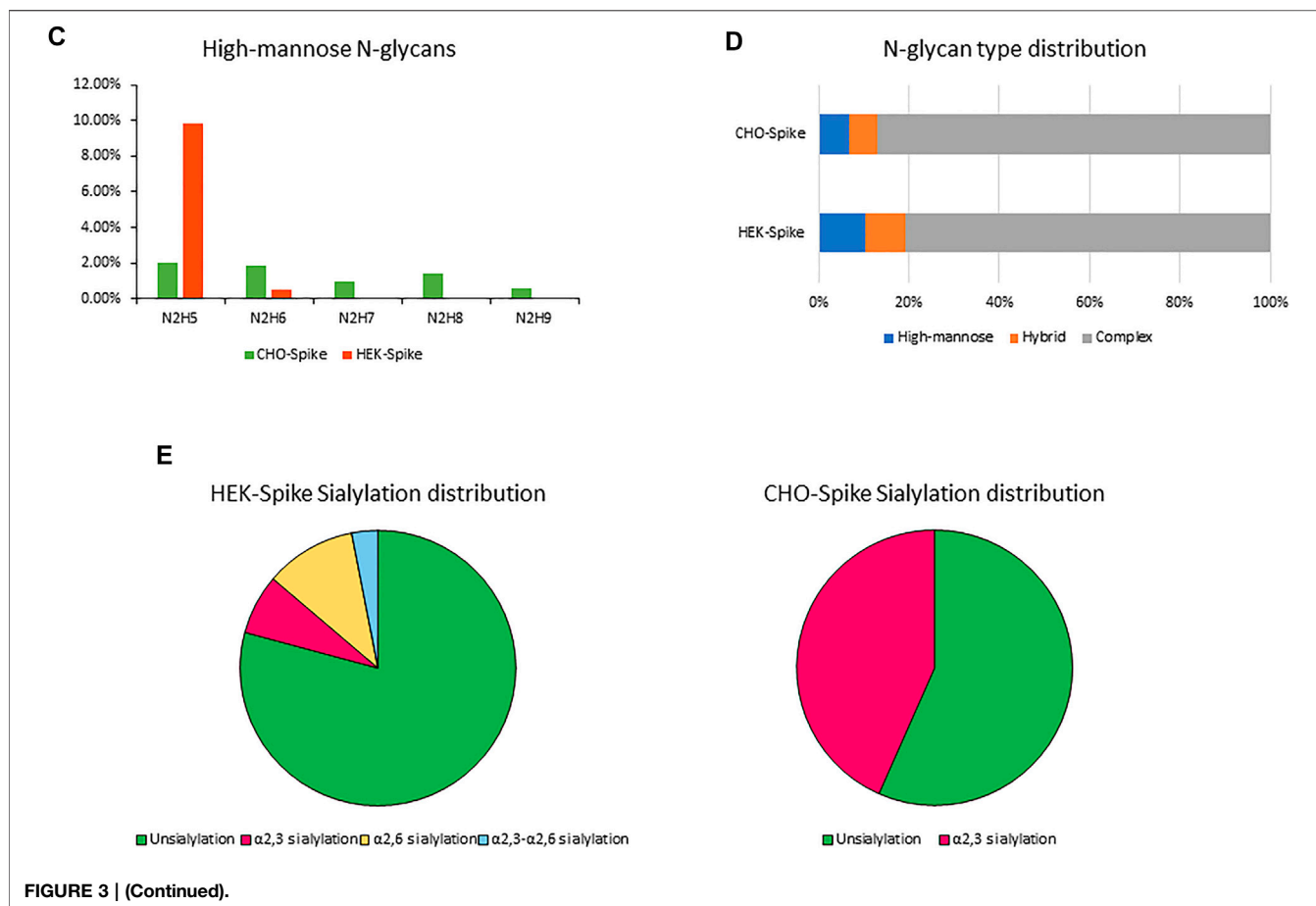
FIGURE 3 | (Continued).

0.5, and the protein false discovery rate (FDR) was 1%. To accommodate for amide-modified Neu5NAc and Neu5Gc masses, as incurred with use of dimethylamine and ammonium hydroxide, we built a mammalian N-glycan database to include these modifications. Dimethylamine can also modify the carboxyl groups on Aspartic acid (D), Glutamic acid (E) and the C-terminal of peptides. We noted this derivatization as a fixed modification in the Byonic search. We performed data searches using dimethylamination only (Option 1) and both dimethylamination and amidation (Option 2) on the same LC-MS/MS experimental dataset using modified fetuin. The search result from Option two showed that D and E are principally modified by dimethylamination and only a trace amount was modified by amidation. According to these results, we use Option one in Byonic as the modification of D, E, and C-terminal of glycopeptides. Using Byonic search results as inputs, Byologic can compute extracted ion chromatograms (XICs) and relative abundances using a label-free quantification approach, a default setting in the Byologic software. The relative abundance of individual glycoforms on specific glycopeptides was calculated by XIC intensities over all charge states. The results of Byologic analysis were further processed and organized using the RStudio program software (Team, 2020). The mass spectrometry glycoproteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD027536 (Deutsch et al., 2017) at http://www.ebi.ac.uk/pride.

RESULTS AND DISCUSSION

Evaluation of Differential Sialic Acid Linkage Labeling of Fetuin Glycopeptides

Several derivatization methods have recently been developed to effectively stabilize sialic acid residues and distinguish the sialyl linkage isomers (Holst et al., 2016; Yang et al., 2018). For example, the application of dimethylamine ($NH(CH_3)_2$) with carboxylic acid activator EDC and the catalyst HOBt in a DMSO solution can result in lactonization and dimethylamidation for individual α2,3- and α2,6-linked sialic acids, respectively (**Figure 1B**) (de Haan et al., 2015). Furthermore, the two-step derivatization by dimethylamine and ammonium hydroxide improved this dimethylamine derivation stability by converting α2,3-sialic acid lactonization into amidation in ammonium hydroxide solution while α2,6-sialic acid dimethylamidation remains stable (Holst et al., 2016; Zhou et al., 2019). This two-step method modified the sialic acid and made it easier to detect different linkages of sialic acids on glycans using MALDI-TOF MS (Zhou et al., 2019). In the current study, we applied this two-step sialic acid derivatization method to glycopeptide analysis

TABLE 1 | Released N-glycans identified from CHO-Spike protein.

Experimental Mass [M + Na]+	Theoretical Mass [M + Na]+	Delta Mass	Composition	Relative percentage
1136.610	1136.40	0.209	N3H3	0.56%
1257.570	1257.43	0.142	N2H5	2.04%
1282.621	1282.46	0.161	N3H3F1	0.58%
1298.591	1298.45	0.137	N3H4	0.77%
1419.630	1419.48	0.149	N2H6	1.86%
1444.633	1444.51	0.121	N3H4F1	0.79%
1460.649	1460.51	0.142	N3H5	0.71%
1485.683	1485.54	0.145	N4H3F1	1.22%
1501.680	1501.53	0.146	N4H4	1.36%
1581.713	1581.53	0.179	N2H7	0.93%
1588.703	1588.57	0.137	S(23)1-N3H4	0.37%
1647.757	1647.59	0.165	N4H4F1	4.00%
1663.759	1663.59	0.172	N4H5	7.09%
1743.668	1743.59	0.082	N2H8	1.41%
1750.754	1750.62	0.135	S(23)1-N3H5	0.62%
1791.775	1791.65	0.129	S(23)1-N4H4	0.85%
1809.834	1809.64	0.190	N4H5F1	22.89%
1825.837	1825.64	0.198	N4H6	0.77%
1850.848	1850.67	0.177	N5H4F1	0.47%
1905.804	1905.64	0.165	N2H9	0.56%
1930.790	1930.67	0.120	N3H7F1	1.78%
1937.910	1937.70	0.207	S(23)1-N4H4F1	2.21%
1953.904	1953.70	0.206	S(23)1-N4H5	4.43%
2012.927	2012.72	0.203	N5H5F1	1.38%
2028.911	2028.72	0.193	N5H6	0.88%
2099.985	2099.76	0.229	S(23)1-N4H5F1	19.53%
2115.973	2115.75	0.222	S(23)1-N4H6	0.79%
2175.009	2174.78	0.232	N5H6F1	3.39%
2244.063	2243.81	0.253	S(23)2-N4H5	0.54%
2303.057	2302.84	0.222	S(23)1-N5H5F1	0.69%
2319.111	2318.83	0.281	S(23)1-N5H6	0.65%
2378.122	2377.86	0.266	N6H6F1	0.37%
2390.136	2389.87	0.268	S(23)2-N4H5F1	3.51%
2465.166	2464.89	0.278	S(23)1-N5H6F1	2.60%
2540.199	2539.91	0.290	N6H7F1	0.79%
2593.117	2592.95	0.170	S(23)2-N5H5F1	0.27%
2609.265	2608.94	0.323	S(23)2-N5H6	0.32%
2668.270	2667.97	0.303	S(23)1-N6H6F1	0.22%
2755.329	2755.00	0.329	S(23)2-N5H6F1	2.17%
2830.358	2830.02	0.338	S(23)1-N6H7F1	0.87%
3045.463	3045.11	0.352	S(23)3-N5H6F1	0.72%
3120.504	3120.13	0.372	S(23)2-N6H7F1	0.97%
3410.492	3410.24	0.249	S(23)3-N6H7F1	0.64%
3700.758	3700.35	0.403	S(23)4-N6H7F1	0.42%

using a solid support system. The specific sialic acid linkages on N-glycans of fetuin from fetal bovine serum were first evaluated as a model protein using a solid support that facilitates glycomic and glycoproteomic measurements. The differential sialic acid linkage labeling method used to generate the fetuin N-glycan profile is shown in **Supplementary Figure S3**. The resulting detailed released N-glycan quantification is listed in **Supplementary Table S2**, which is consistent with the fetuin sialylated N-glycan profiles reported in the literature (Nwosu et al., 2011; Yang and Zhang, 2014), and in agreement with reports that differentiate between terminal sialic acid linkage types identified by alternative labeling strategies (Yang et al., 2017) and high field NMR (Green et al., 1988; Hayase et al., 1992).

Next, this two-step differential sialic acid labeling method was performed on fetuin glycopeptides using the solid support system. To determine whether sialic acids on intact glycopeptides were successfully derivatized by dimethylamine/NH$_4$OH, we inspected the fragment ions of tandem mass spectra. Representative glycopeptides KLC [+57]PD [+27.047]C [+57]PLLAPLN [2886.0791]DSR and VVHAVE [+27.0470]VALATFNAESN [3279.2423]GSYLQLVEISR were observed as shown in **Supplementary Figure S4** (upper lanes), which contain oxonium ions bearing mass shifts with both dimethylamine and NH$_4$OH modified Neu5Ac ions. Other oxonium ions, C6H8NO2, C7H8NO2, HexNAc-H$_2$O, HexNAc, HexNAcHex and HexNAxHex (2) were also observed. These two glycopeptides have only one N-glycan modified by dimethylamine and NH$_4$OH. The fragment peaks are present at 308.289 Da for NeuAc-H$_2$O + NH$_3$ (α2,3-linked) and 336.32 Da for NeuAc-H$_2$O + NH(CH$_3$)$_2$ (α2,6-linked) as shown in **Supplementary Figure S4**. These results verified that sialic acids are successfully

TABLE 2 | Released N-glycans identified from HEK-Spike protein.

Experimental Mass [M + Na]+	Theoretical Mass [M + Na]+	Delta Mass	Composition	Relative percentage
1136.451	1136.402	0.050	N3H3	1.40%
1257.493	1257.428	0.065	N2H5	9.80%
1282.521	1282.459	0.062	N3H3F1	5.18%
1298.492	1298.454	0.037	N3H4	0.56%
1339.535	1339.481	0.054	N4H3	0.88%
1419.536	1419.481	0.056	N2H6	0.54%
1444.576	1444.512	0.064	N3H4F1	1.24%
1485.614	1485.539	0.075	N4H3F1	18.34%
1501.581	1501.534	0.047	N4H4	0.70%
1606.617	1606.565	0.052	N3H5F1	0.45%
1647.670	1647.592	0.078	N4H4F1	4.87%
1688.706	1688.618	0.088	N5H3F1	16.54%
1793.718	1793.65	0.068	N4H4F2	0.77%
1809.739	1809.644	0.094	N4H5F1	2.36%
1834.773	1834.676	0.097	N5H3F2	0.72%
1850.775	1850.671	0.103	N5H4F1	4.73%
1891.802	1891.698	0.104	N6H3F1	3.65%
1937.811	1937.703	0.108	S(23)1-N4H4F1	0.92%
1955.817	1955.702	0.115	N4H5F2	0.61%
1965.831	1965.734	0.097	S(26)1-N4H4F1	1.20%
1996.848	1996.729	0.119	N5H4F2	1.18%
2012.837	2012.724	0.113	N5H5F1	1.44%
2037.871	2037.756	0.115	N6H3F2	0.47%
2053.867	2053.75	0.117	N6H4F1	1.14%
2099.891	2099.756	0.136	S(23)1-N4H5F1	1.26%
2127.901	2127.787	0.114	S(26)1-N4H5F1	2.41%
2140.901	2140.782	0.118	S(23)1-N5H4F1	1.03%
2168.922	2168.813	0.109	S(26)1-N5H4F1	1.54%
2174.914	2174.777	0.137	N5H6F1	0.99%
2215.949	2215.803	0.145	N6H5F1	0.61%
2243.935	2243.809	0.126	S(23)2-N4H5	0.37%
2303.021	2302.835	0.186	S(23)1-N5H5F1	0.85%
2331.018	2330.866	0.152	S(26)1-N5H5F1	0.87%
2372.004	2371.893	0.111	S(26)1-N6H4F1	0.44%
2418.039	2417.898	0.141	S(23)1S(26)1-N4H5F1	0.39%
2446.047	2445.929	0.118	S(26)2-N4H5F1	0.35%
2465.069	2464.888	0.181	S(23)1-N5H6F1	0.91%
2493.084	2492.919	0.165	S(26)1-N5H6F1	1.51%
2506.124	2505.915	0.209	S(23)1-N6H5F1	0.39%
2534.118	2533.946	0.173	S(26)1-N6H5F1	0.60%
2668.099	2667.967	0.132	S(23)1-N6H6F1	0.29%
2755.228	2754.999	0.229	S(23)2-N5H6F1	0.46%
2783.211	2783.03	0.181	S(26)1S(23)1-N5H6F1	0.84%
2811.223	2811.061	0.161	S(26)2-N5H6F1	0.60%
2830.265	2830.02	0.245	S(23)1-N6H7F1	0.30%
2852.277	2852.088	0.189	S(26)2-N6H5F1	0.20%
2858.258	2858.051	0.207	S(26)1-N6H7F1	0.52%
2986.321	2986.11	0.211	S(23)1S(26)1-N6H6F1	0.22%
3101.333	3101.173	0.160	S(23)1S(26)2-N5H6F1	0.24%
3120.382	3120.132	0.250	S(23)2-N6H7F1	0.28%
3148.378	3148.163	0.216	S(26)1S(23)1-N6H7F1	0.59%
3176.373	3176.194	0.180	S(26)2-N6H7F1	0.47%
3438.501	3438.274	0.227	S(26)1S(23)2-N6H7F1	0.30%
3466.510	3466.305	0.205	S(26)2S(23)1-N6H7F1	0.27%
3756.600	3756.416	0.184	S(26)2S(23)2-N6H7F1	0.13%
3784.596	3784.447	0.148	S(26)3S(23)1-N6H7F1	0.09%

modified by dimethylamidation and amidation. Moreover, compared to the native fetuin glycopeptide (**Supplementary Figure S4** lower panels) under the same experimental condition, we observed that more oxonium ions were detected and at higher intensity when dimethylamine and

NH₄OH modified Neu5Ac for the labeled fetuin glycopeptides compared to unmodified Neu5Ac of native fetuin glycopeptides. Additionally, we also noticed that the labeled glycopeptide has a higher mass charge state, which may be due to the dimethylamidation of the carboxyl groups of
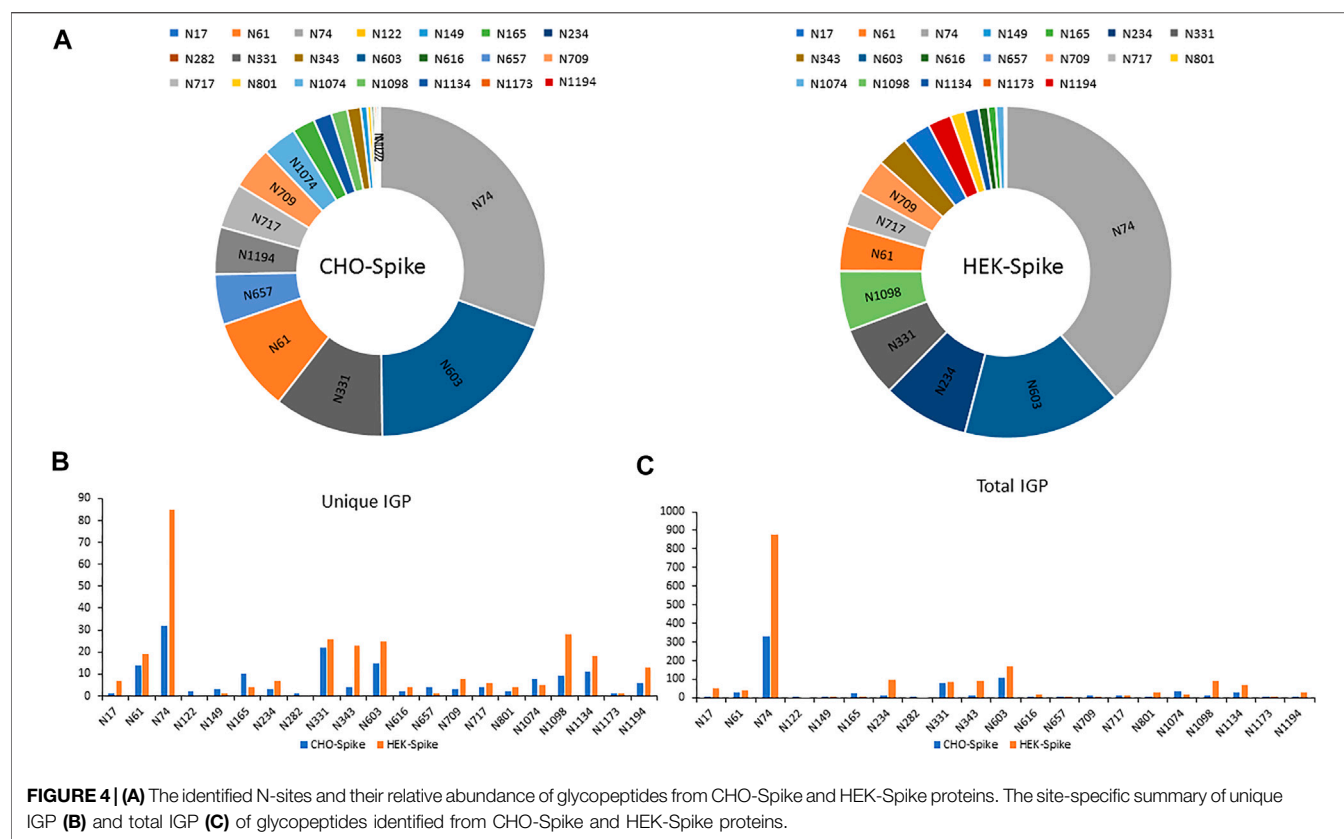
FIGURE 4 | (A) The identified N-sites and their relative abundance of glycopeptides from CHO-Spike and HEK-Spike proteins. The site-specific summary of unique IGP **(B)** and total IGP **(C)** of glycopeptides identified from CHO-Spike and HEK-Spike proteins.

Aspartic acid (D), Glutamic acid (E) and the C-terminal of peptides, where primary amines form $-NH_3^+$ ions at a lower pH (i.e., 0.1% TFA).
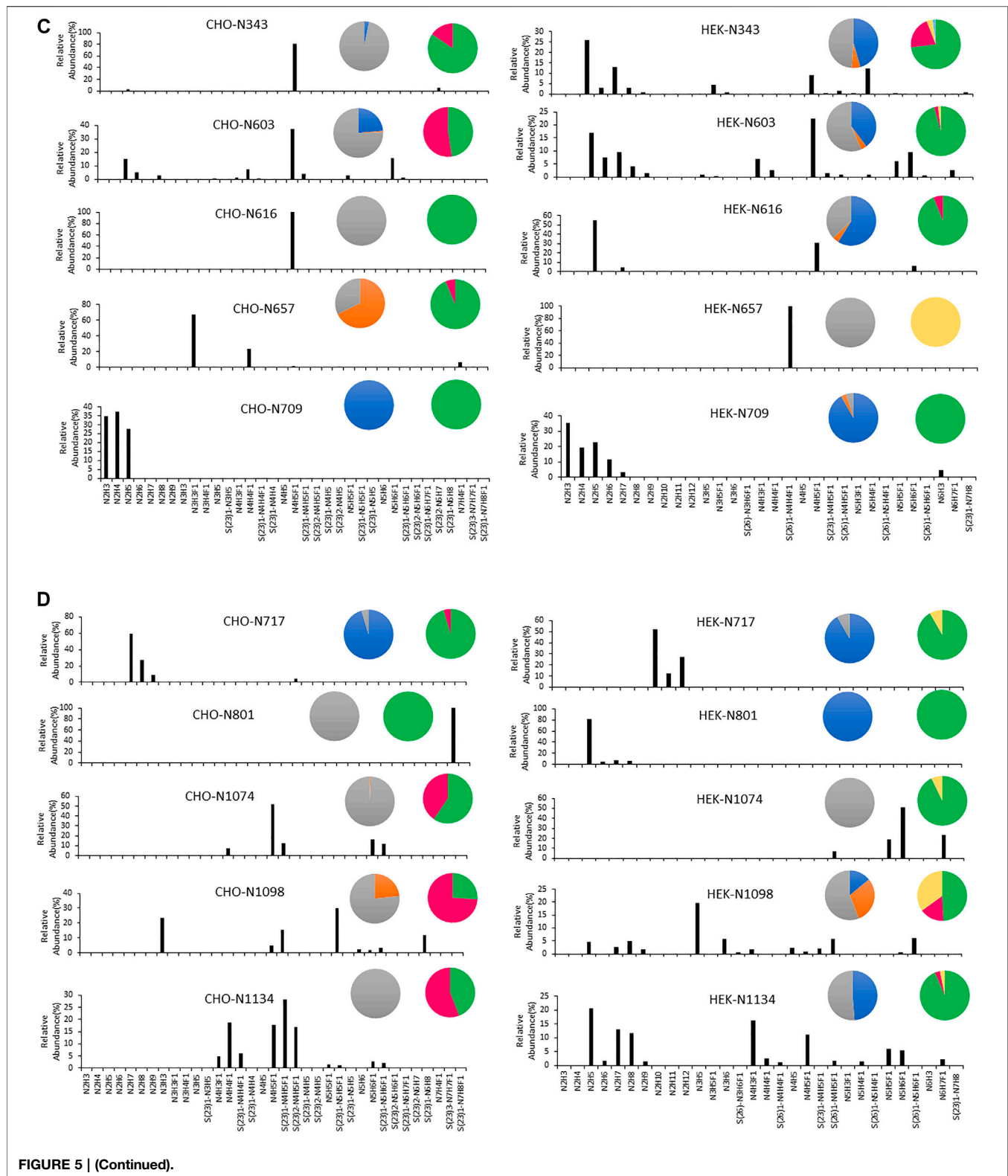
Thus, by using 3 µg of the modified and unmodified intact glycopeptides, performed with three analytic repeats of fetuin, the total and unique intact glycopeptides (IGP) identified for the labeled samples were 1,353 and 126, compared to 747 and 57 in native sample separately. The site-specific total and unique IGP distribution are presented in **Figures 2A,B**, in which the sialic acid labeled glycopeptide showed more IGP identified compared to native glycopeptides at each N-glycan site. As mentioned above, derivatization can also occur on the carboxylic acids of D, E, or the C-terminus. **Supplementary Table S3** lists the relative abundance of glycopeptides containing D and/or E residues modified by dimethylamine in the analysis of fetuin. The glycopeptides at N99 and N179 sites contain a higher number of D and E residues, and the signal strength is higher. On the other hand, the glycopeptide at N156 site have fewer D and E residues, resulting in lower charge state and signal intensity in EThcD fragmentation. Moreover, the site-specific unique sialopeptide analysis (**Figure 2C**) identified more unique sialopeptides in modified glycopeptides versus unmodified fetuin owing to the linkage specific mass shifts imparted by the labeling strategy. Furthermore, the identified linkage specific unique sialopeptides account for the majority of the identified unique IGPs (**Figures 2B,C**), indicating that fetuin is a highly sialylated glycoprotein. Over 75 unique N-glycans were
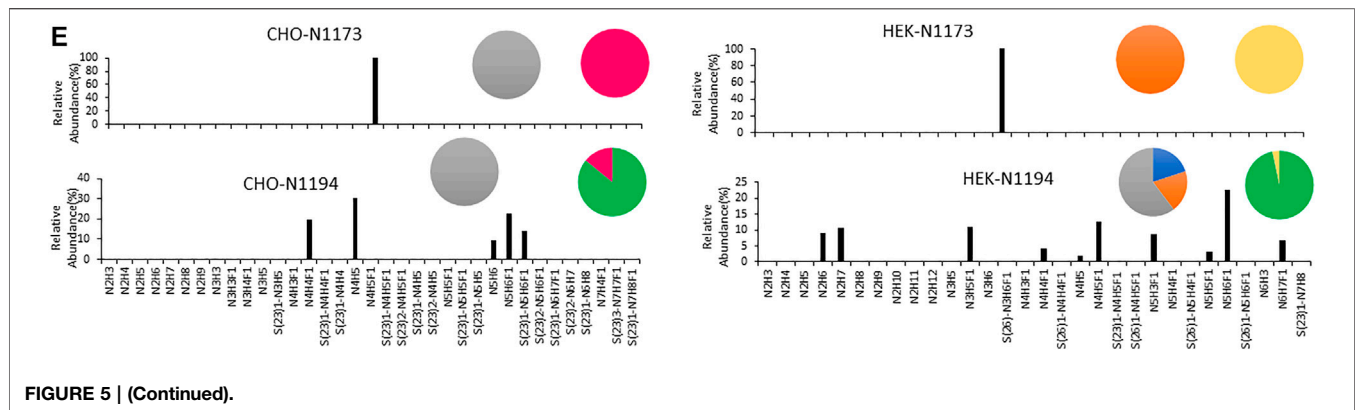
identified in modified fetuin and 24 unique N-glycans without sialylation linkage information were identified for native fetuin with the detailed site-specific N-glycan profile tabulated in **Supplementary Table S4**.

Characterization of Recombinant SARS-CoV-2 Spike Proteins Secreted From CHO or HEK Cell Culture

After evaluating the dimethylamine-NH4OH sialic acid linkage labeling method on fetuin, we next applied this method for the glycomic and glycoproteomic analysis of full-size recombinant SARS-CoV-2 spike protein expressed by either CHO-GS or HEK293 suspension cells. Currently, CHO and HEK cells are the two predominant production platforms for recombinant soluble S protein. Given that CHO and HEK have distinctive glycosylation processing pathways (Croset et al., 2012), the resulting soluble S proteins may produce different N-glycan profiles. Indeed, due to the lack of α2,6-sialyltransferase expression, CHO cells usually only have α2,3-sialylated glycans (Hossler et al., 2009), while HEK cells typically have both α2,3- and α2,6-linked sialic acids. CHO cells may also produce small amounts of Neu5Gc, which can be recognized as a foreign epitope and cause immunogenicity in humans (Borys et al., 2010). A schematic illustration of the different regions of SARS-CoV-2 S protein is presented in **Figure 3A**. According to literature, the SARS-CoV-2 S protein is heavily glycosylated and has 22 N-glycosites (Watanabe et al., 2020b) as designated in **Figure 3A**. Glycan

FIGURE 5 | The site-specific major N-glycan distribution and comparison of SARS-COV-2 spike from CHO and HEK cells, respectively. The pie charts summarized the N-glycan type distribution and the sialylation linkage information identified at each site. The key is shown in **(A)**. Glycosite comparisons are shown as follows: **(A)** Glycosites N17, N61, N74 and N122, **(B)** N149, N165, N234, N282 and N331, **(C)** N343, N603, N616, N657 and N709, **(D)** N717, N801, N1074, N1098 and N1134, and **(E)** N1173 and N1194. The complete lists of N-glycans at each site were presented in **Supplementary Table S5**.

FIGURE 5 | (Continued).

shielding provides a thick sugar-coated barrier against neutralizing antibody recognition. Consequently, global and site-specific deciphering of N-glycosylation of SARS-CoV-2 S

protein from different mammalian expression hosts will help characterize recombinant spike proteins, diagnose COVID-19, and develop protein vaccine.

FIGURE 5 | (Continued).

The released N-glycan MS profile from three glycopeptide analytical replicates of the spike proteins from CHO and HEK hosts are shown in **Figure 3B**. The determined glycan compositions and percentage levels are listed in **Tables 1**, **2**. In total, 44 and 56 N-glycan types were identified in the spike proteins from CHO and HEK cells, respectively. The distribution of high-mannose and N-glycan types is presented in **Figures 3C,D**. CHO-Spike mainly displays unsialylated and α2,3-monosialyated complex N-glycans, while HEK-Spike shows more high-mannose glycans, mainly Man₅GlcNAc₂, with less distribution between large and small MW glycans compared with CHO-Spike. HEK-Spike also mainly contains unsialylated bi- and tri-antennary N-glycans and a small amount of various α2,3- and α2,6-sialylated complex types. In accordance with our result in **Figures 3C,D**, another group also found that high oligomannose and highly fucosylated bi- and tri-antennary N-glycans have low sialylation content on the spikes of HEK cells. However, here we used solid-phase sialic acid linkage method to improve the content of the information (Zhao et al., 2020). Indeed, different expression systems and protein constructs may present different glycan patterns. The α2,3- and α2,6-sialylation distribution summary from glycomic analysis (**Figure 3E**) also reveals that HEK-Spike contained 7 and 11% exclusive α2,3- and α2,6-sialylation individually and 3% mixed sialylation, while CHO-Spike only bears 43% α2,3-sialylation. Meanwhile, the site-specific N-glycosylation analysis of S protein was performed under the co-digestion of trypsin and chymotrypsin. The representative MS/MS spectra are presented in **Supplementary Figure S5**. We identified the glycan compositions at 21 and 19 out of the 22 predicted N-glycosylation sites of SARS-CoV-2 S proteins in CHO and HEK, respectively. The determined N-glycosites and their relative abundance based on LC-MS signal intensity (percentage of total glycopeptide LC-MS intensity) are displayed in **Figure 4A**. We quantified the relative intensity of glycans at each site by comparing the area under the curve of each glycopeptide peak on the LC-MS chromatogram. Interestingly, in this study, the CHO derived spike N-glycosite at 1,158 position (N1158) was unoccupied while the N-glycosites at 122, 282 and 1,158 positions (N122, N282 and N1158) were found unoccupied in the HEK derived spike.

Another spike protein glycopeptide mapping analysis identified 17 out of 22 predicted N-sites from SARS-CoV-2 to be occupied when S1 and S2 protein fragments were expressed separately from HEK cells using a stepped HCD (higher-energy collisional dissociation) triggered CID (collision-induced dissociation) program (Shajahan et al., 2020) with 17, 603, 1,134, 1,158 and 1,173 positions unoccupied. We identified 19 out of 22 N-sites from full-size spike from HEK cells to be occupied. The identified differences may be attributed to the dissimilar LC-MS/MS fragmentation program applied, the sialic acids status as labeled or not, the size of proteins analyzed (full-size spike or individual S1 and S2 proteins), and the polypeptide processing in ER and *Golgi* apparatus. Alternatively, the N74 site represents the most abundantly occupied N-site of Spike from both CHO and HEK cells as displayed in **Figures 4B,C** for site-specific total and unique IGP numbers summed from three analytical repeats. Furthermore, the HEK-Spike protein analysis identified more unique IGPs at N74 site, indicating the site may be more heavily glycosylated on HEK-Spike than CHO-Spike.

In addition, the distribution of site-specific N-glycan on each identified N-glycosite from CHO and HEK cells is displayed in **Figure 5** and compared with the site-specific N-glycan compositions and percentages listed in **Supplementary Table S5**. Consistent with the global N-glycan analysis, we observed more N-glycan types and more high-mannose content on the S glycopeptides from HEK cells, especially in N17, N234, N343, N616, N709, N801, and N1134 sites. The higher abundance of high mannose glycans we observed at the listed sites may be due to differences in the inherent glycosylation propensities of the 2 cell substrates. It should be noted that the absence of complex N-glycans has been reported to reduce the entry of viruses into HEK-ACE2 expressing cells by more than 95% (Yang et al., 2020). Complex N-glycans mainly exist at the N74, N331, N1074, N1098 and N1194 sites on HEK-Spike. Most spike glycopeptides from HEK cells tend to be unsialylated, with a small amount of α2,3- and α2,6- or both α2,3 and α2,6-sialylation linkage types at the N17, N64 and N343 sites. The analysis of the glycans identified on the native-like SARS-CoV-2 virus revealed a high abundance of high-mannose type and complex N-glycans with low sialylation content (Casalino et al., 2020; Grant et al., 2020; Watanabe et al., 2021). Alternatively, CHO-Spike mainly presents complex

N-glycans as the principal glycan form on these and most other sites. Indeed, CHO-Spike is highly sialylated at N61, N74, N343, N1098 and N1134 sites, exclusively containing α2,3-sialylation. Additionally, the level of Neu5Gc detected at the N331 site (0.11% at this site) of spike from CHO cells was negligible. Furthermore, as a heavily glycosylated site on CHO-Spike and HEK-Spike, N74 presents highly α2,3-sialylated complex glycans on CHO-Spike and displays mainly unsialylated complex glycans with some high-mannose and hybrid N-glycans on HEK-Spike. Consequently, our results indicate that the N-glycosylation in full-length spike protein generated by HEK cells, is significantly different from other constructs, such as S1 and S2 independently expressed constructs as well as spike generated from CHO cells. However, separate serum ELISA assay using COVID antigens from either transient HEK and CHO lines revealed that there is no significant difference in the antibody binding response as a function of dilution level, so both antigens may be useful reagents for immunoassay detection (data not shown).

Finally, the S proteins analyzed here were engineered to contain a C-terminal trimerization domain as the transmembrane domain was removed. The trimerization domain was put in place to mimic trimerization in the engineered soluble form that would otherwise occur in the native transmembrane spike. Due to these changes glycosylation patterns may differ from those that would be observed in the TM containing native spike expressed in HEK and CHO cells.

CONCLUSION

In this work, recombinant full-length SARS-CoV-2 spike was produced transiently in suspension CHO and HEK hosts. We then provided global and site-specific analysis of N-linked glycosylation on the soluble full-size S glycoprotein. To evaluate the sialic acid linkages present on the spike glycans, a two-step derivatization by dimethylamine and ammonium hydroxide was performed using a solid support system to differentially modify the sialic acid linkages on the resulting glycans and glycopeptides from the two hosts. We identified the glycan compositions at 21 and 19 out of the 22 predicted N-glycosylation sites of the SARS-CoV-2 S proteins in CHO and HEK, respectively. The N-glycan site at 1,158 position (N1158) was found unoccupied on spike from CHO and the N-glycan sites at 122, 282 and 1,158 positions (N122, N282 and N1158) were found unoccupied on spike secreted from HEK cells. Structural mapping of glycans of recombinant full-length human S proteins revealed that CHO-Spike presented more complex and higher sialylation (α2,3-linked only; ~40% of total) content while HEK-Spike had marginally more high-mannose glycan, almost exclusively as $Man_5GlcNAc_2$, and minor amounts of α2,3- and α2,6 linked sialylation (less than 15% total). The N74 site represents the most heavily and compositionally diverse N-glycosylated site on both spike proteins. The abundant high-mannose sites (N17, N234, N343, N616, N709, N717, N801 and N1134) on HEK-Spike may serve as glycan shields and offer viral evasion properties and/or provide targets for

mannose specific immune system lectins (Hsieh et al., 2018; Rahimi, 2020). Alternatively, the complex type N-glycans presenting at N74, N331, N1074, N1098 and N1194 sites on HEK-Spike may facilitate viral entry into ACE2 expressing cells (Yang et al., 2020) as well as serve as glycan shield. Collectively, these data underscore that certain N-glycan sites offer distinctive glycosylation patterns regardless of the host, while other sites exhibit site-specific differences in glycosylation in different mammalian hosts. These distinct site-specific N-glycan profiles may impact viral behavior *in vivo* and perhaps impact vaccine development, and these differences may have no problem or impact on specific diagnostic applications.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD027536.

AUTHOR CONTRIBUTIONS

QW carried out experiments, provided scientific arguments, processed data and provided major contributions toward manuscript construction. YW designed and performed glycoproteomics experiments and data processing. SY provided expert advice in derivatization chemistries and solid phase design of amidations. CL performed mass spectrometry analysis and data processing. LA performed design, construction and expression of spike proteins YC performed design, construction and expression of spike proteins LP provided expert advise concerning spike structure including glycosylation TY provided expert advise concerning mass spectrometry experimental design. HJ provided expert advice concerning SARS CoV 2 expression and regulation. AP served as led for spike expression and serum testing MB provided expert advice concerning CHO and HEK293 expression systems, glycobiology, and manuscript construction. JC provided overall project focus, experimental design, development of amidation chemistries, project oversight and manuscript construction.

FUNDING

REFERENCES

Amanat, F., Stadlbauer, D., Strohmeier, S., Nguyen, T. H. O., Chromikova, V., McMahon, M., et al. (2020). A Serological Assay to Detect SARS-CoV-2 Seroconversion in Humans. *Nat. Med.* 26, 1033–1036. doi:10.1038/s41591-020-0913-5

A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, et al. (2015). "Essentials of Glycobiology," in *The Consortium of Glycobiology Editors* (La Jolla, California(NY): Cold Spring Harbor Laboratory Press Copyright 2015-2017All rights reserved., Cold Spring Harbor).

Bangaru, S., Ozorowski, G., Turner, H. L., Antanasijevic, A., Huang, D., Wang, X., et al. (2020). Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. *Science* 370, 1089–1094. doi:10.1126/science.abe1502

Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009). Activation of the SARS Coronavirus Spike Protein *via* Sequential Proteolytic Cleavage at Two Distinct Sites. *Proc. Natl. Acad. Sci.* 106, 5871–5876. doi:10.1073/pnas.0809524106

Borys, M. C., Dalal, N. G., Abu-Absi, N. R., Khattak, S. F., Jing, Y., Xing, Z., et al. (2010). Effects of Culture Conditions on N-Glycolylneuraminic Acid (Neu5Gc) Content of a Recombinant Fusion Protein Produced in CHO Cells. *Biotechnol. Bioeng.* 105, 1048–1057. doi:10.1002/bit.22644

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6, 1722–1734. doi:10.1021/acscentsci.0c01056

Croset, A., Delafosse, L., Gaudry, J.-P., Arod, C., Glez, L., Losberger, C., et al. (2012). Differences in the Glycosylation of Recombinant Proteins Expressed in HEK and CHO Cells. *J. Biotechnol.* 161, 336–348. doi:10.1016/j.jbiotec.2012.06.038

de Haan, N., Reiding, K. R., Haberger, M., Reusch, D., Falck, D., and Wuhrer, M. (2015). Linkage-Specific Sialic Acid Derivatization for MALDI-TOF-MS Profiling of IgG Glycopeptides. *Anal. Chem.* 87, 8284–8291. doi:10.1021/acs.analchem.5b02426

De Leoz, M. L. A., Duewer, D. L., Fung, A., Liu, L., Yau, H. K., Potter, O., et al. (2020). NIST Interlaboratory Study on Glycosylation Analysis of Monoclonal Antibodies: Comparison of Results from Diverse Analytical Methods. *Mol. Cell Proteomics* 19, 11–30. doi:10.1074/mcp.ra119.001677

Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., et al. (2017). The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition. *Nucleic Acids Res.* 45, D1100–d1106. doi:10.1093/nar/gkw936

Ghorbani, M., Brooks, B. R., and Klauda, J. B. (2021). Exploring Dynamics and Network Analysis of Spike Glycoprotein of SARS-COV-2. *Biophysical J.* 120 (14), 2902–2913. doi:10.1016/j.bpj.2021.02.047

Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., et al. (2011). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol.* 85, 4122–4134. doi:10.1128/jvi.02232-10

Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020). Analysis of the SARS-CoV-2 Spike Protein Glycan Shield Reveals Implications for Immune Recognition. *Scientific Rep.* 10, 14991. doi:10.1038/s41598-020-71748-7

Green, E. D., Adelt, G., Baenziger, J. U., Wilson, S., and Van Halbeek, H. (1988). The Asparagine-Linked Oligosaccharides on Bovine Fetuin. Structural Analysis of N-Glycanase-Released Oligosaccharides by 500-megahertz 1H NMR Spectroscopy. *J. Biol. Chem.* 263, 18253–18268. doi:10.1016/s0021-9258(19)81354-6

Guo, Y.-R., Cao, Q.-D., Hong, Z.-S., Tan, Y.-Y., Chen, S.-D., Jin, H.-J., et al. (2020). The Origin, Transmission and Clinical Therapies on Coronavirus Disease 2019 (COVID-19) Outbreak – an Update on the Status. *Mil. Med. Res.* 7, 11. doi:10.1186/s40779-020-00240-0

Han, L., and Costello, C. E. (2013). Mass Spectrometry of Glycans. *Biochem. Mosc.* 78, 710–720. doi:10.1134/s0006297913070031

Hayase, T., Rice, K. G., Dziegielewska, K. M., Kuhlenschmidt, M., Reilly, T., and Lee, Y. C. (1992). Comparison of N-Glycosides of Fetuins from Different Species and Human .alpha.2-HS-Glycoprotein. *Biochemistry* 31, 4915–4921. doi:10.1021/bi00135a024

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271–280. doi:10.1016/j.cell.2020.02.052

Holst, S., Heijs, B., de Haan, N., van Zeijl, R. J. M., Briaire-de Bruijn, I. H., van Pelt, G. W., et al. (2016). Linkage-Specific *In Situ* Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Anal. Chem.* 88, 5904–5913. doi:10.1021/acs.analchem.6b00819

Hossler, P., Khattak, S. F., and Li, Z. J. (2009). Optimal and Consistent Protein Glycosylation in Mammalian Cell Culture. *Glycobiology* 19, 936–949. doi:10.1093/glycob/cwp079

Hsieh, I. N., De Luna, X., White, M. R., and Hartshorn, K. L. (2018). The Role and Molecular Mechanism of Action of Surfactant Protein D in Innate Host Defense against Influenza A Virus. *Front. Immunol.* 9, 1368. doi:10.3389/fimmu.2018.01368

Huang, Y., Yang, C., Xu, X.-f., Xu, W., and Liu, S.-w. (2020). Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol. Sin* 41, 1141–1149. doi:10.1038/s41401-020-0485-4

Leung, H. S. Y., Li, O. T. W., Chan, R. W. Y., Chan, M. C. W., Nicholls, J. M., and Poon, L. L. M. (2012). Entry of Influenza A Virus with a 2,6-Linked Sialic Acid Binding Preference Requires Host Fibronectin. *J. Virol.* 86, 10704–10713. doi:10.1128/jvi.01166-12

Longo, P. A., Kavran, J. M., Kim, M.-S., and Leahy, D. J. (2013). Transient Mammalian Cell Transfection with Polyethylenimine (PEI). *Methods Enzymol.* 529, 227–240. doi:10.1016/b978-0-12-418687-3.00018-5

Nwosu, C. C., Seipert, R. R., Strum, J. S., Hua, S. S., An, H. J., Zivkovic, A. M., et al. (2011). Simultaneous and Extensive Site-specific N- and O-Glycosylation Analysis in Protein Mixtures. *J. Proteome Res.* 10, 2612–2624. doi:10.1021/pr2001429

Pritchard, L. K., Spencer, D. I., Royle, L., Bonomelli, C., Seabright, G. E., Behrens, A.-J., et al. (2015). Glycan Clustering Stabilizes the Mannose Patch of HIV-1 and Preserves Vulnerability to Broadly Neutralizing Antibodies. *Nat. Commun.* 6, 1–11. doi:10.1038/ncomms8479

Rahimi, N. (2020). C-type Lectin CD209L/L-SIGN and CD209/DC-SIGN: Cell Adhesion Molecules Turned to Pathogen Recognition Receptors. *Biology (Basel)* 10. doi:10.3390/biology10010001

Sanda, M., Morrison, L., and Goldman, R. (2021). N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal. Chem.* 93, 2003–2009. doi:10.1021/acs.analchem.0c03173

Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* 30, 981–988. doi:10.1093/glycob/cwaa042

Sommerstein, R., Flatz, L., Remy, M. M., Malinge, P., Magistrelli, G., Fischer, N., et al. (2015). Arenavirus Glycan Shield Promotes Neutralizing Antibody Evasion and Protracted Infection. *PLoS Pathog.* 11, e1005276. doi:10.1371/journal.ppat.1005276

Spriestersbach, A., Kubicek, J., Schäfer, F., Block, H., and Maertens, B. (2015). Purification of His-Tagged Proteins. *Methods Enzymol.* 559, 1–15. doi:10.1016/bs.mie.2014.11.003

Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmeier, S., Arunkumar, G. A., et al. (2020). SARS-CoV-2 Seroconversion in Humans: A Detailed

Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr. Protoc. Microbiol.* 57, e100. doi:10.1002/cpmc.100

Sun, S., Shah, P., Eshghi, S. T., Yang, W., Trikannad, N., Yang, S., et al. (2016). Comprehensive Analysis of Protein Glycosylation by Solid-phase Extraction of N-Linked Glycans and Glycosite-Containing Peptides. *Nat. Biotechnol.* 34, 84–88. doi:10.1038/nbt.3403

Team, R. (2020). *RStudio*. Boston, MA: Integrated Development for R.

Van Breedam, W., Pöhlmann, S., Favoreel, H. W., de Groot, R. J., and Nauwynck, H. J. (2014). Bitter-sweet Symphony: Glycan-Lectin Interactions in Virus Biology. *FEMS Microbiol. Rev.* 38, 598–632. doi:10.1111/1574-6976.12052

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 Update of the PRIDE Database and its Related Tools. *Nucleic Acids Res.* 44, D447–D456. doi:10.1093/nar/gkv1145

Wang, Q., Stuczynski, M., Gao, Y., and Betenbaugh, M. J. (2015). *Strategies for Engineering Protein N-Glycosylation Pathways in Mammalian Cells*. New York, NY: Humana Press, 287–305. doi:10.1007/978-1-4939-2760-9_20

Wang, Y., Wu, Z., Hu, W., Hao, P., and Yang, S. (2021). Impact of Expressing Cells on Glycosylation and Glycan of the SARS-CoV-2 Spike Glycoprotein. *ACS Omega* 6, 15988–15999. doi:10.1021/acsomega.1c01785

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 369, 330–333. doi:10.1126/science.abb9983

Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., et al. (2020). Vulnerabilities in Coronavirus Glycan Shields Despite Extensive Glycosylation. *Nat. Commun.* 11, 2688. doi:10.1038/s41467-020-16567-0

Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of Glycosylation in Enveloped Virus Pathobiology. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1863, 1480–1497. doi:10.1016/j.bbagen.2019.05.012

Watanabe, Y., Mendonça, L., Allen, E. R., Howe, A., Lee, M., Allen, J. D., et al. (2021). Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Cent. Sci.* 7, 594–602. doi:10.1021/acscentsci.1c00080

Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., and Prescott, H. C. (2020). Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19). *JAMA* 324, 782–793. doi:10.1001/jama.2020.12839

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* 367, 1260–1263. doi:10.1126/science.abb2507

Yang, Q., Hughes, T. A., Kelkar, A., Yu, X., Cheng, K., Park, S., et al. (2020). Inhibition of SARS-CoV-2 Viral Entry upon Blocking N-And O-Glycan Elaboration. *Elife* 9, e61552. doi:10.7554/elife.61552

Yang, S., Jankowska, E., Kosikova, M., Xie, H., and Cipollo, J. (2017). Solid-Phase Chemical Modification for Sialic Acid Linkage Analysis: Application to Glycoproteins of Host Cells Used in Influenza Virus Propagation. *Anal. Chem.* 89, 9508–9517. doi:10.1021/acs.analchem.7b02514

Yang, S., Wu, W. W., Shen, R.-F., Bern, M., and Cipollo, J. (2018). Identification of Sialic Acid Linkages on Intact Glycopeptides *via* Differential Chemical Modification Using IntactGIG-HILIC. *J. Am. Soc. Mass. Spectrom.* 29, 1273–1283. doi:10.1007/s13361-018-1931-0

Yang, S., and Zhang, H. (2014). Glycomic Analysis of Glycans Released from Glycoproteins Using Chemical Immobilization and Mass Spectrometry. *Curr. Protoc. Chem. Biol.* 6, 191–208. doi:10.1002/9780470559277.ch140085

Yehuda, S., and Padler-Karavani, V. (2020). Glycosylated Biotherapeutics: Immunological Effects of N-Glycolylneuraminic Acid. *Front. Immunol.* 11, 21. doi:10.3389/fimmu.2020.00021

Yuen, K.-S., Ye, Z. W., Fung, S.-Y., Chan, C.-P., and Jin, D.-Y. (2020). SARS-CoV-2 and COVID-19: The Most Important Research Questions. *Cel Biosci.* 10, 40. doi:10.1186/s13578-020-00404-4

Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., et al. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* 28, 586–601. doi:10.1016/j.chom.2020.08.004

Zhou, X., Yang, G., and Guan, F. (2020). Biological Functions and Analytical Strategies of Sialic Acids in Tumor. *Cells* 9, 676–680. doi:10.3390/cells9020273

Zhou, X., Yang, S., Yang, G., Tan, Z., and Guan, F. (2019). Two-step Derivatization and Mass Spectral Distinction of α2,3 and α2,6 Sialic Acid Linkages on N-Glycans by MALDI-TOF. *Chin. Chem. Lett.* 30, 676–680. doi:10.1016/j.cclet.2018.12.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past co-authorship with the authors (QW, MB, YC).

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Direct N-Glycosylation Profiling of Urine and Prostatic Fluid Glycoproteins and Extracellular Vesicles

Calvin R. K. Blaschke[1], Jordan P. Hartig[1], Grace Grimsley[1], Liping Liu[1], O. John Semmes[2,3], Jennifer D. Wu[4], Joseph E. Ippolito[5], Chanita Hughes-Halbert[6,7], Julius O. Nyalwidhe[2,3] and Richard R. Drake[1,5]*

[1]Department of Cell and Molecular Pharmacology and Experimental Therapeutics, Medical University of South Carolina, Charleston, SC, United States, [2]Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, Norfolk, VA, United States, [3]The Leroy T. Canoles Jr., Cancer Research Center, Eastern Virginia Medical School, Norfolk, VA, United States, [4]Departments of Urology and Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL, United States, [5]Department of Radiology, Washington University School of Medicine, St. Louis, MO, United States, [6]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, SC, United States, [7]Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, United States

Expressed prostatic secretions (EPS), also called post digital rectal exam urines, are proximal fluids of the prostate that are widely used for diagnostic and prognostic assays for prostate cancer. These fluids contain an abundant number of glycoproteins and extracellular vesicles secreted by the prostate gland, and the ability to detect changes in their N-glycans composition as a reflection of disease state represents potential new biomarker candidates. Methods to characterize these N-glycan constituents directly from clinical samples in a timely manner and with minimal sample processing requirements are not currently available. In this report, an approach is described to directly profile the N-glycan constituents of EPS urine samples, prostatic fluids and urine using imaging mass spectrometry for detection. An amine reactive slide is used to immobilize glycoproteins from a few microliters of spotted samples, followed by peptide N-glycosidase digestion. Over 100 N-glycan compositions can be detected with this method, and it works with urine, urine EPS, prostatic fluids, and urine EPS-derived extracellular vesicles. A comparison of the N-glycans detected from the fluids with tissue N-glycans from prostate cancer tissues was done, indicating a subset of N-glycans present in fluids derived from the gland lumens. The developed N-glycan profiling is amenable to analysis of larger clinical cohorts and adaptable to other biofluids.

Keywords: glycosylation, N-glycan, prostate, urine, MALDI

INTRODUCTION

In the search and characterization of disease biomarkers for use in liquid biopsy applications, proximal fluids like blood and urine are commonly used. Proximal fluids are found adjacent to a given tissue or organ and represent a repertoire of secreted proteins and shed cells reflective of the physiological state of that tissue. For prostate cancer and other genitourinary diseases, proximal fluids are represented by seminal plasma and expressed-prostatic secretion in urine (EPSu) (Drake

and Kislinger, 2014; Nawaz et al., 2014). EPSu, also termed post-digital rectal exam (DRE) urine, represents the fluid being secreted by the prostate following a digital rectal prostate massage, which in turn can be collected in voided urine post-exam (Drake et al., 2009; Drake and Kislinger, 2014). The prostate gland secretes many proteins and other biomolecules in a prostatic fluid that combines with seminal fluid and sperm from the seminal vesicles during ejaculation. Many of these prostatic proteins are glycoproteins, like prostate specific antigen (PSA), that perform functions to activate sperm and suppress the vaginal immune micro-environment. Our group has previously characterized the proteomic composition of EPSu and prostatic secretions, identifying hundreds of different prostate-derived glycoproteins (Drake et al., 2009; Drake et al., 2010; Kim et al., 2012; Principe et al., 2012). Development and evaluation of extensive targeted proteomic assays to these proteins in EPSu are in progress for use in prostate cancer diagnosis (Kim et al., 2016; Otto et al., 2020). The prostatic fluids, EPSu and urine are also rich in extracellular vesicles (EV), which are a source for many ongoing non-coding RNA and related oligonucleotide-targeted diagnostic assays for prostate cancer and multiple diseases (Van Gils et al., 2007; Laxman et al., 2008; Linxweiler and Junker, 2020). EV obtained from urine and EPSu continues to be a highly active area for diagnostic assay development (Wang et al., 2020; Erdbrügger et al., 2021).

The majority of proteins in EPSu, urine, and associated EVs are glycosylated (Drake et al., 2009; Drake et al., 2010; Kim et al., 2012; Principe et al., 2012), either on asparagine residues, termed N-linked glycosylation, or on serine or threonine residues, termed O-glycosylation. Changes in glycosylation have been well documented in prostate cancer associated tissues, cells and biofluids (Drake et al., 2015; Scott and Munkley, 2019; Tkac et al., 2019). Extensive glycoproteomic approaches, i.e., characterization of the glycan structures at the peptide sites of modification, have been reported for urine and EPSu glycoprotein targets (Leymarie et al., 2013; Saraswat et al., 2015; Brown et al., 2020). One of the most highly characterized glycoproteins is PSA (White et al., 2009; Leymarie et al., 2013; Nyalwidhe et al., 2013; Jia et al., 2017; Kammeijer et al., 2018; Hatakeyama et al., 2021), due to its known role in prostate cancer diagnosis and relatively simple glycosylation pattern of having a single N-linked glycosylation site. Although changes in glycosylation of PSA and many other prostatic glycoproteins have diagnostic potential, assays to efficiently characterize N-glycans in urine can be lengthy and require multiple processing steps, precluding large scale clinical utility (Song et al., 2019; Hanzawa et al., 2021). Lectin arrays have been effectively used to profile glycan motifs in large cohorts of clinical urine samples (Kawakita et al., 2021; Mise et al., 2021), however this approach cannot determine full glycan compositions or distinguish N-linked or O-linked origins. Based on an adaptation of a recently published workflow for rapid characterization of serum and plasma N-glycans (Blaschke et al., 2020), we report herein a more efficient slide-based approach combined with MALDI imaging mass spectrometry (IMS) workflows to detect total N-glycan profiles of urine, EPSu and prostatic fluid samples. Comparative results with N-glycans

detected and histologically mapped in prostate cancer tissues by MALDI IMS are also included.

MATERIALS AND METHODS

Materials

Amicon Ultra 10 k centrifugal filters were obtained from Merk Millipore (Carrigtwohill, IRL). Hydrogel coated slides (Nexterion® Slide H) were obtained from Applied Microarrays (Tempe, AZ). The rotary tool was a Dremel 200 series. The well slide module (ProPlate Multi-Array Slide System, 64-well) was obtained from Grace Bio-Laboratories (Bend, OR). Sodium bicarbonate, trifluoroacetic acid (TFA), and α-cyano-4-hydroxycinnamic acid (CHCA) were obtained from Sigma-Aldrich (St. Louis, MO). HPLC grade water, 1X phosphate buffered saline (PBS), acetonitrile, citraconic anhydride, glacial acetic acid, methanol, xylene, and chloroform were obtained from Fisher Scientific (Hampton, NH). Ethanol was obtained from Decon Labs (King of Prussia, PA). Peptide-N-glycosidase F (PNGase F) PRIME™ was from N-Zyme Scientifics (Doylestown, PA). H&E stains were obtained from Cancer Diagnostics (Durham, NC).

Expressed Prostatic Secretion Urine Samples, Extracellular Vesicles, and Tissue

All samples were collected from patients and utilized after informed consent following Institutional Review Board-approved protocols at Urology of Virginia, Sentara Medical School, and the Eastern Virginia Medical School. All personal information or identifiers beyond diagnosis and lab results were not available to the laboratory investigators. EPS-urine samples were collected performing a gentle massage of the prostate gland during DRE prior to biopsy, as previously described (Drake et al., 2009). The massage consisted of three strokes on each side of the median sulcus of the prostate and the expressed fluid from the glandular network of the prostate was subsequently voided in urine. Pools (25–50 ml/sample) of EPSu were derived from 10 patients classified as having high grade, Gleason 8–10 tumors and 10 patients with low grade, Gleason 6, organ-confined prostate cancer as described previously (Nyalwidhe et al., 2013). For isolation of EPS-derived extracellular vesicles (EPSev), the two EPSu pools (45 ml) were centrifuged at $25,000 \times g$ for 30 min, and the supernatant centrifuged at $100,000 \times g$ for 4 h. The pelleted exosomes were washed twice with PBS and resuspended in 0.5 ml PBS, as previously described (Nyalwidhe et al., 2013). Direct EPS fluids (EPSd) were obtained under anesthesia prior to prostatectomy as previously described (Drake et al., 2010). A subset of 10 pairs of patient samples who provided both EPSu and EPSd were selected. Prior to glycomic analysis, 0.125 ml aliquots of each EPSu and EPSd sample were concentrated in a 10,000 MW filter cut-off 0.5 ml Amicon tube by centrifugation at 11,000 rpm in a Sorvall Legend Micro 21 benchtop microcentrifuge for 25 min. To each filtration tube was added 0.125 ml of 1X PBS, and centrifugation was repeated for 25 min. The remaining concentrated fluid, approximately

15–20 microliters, was removed to a separate vial. Each tube was rinsed with 20 microliters of PBS, and added to the concentrated sample vial (final volume 35–40 microliters). A de-identified prostate tumor tissue pair of Gleason grade 8 (4 + 4)/stage pT3b and patient-matched distal non-tumor tissue was obtained from the Hollings Cancer Center Tissue and Analysis Biorepository at the Medical University of South Carolina. A serial section of each tissue was H&E-stained according to a standardized protocol.

Control Urine Standards

Commercial urine samples representing pooled samples from four healthy males and four females were purchased from Lee BioSolutions (Maryland Heights, MO). Prior to glycomic analysis, the control urine samples were filtered and rinsed as described for the EPS samples, except 4 ml starting volume was used with larger Amicon tubes.

Expressed Prostatic Secretions Fluids and Urine Preparation for MALDI-IMS

The sample preparation and analysis of the EPS fluids (EPSu, EPSd, and EPSev) and urine samples were adapted from a workflow established for the glycomic analysis of serum and plasma (Blaschke et al., 2020). After a 30-min temperature equilibration in a moisture resistant pouch, an amine-reactive hydrogel coated slide was ground down with a rotary tool until if could fit into a Bruker MTP Slide Adapter II. A 64 well module was attached and outlined on to the back of the slide. Then the well module was unattached. Two microliters of sodium bicarbonate (100 mM, pH 8.0) was mixed with 1 microliter of the sample and briefly mixed. Within the outline of a well, 1 microliter was spotted onto the slide. EPSu, EPSd, and EPSev samples were spotted in technical triplicates, and the control urine samples were spotted in technical quadruplicates. The slide was placed in a humidity chamber, made from a culture dish with a Wypall × 60 paper towel lining the bottom and two rolled KimWipes saturated with distilled water on opposite sides, for 1 h on the benchtop to immobilize the samples to the slide. The slide was then dried in a desiccator for 15 min. The well module was reattached to the slide, matching the wells with the outlines drawn on previously. The samples were washed with Carnoy's solution (10% glacial acetic acid, 30% chloroform, and 60% 200 proof ethanol) three times for 3 min each, and subsequently washed with HPLC-grade water once for 1 min. For the washing and rinsing steps, 50 microliters of solution was added to each well and dumped out of the well by inverting the slide. Following the water wash, the slide was dried in a desiccator for 30 min with the slide module attached. After detaching the slide module, a M5 TM-Sprayer (HTX Technologies) was used to spray a 0.1 mg/ml PNGase F PRIME solution in water on to the slide for 15 passes at 25 microliters/min, 1,200 mm/min, 45°C, and 3 mm spacing between passes with 10 psi nitrogen gas. The slide was then incubated in a preheated humidity chamber at 37°C for 2 h. A M5 TM-Sprayer was also used to apply the MALDI matrix solution (7 mg/ml of CHCA in 50% acetonitrile/0.1% TFA) on the slide for 10 passes at 100 microliters/min,

1,200 mm/min, 79°C, and 2.5 mm spacing between passes with 10 psi nitrogen gas.

Prostate Tissue Preparation for MALDI-IMS

The tissues were prepared as described previously (Drake et al., 2018a). Briefly, the tissues were dewaxed by 1 h in 60°C and xylene washes, rehydrated with a gradation of ethanol and water washes, and underwent antigen retrieval in citraconic anhydride buffer (25-µL citraconic anhydride, 2-µL 12 M HCl, 50-ml HPLC-grade water, pH 3.0 ± 0.5) in a decloaking chamber at 95°C for 30 min. A M5 TM-Sprayer (HTX Technologies) was used to spray a 0.1 mg/ml PNGase F PRIME solution in water on to the slide for 15 passes at 25 microliters/min, 1,200 mm/min, 45°C, and 3 mm spacing between passes with 10 psi nitrogen gas. The slide was then incubated in a preheated humidity chamber at 37°C for 2 h. A M5 TM-Sprayer was also used to apply the MALDI matrix solution (7 mg/ml of CHCA in 50% acetonitrile/ 0.1% TFA) on the slide for 10 passes at 100 microliters/min, 1,200 mm/min, 79°C, and 2.5 mm spacing between passes with 10 psi nitrogen gas.

MALDI Imaging Mass Spectrometry

A dual source timsTOF fleX MALDI-QTOF mass spectrometer (Bruker) was used to image the slides as previously described (McDowell et al., 2021). Images were collected with a SmartBeam 3D laser operating at 10,000 Hz with a 20 µm laser spot size at a 150 µm raster with 300 laser shots per pixel. Samples were analyzed in positive ion mode spanning a m/z range of 700–4,000.

Data Processing and Analysis

Mass spectra were imported in to SCiLS Lab software 2021a (Bruker), normalized to total ion current, and manually peak selected for N-glycans based on theoretical mass values. SCiLS was also used for individual peak visualization and quantification. Maximum mean values for each peak were exported for each sample region. Each N-glycan measurement for each sample was subtracted by the background signal in the blank well to find the absolute intensity. To account for differences in protein concentrations that could lead to higher signal intensities and detection of more low-abundance N-glycan species, N-glycan relative intensities were calculated as the absolute intensity divided by the sum of all the absolute intensities of the N-glycans found in each of the samples being compared. Comparisons of the number of N-glycans detected in each sample is also discussed, and the presence/absence of a N-glycan in each sample is noted in **Supplementary Table S1**. N-glycan structures were labelled with a N-glycan class or classes depending on their putative structures. Quantifications of the N-glycan classes were calculated by summing the relative intensities of the individual N-glycans belonging to each class. N-glycan profiles were also examined by grouping each N-glycan into a group depending on the presence and/or absence of mannose, fucose, sialic acid, and sulphate and comparing the summed relative intensities of the classes. When comparing individual N-glycan intensities across samples, the multiply sodiated species of sialylated and sulfated N-glycans were added together.
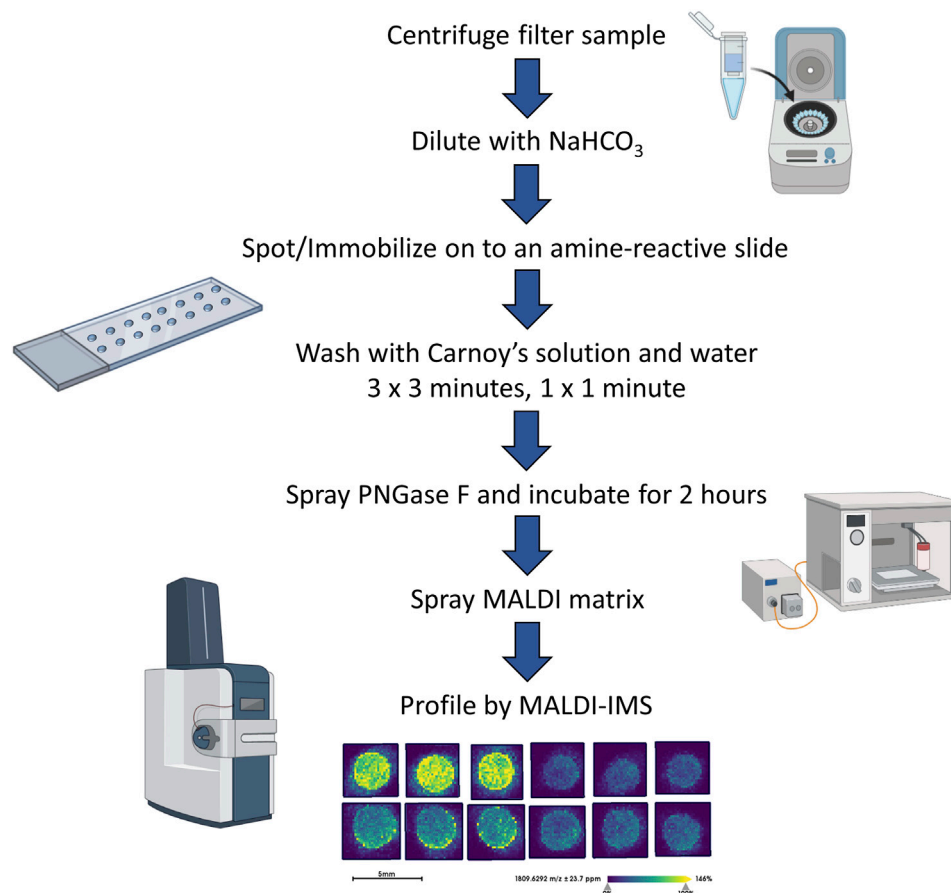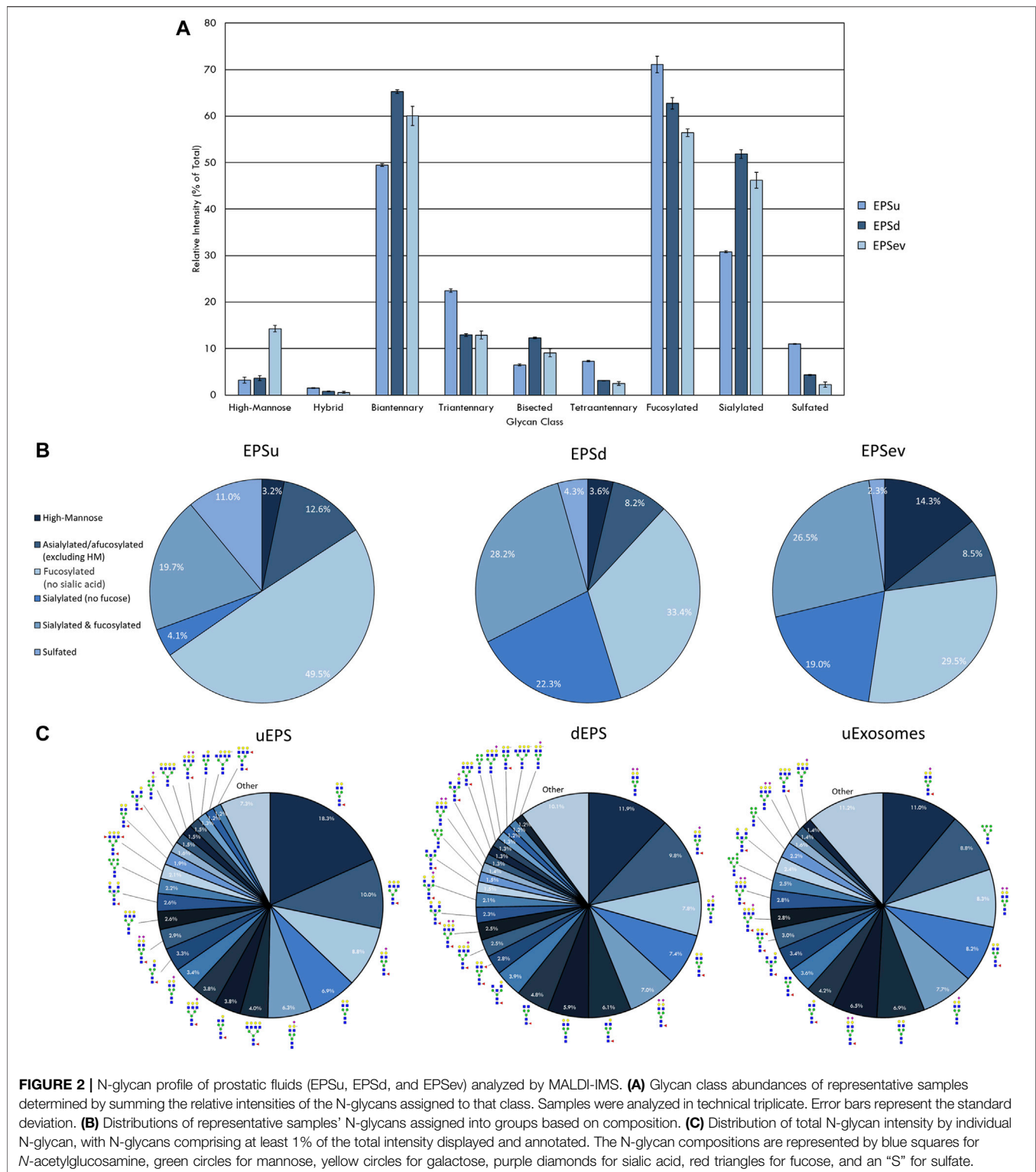
FIGURE 1 | Workflow of the EPS fluid and urine N-glycan analysis.

RESULTS

To complement previous proteomic studies of proximal prostatic fluids obtained in the urology clinic as related to prostate cancers (Drake et al., 2009; Drake et al., 2010; Kim et al., 2012; Principe et al., 2012) a series of different EPSu, EPSd and EPSev samples were used to develop a MALDI-based N-glycan profiling method. The goal was to have a workflow that required minimal sample processing and could be completed in a 6–8 h timeline, in contrast to current glycomic analysis workflows for urine and prostatic fluids that require multiple processing, derivatization and purification steps prior to analysis. A previous slide-based approach used for serum and plasma N-glycan profiling (Blaschke et al., 2020) was the starting point, and a workflow summarized in **Figure 1** was developed for EPSu. A key feature is use of an amine reactive slide chemistry that covalently binds target glycoproteins, facilitating washing steps to remove lipid and salts prior to spraying of a molecular coating of PNGase F PRIME to release N-glycans. An additional concentration and buffer exchange step was added for EPSu and EPSd, using a 10,000 MW cut-off spin cartridge to concentrate and allow buffer exchange of the sample prior to addition to the amine-reactive slide. This step results in a three to four fold increase in

concentration of glycoproteins in the biofluid. An SDS-polyacrylamide gel image showing protein loading examples for the EPSu, EPSd and EPSev samples are provided in **Supplementary Figure S1**.
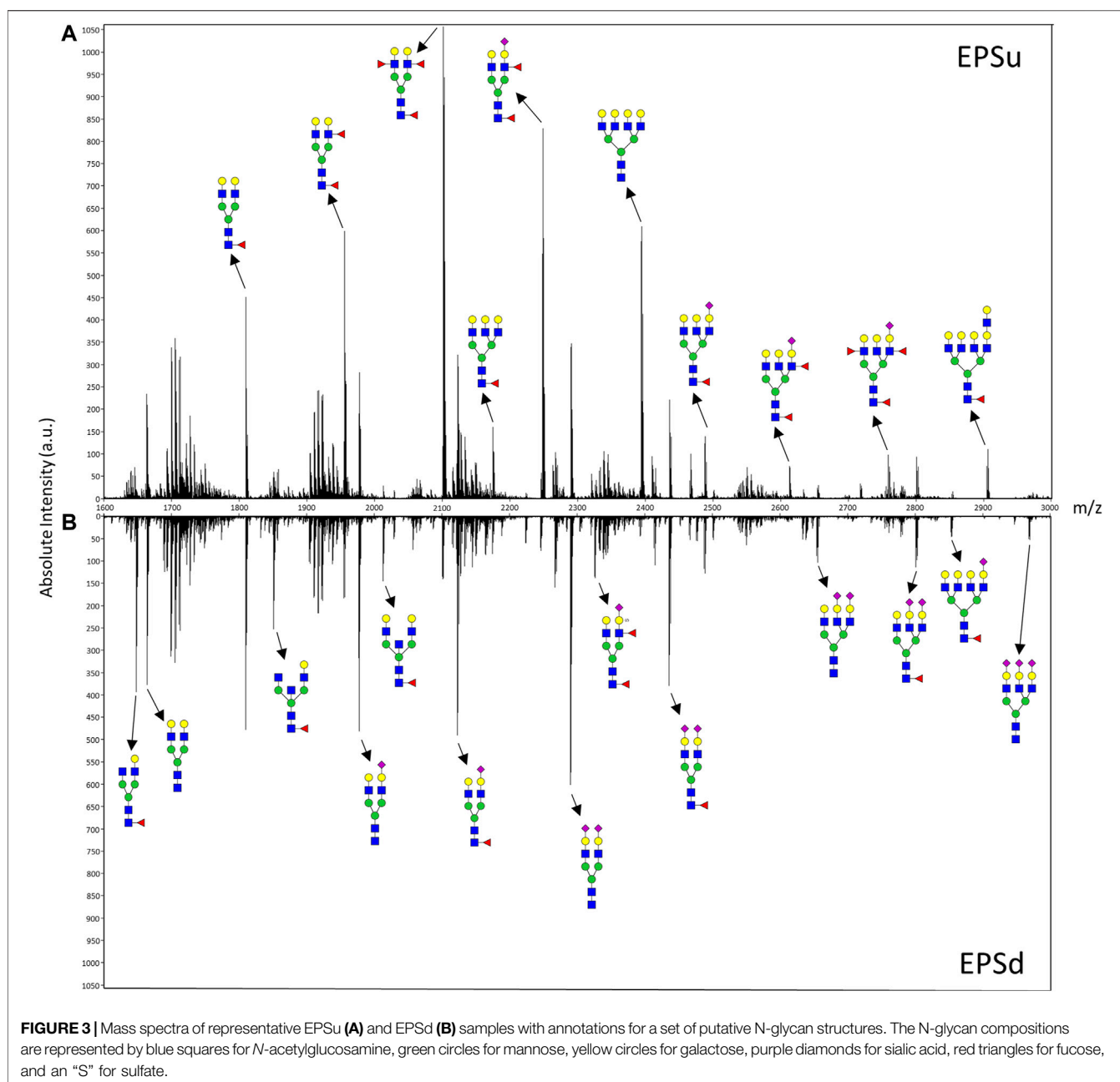
Initial experiments focused on optimizing detection of N-glycans in EPSu and EPSd sample pairs obtained from the same donors, as well as existing EPSev pool samples. Similar to what was previously determined for serum or plasma preparation on the amine-reactive hydrogel slides, the key for optimal N-glycan detection was inclusion of the Carnoy's solution wash after spotting, which serves to remove lipids and salts, as well as denature the bound glycoproteins facilitating access for PNGase F. Thus far, a total of 35 EPSu, 10 EPSd and 8 EPSev samples have been analyzed with the workflow shown in **Figure 1**. Cumulatively, the resulting N-glycans detected in each sample type are summarized in **Supplementary Table S1**, and structural class groupings are shown in **Figure 2**. Broadly, the EPSu samples had the most N-glycan species detected ($n = 182$) versus EPSd ($n = 135$). These numbers include multiple versions of the same N-glycan compositions for sialylated and sulfates species, which can vary in mass due to varying numbers of sodium ions associating with the charged groups. These glycoforms were included in

FIGURE 2 | N-glycan profile of prostatic fluids (EPSu, EPSd, and EPSev) analyzed by MALDI-IMS. **(A)** Glycan class abundances of representative samples determined by summing the relative intensities of the N-glycans assigned to that class. Samples were analyzed in technical triplicate. Error bars represent the standard deviation. **(B)** Distributions of representative samples' N-glycans assigned into groups based on composition. **(C)** Distribution of total N-glycan intensity by individual N-glycan, with N-glycans comprising at least 1% of the total intensity displayed and annotated. The N-glycan compositions are represented by blue squares for *N*-acetylglucosamine, green circles for mannose, yellow circles for galactose, purple diamonds for sialic acid, red triangles for fucose, and an "S" for sulfate.

Supplementary Table S1, but were generally detected at lower intensity values.

There was a range of protein concentrations across the samples in each sample type examined. While this created differences in the total intensity of the N-glycan profile and

number of N-glycans that could be detected, these differences were accounted for by only comparing relative intensities, i.e. an individual N-glycan's intensity relative to the total intensity of the N-glycans in that sample that were also seen in all sample types being compared. Representative samples with the most N-glycans

FIGURE 3 | Mass spectra of representative EPSu **(A)** and EPSd **(B)** samples with annotations for a set of putative N-glycan structures. The N-glycan compositions are represented by blue squares for *N*-acetylglucosamine, green circles for mannose, yellow circles for galactose, purple diamonds for sialic acid, red triangles for fucose, and an "S" for sulfate.

detected were selected and compared from each sample type to display the breadth the N-glycan profiles.

The intensity of the N-glycan classes varied across the prostatic fluid samples (**Figure 2A**). EPSd and EPSev N-glycan classes had similar intensities, except for the higher amount of high-mannose N-glycans in EPSev. For all samples, the majority of the N-glycans were biantennary and/or fucosylated. EPSu had approximately 15 and 20% less sialylation than the EPSd and EPSev, respectively, but had an increased amount of tetraantennary and sulfated N-glycans. Many of these findings were replicated when grouping the N-glycans detected in each sample based on composition (**Figure 2B**). About half of the N-glycans in EPSu were fucosylated with no sialic acid compared

to approximately 30% in the EPSd and EPSev samples. The EPSd and EPSev samples had higher levels of N-glycans with sialic acids and no fucose. When examining the intensity of individual N-glycans in the prostatic fluid samples, the most abundant N-glycans were typically biantennary with two galactoses (**Figure 2C**). In concordance with the N-glycan class comparison, EPSev had more high abundance high-mannose N-glycans than the other samples, and m/z 1419.4755 (Hex6HexNAc2 + 1Na) was the second most abundant N-glycan. The sulfated N-glycan m/z 2056.6156 (Hex5HexNAc4NeuAc1 + 1SO4 + 2Na) had a relative intensity of 6.3% in the EPSu, compared to 1.2% in EPSd and less than 1% in EPSev.
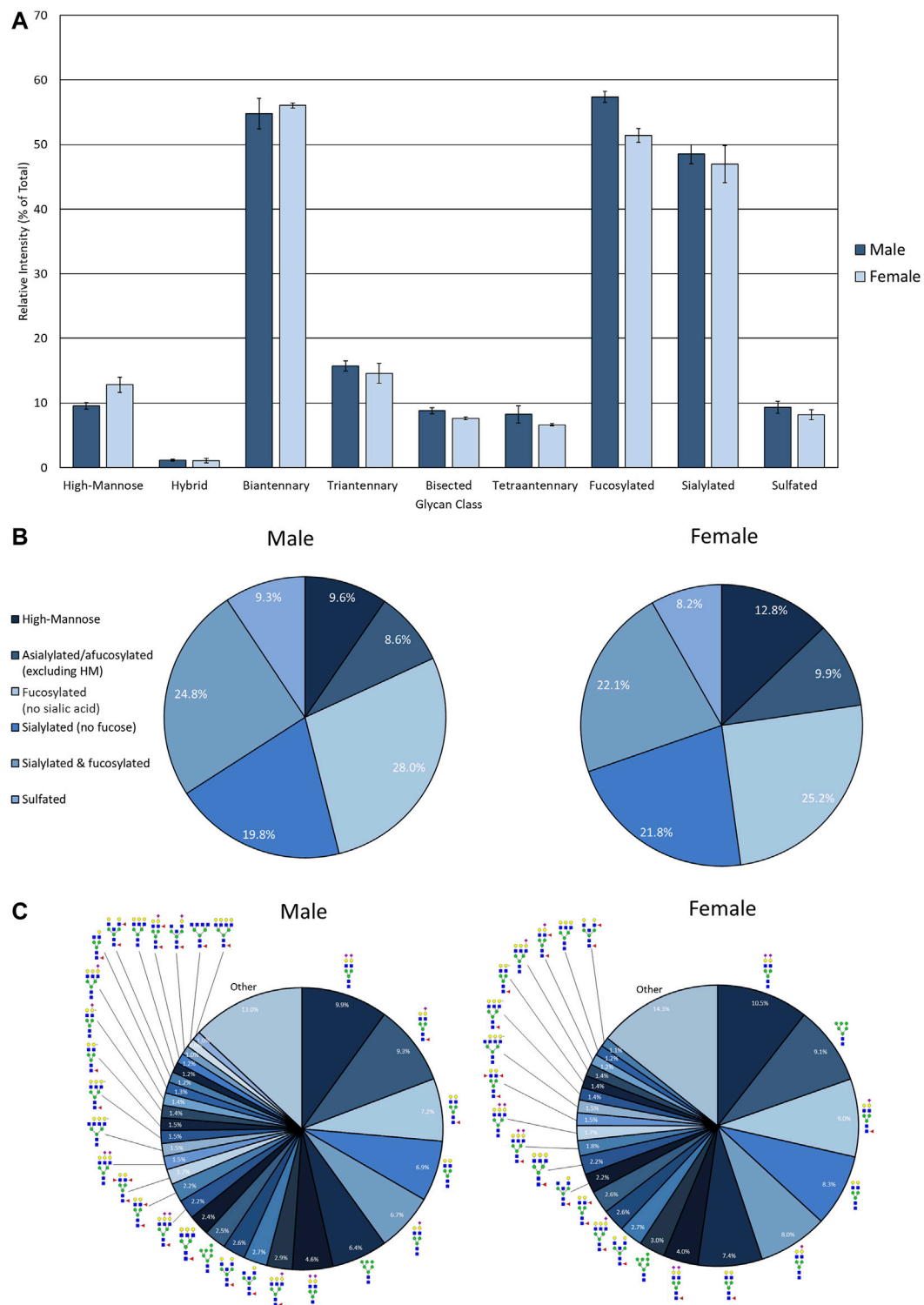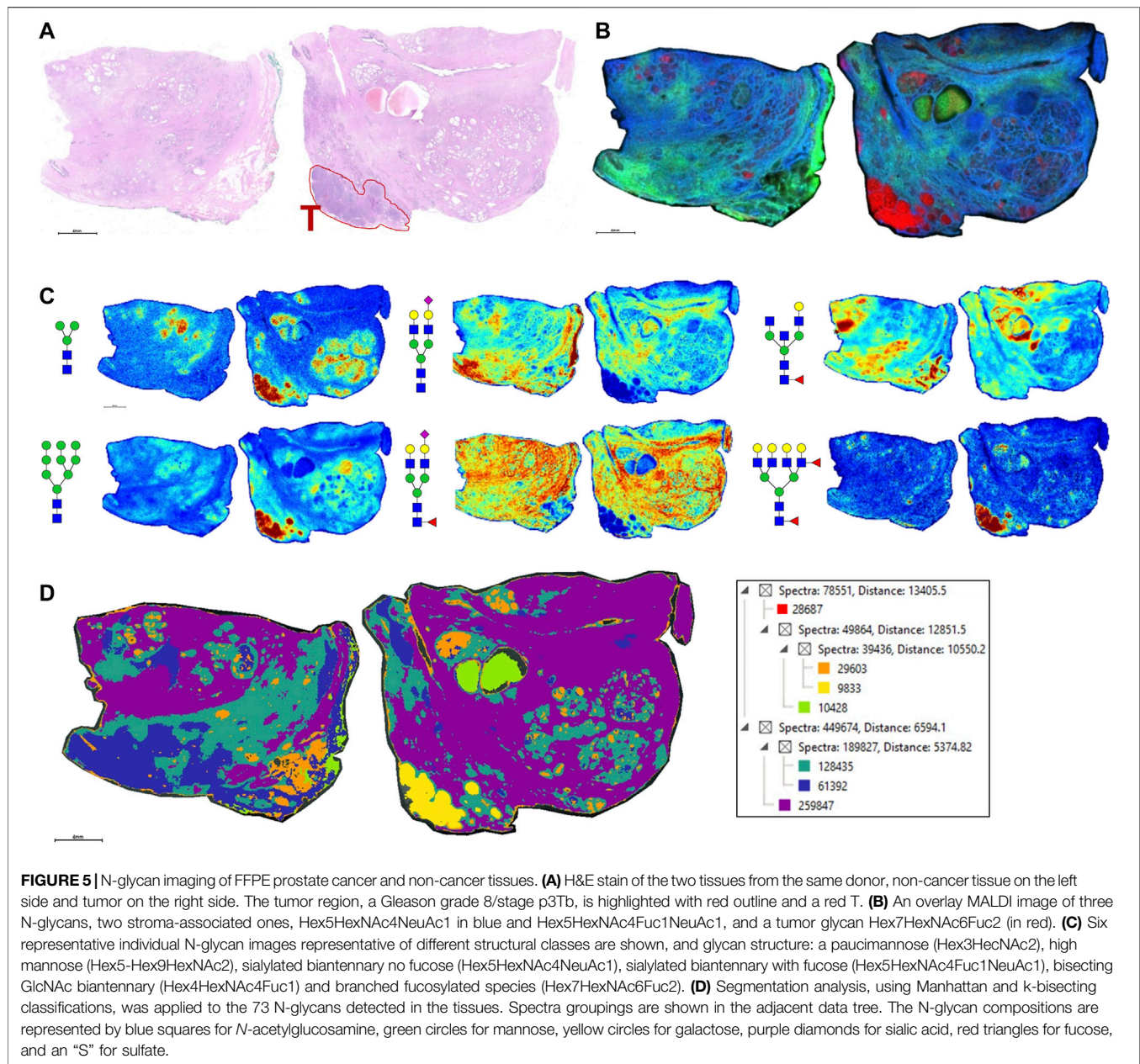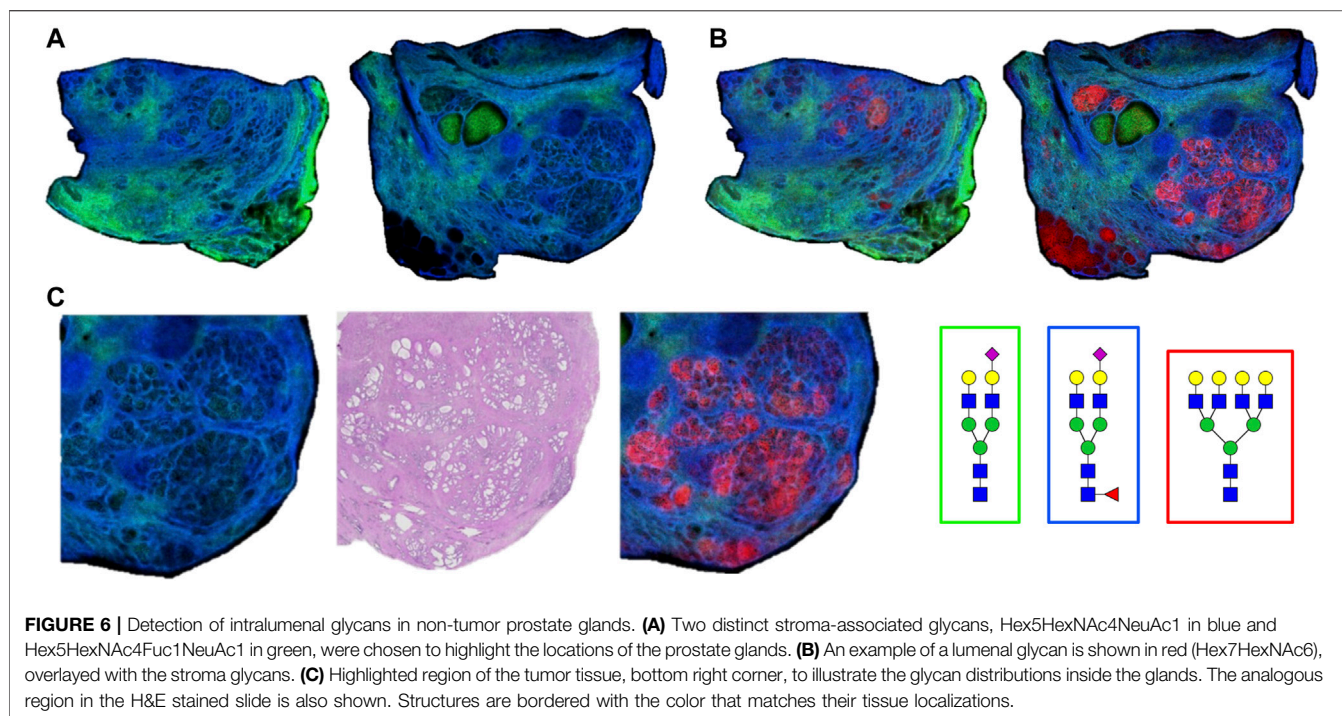
FIGURE 4 | N-glycan profile of pooled healthy male and female urine analyzed my MALDI-IMS. **(A)** Glycan class abundances of representative samples determined by summing the relative intensities of the N-glycans assigned to that class. Sample were analyzed in technical quadruplicate. Error bars represent the standard deviation. **(B)** Distributions of representative samples' N-glycans assigned into groups based on composition. **(C)** Distribution of total glycan intensity by individual N-glycan, with N-glycans comprising at least 1% of the total intensity displayed and annotated. The N-glycan compositions are represented by blue squares for N-acetylglucosamine, green circles for mannose, yellow circles for galactose, purple diamonds for sialic acid, red triangles for fucose, and an "S" for sulfate.

FIGURE 5 | N-glycan imaging of FFPE prostate cancer and non-cancer tissues. **(A)** H&E stain of the two tissues from the same donor, non-cancer tissue on the left side and tumor on the right side. The tumor region, a Gleason grade 8/stage p3Tb, is highlighted with red outline and a red T. **(B)** An overlay MALDI image of three N-glycans, two stroma-associated ones, Hex5HexNAc4NeuAc1 in blue and Hex5HexNAc4Fuc1NeuAc1, and a tumor glycan Hex7HexNAc6Fuc2 (in red). **(C)** Six representative individual N-glycan images representative of different structural classes are shown, and glycan structure: a paucimannose (Hex3HecNAc2), high mannose (Hex5-Hex9HexNAc2), sialylated biantennary no fucose (Hex5HexNAc4NeuAc1), sialylated biantennary with fucose (Hex5HexNAc4Fuc1NeuAc1), bisecting GlcNAc biantennary (Hex4HexNAc4Fuc1) and branched fucosylated species (Hex7HexNAc6Fuc2). **(D)** Segmentation analysis, using Manhattan and k-bisecting classifications, was applied to the 73 N-glycans detected in the tissues. Spectra groupings are shown in the adjacent data tree. The N-glycan compositions are represented by blue squares for *N*-acetylglucosamine, green circles for mannose, yellow circles for galactose, purple diamonds for sialic acid, red triangles for fucose, and an "S" for sulfate.

In order to compare mass spectra and absolute intensity values of N-glycan peaks, an EPSu and EPSd sample with similar protein levels from the same cohort were prepped together and imaged in the same run (**Figure 3**). Substantial differences were seen in most of the high-intensity N-glycan peaks, except for m/z 1809.6393 (Hex5dHex1HexNAc4 + 1Na), m/z 2465.8669 (Hex6dHex1HexNAc5NeuAc1 + 1Na), and m/z 2800.9263 (Hex6dHex1HexNAc5NeuAc2 + 3Na). The increased levels of sialylated N-glycans in the EPSd sample and increased fucosylated species in the EPSu sample is also evident.

The advantage of developing the assay using EPSu samples are the inherently higher protein concentrations present from the prostatic fluid mixture in these samples relative to normal urine. Therefore, for comparison, control urine samples from pools of four healthy male and four healthy female donors were processed and analyzed using this workflow. This required 4 ml of starting fluid for concentration and desalting using larger filtration tubes, but was otherwise the same workflow as described for EPSu/d samples. Overall, N-glycan class intensities were similar for both samples (**Figure 4A**). There were high levels of biantennary, fucosylated, and sialylated N-glycans and low levels of hybrid N-glycans. The rest of the N-glycan classes had approximately 10% intensity. The biggest difference between the male and female samples was the slightly lower level of fucosylation in the female sample. The distribution of N-glycans based on composition displayed very similar profiles (**Figure 4B**). The biggest difference was a 3.2% increase in high-mannose N-glycans in the female sample. Accordingly, m/z 1419.4755

FIGURE 6 | Detection of intralumenal glycans in non-tumor prostate glands. **(A)** Two distinct stroma-associated glycans, Hex5HexNAc4NeuAc1 in blue and Hex5HexNAc4Fuc1NeuAc1 in green, were chosen to highlight the locations of the prostate glands. **(B)** An example of a lumenal glycan is shown in red (Hex7HexNAc6), overlayed with the stroma glycans. **(C)** Highlighted region of the tumor tissue, bottom right corner, to illustrate the glycan distributions inside the glands. The analogous region in the H&E stained slide is also shown. Structures are bordered with the color that matches their tissue localizations.

(Hex6HexNAc2 + 1Na) was the sixth most abundant N-glycan in males and the second most abundant in females (**Figure 4C**). Similar to the EPSu samples, the most abundant N-glycans are primarily biantennary with two galactoses.

As the glycoprotein constituents of the EPSu and EPSd samples are secreted by prostate glands, a comparative N-glycan comparison of prostate tissues was done using previously reported N-glycan MALDI IMS approaches (Drake et al., 2017, 2018b; McDowell et al., 2021). In the example shown in **Figure 5A**, a pair of non-tumor and tumor tissues from the same donor were evaluated. Use of the non-tumor tissue from the same donor is done to demonstrate the specificity of tumor specific N-glycans detected in the tumor containing tissue [a Gleason grade 8 (4 + 4), stage pT3b]. The tumor region is quite distinct, localized to the bottom left corner of the tissue. Additional H&E images of both tissues with increasing magnification are provided in **Supplementary Figure S2**. In **Figures 5B,C** are shown representative N-glycan classes and their tissue distributions, from a total of 73 N-glycans detected. These structures include tumor-associated paucimannose (Hex3HecNAc2), high mannose (Hex5-Hex9HexNAc2), and branched fucosylated species (Hex7HexNAc6Fuc2) (**Figure 5C**), consistent with previous reports (Drake et al., 2018b). Two of the most abundant sialylated biantennary N-glycans detected are stroma-associated, as shown in the overlay image with a tumor N-glycan (in red), Hex5HexNAc4NeuAc1 in blue and Hex5HexNAc4Fuc1NeuAc1 in green (**Figure 5B**, and individually in **Figure 5C**). A segmentation analysis of the 73 N-glycans is shown in **Figure 5D**, illustrating how different N-glycan classes are associated with different histopathology features. This representative tissue was selected for another

feature, as it was noted that there was a distinct intra-lumen glandular N-glycan signature that could be detected. As shown in **Figure 6A**, the two sialylated biantennary N-glycans (Hex5HexNAc4NeuAc1 in blue and Hex5HexNAc4Fuc1NeuAc1 in green) provide a stromal scaffold image. This was used to detect which N-glycan species were present in the lumen regions, as illustrated in **Figure 6B** for Hex7HexNAc6 in red, and in **Figure 6C** for a highlighted gland region. Doing this, 38 N-glycans were detected in the lumen of glands. Many of these same N-glycans are also tumor associated, but the histopathology differences are significant between tumor and lumen.

Using the lists of 73 tissue N-glycans and the subset of 38 lumen N-glycans, these were compared to the N-glycans detected in EPSu and EPSd samples. As shown in the Venn diagram in Figure 7, 44 N-glycans were present in the tissues and both EPS samples (**Figure 7A**), and 18 N-glycans were shared in the lumen and EPS samples (**Figure 7B**). Structurally, these shared N-glycans were the high mannose and the most abundant branched fucosylated N-glycans.

DISCUSSION

The goal of our study was to develop and optimize a more efficient and rapid method to evaluate N-linked glycosylation by MALDI-MS in urine and EPSu samples. The workflow is similar to what was developed for serum and plasma (Blaschke et al., 2020), utilizing an amine reactive slide to capture glycoproteins on a solid surface. An additional 1 h step of sample processing by molecular weight filtration and centrifugation for urine was included in order to concentrate

FIGURE 7 | Comparison of detected N-glycans across sample sets. **(A)** The number of unique and common N-glycans detected in EPSu, EPSd, and across the entire prostate tissues. **(B)** The number of unique and common N-glycans detected in EPSu, EPSd, and within the lumen of the prostate tissues.
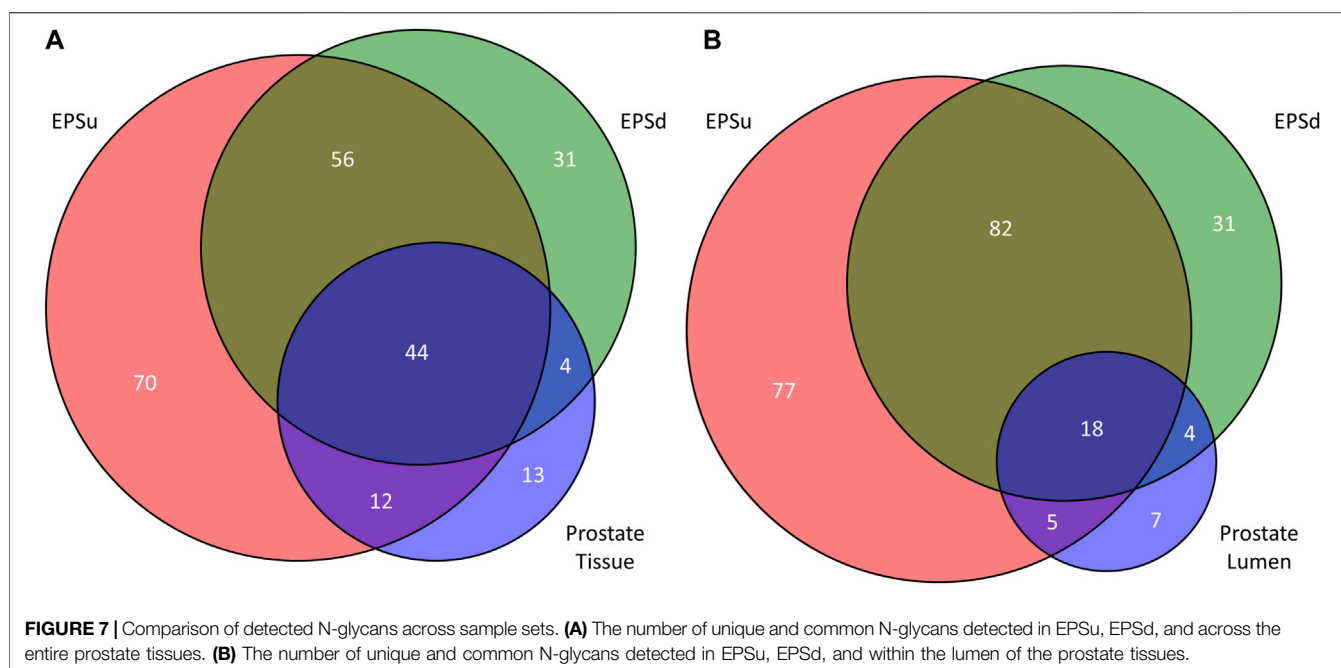
and solvent exchange. Spotting only 1 ul, the urine glycoprotein spots are sprayed with a molecular coating of PNGase F, followed by analysis on a MALDI-FTICR or MALDI-QTOF mass spectrometer. The method is applicable to urine, EPSu and EPSd samples. An EPSev prep was spotted directly onto the amine reactive slides without any pre-processing. Routinely, 100 or more N-glycans can be detected depending on the sample type.

The N-glycan profiling of the prostatic fluid samples displayed more high-mannose N-glycans in the EPSev compared to the EPSu and EPSd, reflected in the intensities of the N-glycan classes and the most abundant individual N-glycans (**Figure 1**). Relatively high intensities of high-mannose N-glycans have been noted in non-EPS urinary exosomes as well (Zou et al., 2017). The similarity of the EPSd and EPSev N-glycan profiles suggests that a significant portion of the EPS glycoproteins may be within exosomes. In conjunction with these findings, direct comparison of mass spectra from EPSu and EPSd with similar protein levels displayed distinct N-glycan profiles for EPSu and EPSd (**Figure 2**). While male and female urine N-glycan profiles and individual N-glycan abundances displayed few differences, there was a slightly higher amount of fucosylation in the male sample. This trend has previously been identified in male and female plasma samples (Dotz and Wuhrer, 2019). Similar levels of sulfation were detected in the urine controls and EPSu, which was higher than the EPSd and EPSev, indicating that the sulfated N-glycans in EPSu may be originating from urine glycoproteins.

A comprehensive MS analysis of the N-glycans present in adult and pediatric urine samples has been published (Li et al., 2020). In this study, PNGase F released N-glycans were isolated, labeled with aminobenzoic acid, and methylamidated prior to analysis by LC-MS/MS. The authors reported 116 N-glycan compositions could be detected, and a subset of 46 N-glycans that were reproducibly detected and further quantified (Li et al.,

2020). This study provides an optimal benchmark for our current study to evaluate how many of the 46 N-glycans could be detected by our solid-phase and single day MALDI MS workflow. Comparing the MALDI data from the male and female urine standards, the majority of N-glycans are detected, especially the biantennary, high mannose and fucosylated species. The MALDI-based assay is not as effective for detection of larger tri- and tetra-sialylated N-glycans, and for N-glycans with molecular masses above 3,700 m/z.

One observation noted from our data was that the intensity levels of N-glycans detected were variable, even when protein concentrations were comparable. There was also not a large variation in the types of N-glycan structures detected in the samples, especially for the most abundant species. These variations could reflect significantly different protein levels of individual glycoprotein species, or variations in the amount of the most abundant protein in human urine, uromodulin. Uromodulin, also called Tamm–Horsfall protein, has eight N-glycosylation sites and has been comprehensively studied for N-glycosylation content and composition (Li et al., 2021). Interestingly in this study, N-glycan analysis comparisons of a uromodulin depleted urine versus non-depleted sample indicated little difference in the N-glycan compositions detected in both samples (Li et al., 2021). Our study only represents a few sample numbers, so no conclusions can be reached until statistically relevant sample numbers are analyzed. The rapidity and efficiency of the slide-based MALDI assay will facilitate evaluation of larger cohorts. Possibly pairing the N-glycan data with the quantitative MRM proteomic assay developed for EPSu proteins could address both the changes in N-glycan levels and protein concentrations in the same sample (Kim et al., 2016; Otto et al., 2020). It also does not preclude direct targeting of specific glycoproteins present for N-glycan content analysis, as

is done so often for PSA. We have already reported a solid-phase antibody array approach to N-glycan profile individual serum glycoproteins captured by their specific antibodies (Black et al., 2019a; 2019b). A similar approach targeting the abundant urine and EV glycoproteins like PSA is in progress.

The N-glycans obtained from a representative FFPE prostate cancer tissue were also included to illustrate the N-glycans that can be detected in the lumen of the prostate glands, representing prostatic fluid still present during tissue fixation. Based on hundreds of prostate cancer tissues analyzed by N-glycan IMS, to be reported separately, detection of luminal N-glycans in these tissues is highly variable. Fluid remnants can be seen in the H&E stains, but not all gland lumens contain this, and can be completely absent in many tissues. Presumably this variation reflects how much fluid was present at the time of prostatectomy, but could also reflect differences in FFPE tissue preparation and processing. While this is a variable that may lack clinical diagnostic significance, the selected tissue in **Figure 5** highlights the tumor-associated and secreted N-glycans ($n = 38$) in tissue that can potentially be detected in the EPSu, EPSd and EPSev samples. These are primarily paucimannose, high mannose and branched multi-fucosylated N-glycans.

In summary, we present an initial N-glycan profiling workflow applicable to urine and prostatic fluids. It also represents a scalable framework, depending on the application and intent of the assay. There is already demonstrable sensitivity for detecting N-glycans efficiently in low concentrations of starting material, hence analysis of larger clinical cohorts is feasible. If time of preparation is less of a concern, an increased amount of starting fluid can be concentrated to increase protein concentrations prior to spotting. Additionally, the use of the concentration and solvent/buffer exchange step in the workflow can be adapted to other biofluids with protein concentrations similar to urine, i.e., saliva, cerebrospinal fluid, bronchial lavage. The method can also be used with other enzymes besides PNGase F, like endoglycosidase F3 specific to core fucoses (West et al., 2020), or other glycosidases specific to other N-glycan structural classes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

REFERENCES

Black, A. P., Angel, P. M., Drake, R. R., and Mehta, A. S. (2019a). Antibody Panel Based N -Glycan Imaging for N -Glycoprotein Biomarker Discovery. *Curr. Protoc. Protein Sci.* 98, e99. doi:10.1002/cpps.99

Black, A. P., Liang, H., West, C. A., Wang, M., Herrera, H. P., Haab, B. B., et al. (2019b). A Novel Mass Spectrometry Platform for Multiplexed N-Glycoprotein Biomarker Discovery from Patient Biofluids by Antibody Panel Based N-Glycan Imaging. *Anal. Chem.* 91, 8429–8435. doi:10.1021/acs.analchem.9b01445

Blaschke, C. R. K., Black, A. P., Mehta, A. S., Angel, P. M., and Drake, R. R. (2020). Rapid N-Glycan Profiling of Serum and Plasma by a Novel Slide-Based Imaging Mass Spectrometry Workflow. *J. Am. Soc. Mass. Spectrom.* 31, 2511–2520. doi:10.1021/jasms.0c00213

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

CB, JH, GG, LL, and RD carried out experiments. CB performed MS data processing and sample analysis assisted by RD. OS, JW, JI, CH-H, and JN provided the study sample cohorts. CB and RD prepared the manuscript with contributions from all authors.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.734280/full#supplementary-material

Brown, C. J., Gaunitz, S., Wang, Z., Strindelius, L., Jacobson, S. C., Clemmer, D. E., et al. (2020). Glycoproteomic Analysis of Human Urinary Exosomes. *Anal. Chem.* 92, 14357–14365. doi:10.1021/acs.analchem.0c01952

Dotz, V., and Wuhrer, M. (2019). N -glycome Signatures in Human Plasma: Associations with Physiology and Major Diseases. *FEBS Lett.* 593, 2966–2976. doi:10.1002/1873-3468.13598

Drake, R. R., Elschenbroich, S., Lopez-Perez, O., Kim, Y., Ignatchenko, V., Ignatchenko, A., et al. (2010). In-depth Proteomic Analyses of Direct Expressed Prostatic Secretions. *J. Proteome Res.* 9, 2109–2116. doi:10.1021/pr1001498

Drake, R. R., Jones, E. E., Powers, T. W., and Nyalwidhe, J. O. (2015). "Altered Glycosylation in Prostate Cancer," in *Advances In Cancer Research* (San Diego: Academic Press), 345–382. doi:10.1016/bs.acr.2014.12.001

Drake, R. R., and Kislinger, T. (2014). The Proteomics of Prostate Cancer Exosomes. *Expert Rev. Proteomics* 11, 167–177. doi:10.1586/14789450.2014.890894

Drake, R. R., Powers, T. W., Jones, E. E., Bruner, E., Mehta, A. S., and Angel, P. M. (2017). MALDI Mass Spectrometry Imaging of N-Linked Glycans in Cancer Tissues. *Adv. Cancer Res.* (Singapore: Springer), 85–116. doi:10.1016/bs.acr.2016.11.009

Drake, R. R., Powers, T. W., Norris-Caneda, K., Mehta, A. S., and Angel, P. M. (2018a). *In Situ* Imaging of N-Glycans by MALDI Imaging Mass Spectrometry of Fresh or Formalin-Fixed Paraffin-Embedded Tissue. *Curr. Protoc. Protein Sci.* 94, e68–21. doi:10.1002/cpps.68

Drake, R. R., West, C. A., Mehta, A. S., and Angel, P. M. (2018b). "MALDI Mass Spectrometry Imaging of N-Linked Glycans in Tissues," in *Advances In Experimental Medicine And Biology* (Springer New York LLC), 59–76. doi:10.1007/978-981-13-2158-0_4

Drake, R. R., White, K. Y., Fuller, T. W., Igwe, E., Clements, M. A., Nyalwidhe, J. O., et al. (2009). Clinical Collection and Protein Properties of Expressed Prostatic Secretions as a Source for Biomarkers of Prostatic Disease. *J. Proteomics* 72, 907–917. doi:10.1016/j.jprot.2009.01.007

Erdbrügger, U., Blijdorp, C. J., Bijnsdorp, I. V., Borràs, F. E., Burger, D., Bussolati, B., et al. (2021). Urinary Extracellular Vesicles: A Position Paper by the Urine Task Force of the International Society for Extracellular Vesicles. *J. Extracellular Vesicles* 10, e12093. doi:10.1002/jev2.12093

Hanzawa, K., Tanaka-Okamoto, M., Murakami, H., Mukai, M., Takahashi, H., Omori, T., et al. (2021). Investigation of Acidic Free-Glycans in Urine and Their Alteration in Cancer. *Glycobiology* 31, 391–409. doi:10.1093/GLYCOB/CWAA100

Hatakeyama, S., Yoneyama, T., Tobisawa, Y., Yamamoto, H., and Ohyama, C. (2021). Narrative Review of Urinary Glycan Biomarkers in Prostate Cancer. *Transl. Androl. Urol.* 10, 1850–1864. doi:10.21037/tau-20-964

Jia, G., Dong, Z., Sun, C., Wen, F., Wang, H., Guo, H., et al. (2017). Alterations in Expressed Prostate Secretion-Urine PSA N-Glycosylation Discriminate Prostate Cancer from Benign Prostate Hyperplasia. *Oncotarget* 8, 76987–76999. doi:10.18632/oncotarget.20299

Kammeijer, G. S. M., Nouta, J., De La Rosette, J. J. M. C. H., De Reijke, T. M., and Wuhrer, M. (2018). An In-Depth Glycosylation Assay for Urinary Prostate-specific Antigen. *Anal. Chem.* 90, 4414–4421. doi:10.1021/acs.analchem.7b04281

Kawakita, C., Mise, K., Onishi, Y., Sugiyama, H., Yoshida, M., Yamada, M., et al. (2021). Novel Urinary Glycan Profiling by Lectin Array Serves as the Biomarkers for Predicting Renal Prognosis in Patients with IgA Nephropathy. *Sci. Rep.* 11, 1. doi:10.1038/s41598-020-77736-1

Kim, Y., Ignatchenko, V., Yao, C. Q., Kalatskaya, I., Nyalwidhe, J. O., Lance, R. S., et al. (2012). Identification of Differentially Expressed Proteins in Direct Expressed Prostatic Secretions of Men with Organ-Confined versus Extracapsular Prostate Cancer. *Mol. Cell Proteomics* 11, 1870–1884. doi:10.1074/mcp.M112.017889

Kim, Y., Jeon, J., Mejia, S., Yao, C. Q., Ignatchenko, V., Nyalwidhe, J. O., et al. (2016). Targeted Proteomics Identifies Liquid-Biopsy Signatures for Extracapsular Prostate Cancer. *Nat. Commun.* 7, 1–10. doi:10.1038/ncomms11906

Laxman, B., Morris, D. S., Yu, J., Siddiqui, J., Cao, J., Mehra, R., et al. (2008). A First-Generation Multiplex Biomarker Analysis of Urine for the Early Detection of Prostate Cancer. *Cancer Res.* 68, 645–649. doi:10.1158/0008-5472.CAN-07-3224

Leymarie, N., Griffin, P. J., Jonscher, K., Kolarich, D., Orlando, R., McComb, M., et al. (2013). Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: The ABRF Glycoprotein Research Multi-Institutional Study 2012. *Mol. Cell Proteomics* 12, 2935–2951. doi:10.1074/mcp.M113.030643

Li, H., Kostel, S. A., Dimartino, S. E., Hashemi Gheinani, A., Froehlich, J. W., and Lee, R. S. (2021). Uromodulin Isolation and its N-Glycosylation Analysis by NanoLC-MS/MS. *J. Proteome Res.* 20, 2662–2672. doi:10.1021/acs.jproteome.0c01053

Li, H., Patel, V., DiMartino, S. E., Froehlich, J. W., and Lee, R. S. (2020). An In-Depth Comparison of the Pediatric and Adult Urinary N-Glycomes. *Mol. Cell Proteomics* 19, 1767–1776. doi:10.1074/mcp.RA120.002225

Linxweiler, J., and Junker, K. (2020). Extracellular Vesicles in Urological Malignancies: an Update. *Nat. Rev. Urol.* 17, 11–27. doi:10.1038/s41585-019-0261-8

McDowell, C. T., Klamer, Z., Hall, J., West, C. A., Wisniewski, L., Powers, T. W., et al. (2021). Imaging Mass Spectrometry and Lectin Analysis of N-Linked Glycans in Carbohydrate Antigen-Defined Pancreatic Cancer Tissues. *Mol. Cell Proteomics* 20, 100012. doi:10.1074/mcp.ra120.002256

Mise, K., Imamura, M., Yamaguchi, S., Watanabe, M., Higuchi, C., Katayama, A., et al. (2021). Novel Urinary Glycan Biomarkers Predict Cardiovascular Events

in Patients with Type 2 Diabetes: A Multicenter Prospective Study with 5-Year Follow up (U-CARE Study 2). *Front. Cardiovasc. Med.* 8, 373. doi:10.3389/FCVM.2021.668059

Nawaz, M., Camussi, G., Valadi, H., Nazarenko, I., Ekström, K., Wang, X., et al. (2014). The Emerging Role of Extracellular Vesicles as Biomarkers for Urogenital Cancers. *Nat. Rev. Urol.* 11, 688–701. doi:10.1038/nrurol.2014.301

Nyalwidhe, J. O., Betesh, L. R., Powers, T. W., Jones, E. E., White, K. Y., Burch, T. C., et al. (2013). Increased Bisecting N-acetylglucosamine and Decreased Branched Chain Glycans of N-linked Glycoproteins in Expressed Prostatic Secretions Associated with Prostate Cancer Progression. *Prot. Clin. Appl.* 7, 677–689. doi:10.1002/prca.201200134

Otto, J. J., Correll, V. L., Engstroem, H. A., Hitefield, N. L., Main, B. P., Albracht, B., et al. (2020). Targeted Mass Spectrometry of a Clinically Relevant PSA Variant from Post-DRE Urines for Quantitation and Genotype Determination. *Prot. Clin. Appl.* 14, 2000012. doi:10.1002/prca.202000012

Principe, S., Kim, Y., Fontana, S., Ignatchenko, V., Nyalwidhe, J. O., Lance, R. S., et al. (2012). Identification of Prostate-Enriched Proteins by In-Depth Proteomic Analyses of Expressed Prostatic Secretions in Urine. *J. Proteome Res.* 11, 2386–2396. doi:10.1021/pr2011236

Saraswat, M., Joenväära, S., Musante, L., Peltoniemi, H., Holthofer, H., and Renkonen, R. (2015). N-linked (N-) Glycoproteomics of Urinary Exosomes. *Mol. Cel. Proteomics* 14, 2298. doi:10.1074/mcp.A114.040345

Scott, E., and Munkley, J. (2019). Glycans as Biomarkers in Prostate Cancer. *Int. J. Mol. Sci.* 20, 1389. doi:10.3390/ijms20061389

Song, W., Zhou, X., Benktander, J. D., Gaunitz, S., Zou, G., Wang, Z., et al. (2019). In-Depth Compositional and Structural Characterization of N-Glycans Derived from Human Urinary Exosomes. *Anal. Chem.* 91, 13528–13537. doi:10.1021/ACS.ANALCHEM.9B02620

Tkac, J., Bertok, T., Hires, M., Jane, E., Lorencova, L., and Kasak, P. (2019). Glycomics of Prostate Cancer: Updates. *Expert Rev. Proteomics* 16, 65–76. doi:10.1080/14789450.2019.1549993

Van Gils, M. P. M. Q., Cornel, E. B., Hessels, D., Peelen, W. P., Witjes, J. A., Mulders, P. F. A., et al. (2007). Molecular PCA3 Diagnostics on Prostatic Fluid. *Prostate* 67, 881–887. doi:10.1002/pros.20564

Wang, Y.-T., Shi, T., Srivastava, S., Kagan, J., Liu, T., and Rodland, K. D. (2020). Proteomic Analysis of Exosomes for Discovery of Protein Biomarkers for Prostate and Bladder Cancer. *Cancers* 12, 2335–2419. doi:10.3390/cancers12092335

West, C. A., Liang, H., Drake, R. R., and Mehta, A. S. (2020). New Enzymatic Approach to Distinguish Fucosylation Isomers of N-Linked Glycans in Tissues Using MALDI Imaging Mass Spectrometry. *J. Proteome Res.* 19, 2989–2996. doi:10.1021/acs.jproteome.0c00024

White, K. Y., Rodemich, L., Nyalwidhe, J. O., Comunale, M. A., Clements, M. A., Lance, R. S., et al. (2009). Glycomic Characterization of Prostate-specific Antigen and Prostatic Acid Phosphatase in Prostate Cancer and Benign Disease Seminal Plasma Fluids. *J. Proteome Res.* 8, 620–630. doi:10.1021/pr8007545

Zou, G., Benktander, J. D., Gizaw, S. T., Gaunitz, S., and Novotny, M. V. (2017). Comprehensive Analytical Approach toward Glycomic Characterization and Profiling in Urinary Exosomes. *Anal. Chem.* 89, 5364–5372. doi:10.1021/acs.analchem.7b00062

Glycoproteomic Characterization of FUT8 Knock-Out CHO Cells Reveals Roles of FUT8 in the Glycosylation

Ganglong Yang[1]*, Qiong Wang[2], Lijun Chen[1], Michael J. Betenbaugh[2] and Hui Zhang[1,2]

[1]Department of Pathology, Johns Hopkins University, Baltimore, MD, United States, [2]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, United States

The α1,6-fucosyltransferase (encoded by FUT8 gene) is the key enzyme transferring fucose to the innermost GlcNAc residue on an N-glycan through an α-1,6 linkage in the mammalian cells. The presence of core fucose on antibody Fc region can inhibit antibody-dependent cellular cytotoxicity (ADCC) and reduce antibody therapeutic efficiency *in vivo*. Chinese hamster ovary (CHO) cells are the predominant production platform in biopharmaceutical manufacturing. Therefore, the generation of FUT8 knock-out (FUT8KO) CHO cell line is favorable and can be applied to produce completely non-fucosylated antibodies. The characterization of monoclonal antibodies as well as host cell glycoprotein impurities are required for quality control purposes under regulation rules. To understand the role of FUT8 in the glycosylation of CHO cells, we generated a FUT8 knock-out CHO cell line and performed a large-scale glycoproteomics to characterize the FUT8KO and wild-type (WT) CHO cells. The glycopeptides were enriched by hydrophilic chromatography and fractionated 25 fractions by bRPLC followed by analysis using high-resolution liquid chromatography mass spectrometry (LC-MS). A total of 7,127 unique N-linked glycosite-containing intact glycopeptides (IGPs), 928 glycosites, and 442 glycoproteins were identified from FUT8KO and WT CHO cells. Moreover, 28.62% in 442 identified glycoproteins and 26.69% in 928 identified glycosites were significantly changed in the FUT8KO CHO compared to wild-type CHO cells. The relative abundance of all the three N-glycan types (high-mannose, hybrid, and complex) was determined in FUT8KO comparing to wild-type CHO cells. Furthermore, a decrease in fucosylation content was observed in FUT8KO cells, in which core-fucosylated glycans almost disappeared as an effect of FUT8 gene knockout. Meantime, a total of 51 glycosylation-related enzymes were also quantified in these two cell types and 16 of them were significantly altered in the FUT8KO cells, in which sialyltransferases and glucosyltransferases were sharply decreased. These glycoproteomic results revealed that the knock-out of FUT8 not only influenced the core-fucosylation of proteins but also altered other glycosylation synthesis processes and changed the relative abundance of protein glycosylation.

Keywords: glycoproteomics, FUT8 knock-out CHO cells, mass spectrometry, intact glycopeptides, glycosylation related enzyme

Abbreviations: IGP, intact glycopeptide; CHO, Chinese hamster ovary; PSM, peptide spectrum match; LC-MS, liquid chromatography mass spectrometry.

1 INTRODUCTION

Chinese hamster ovary (CHO) cells have been extensively applied in pharmaceutical manufacturing to produce therapeutic recombinant proteins, especially monoclonal antibodies. The merits of CHO cell culture include the adaption ability to grow at high densities in suspension cultures, high levels of protein expression over prolonged fermentation cycles, the incorporation of complex glycans on exogenously expressed proteins and the capacity to perform post-translational modifications compatible with humans (Byrne et al., 2018; Mizukami et al., 2018). Currently, CHO cells are involved in the production of more than 70% of recombinant biopharmaceutical proteins, most of them being monoclonal antibodies (mAbs) (Lalonde and Durocher, 2017). The glycan profiles on recombinant glycoproteins, including monoclonal antibodies, have shown different influences on their biological, pharmacokinetic, and immunogenic properties (Sethuraman and Stadheim, 2006). The diversity and complexity of these N-glycans can be attributed to the high number of different sugar moieties and the multitude of possible linkages. For example, mannosylation affects pharmacokinetics (Liu, 2015). Sialylation is associated with anti-inflammatory properties, and terminal galactosylation involves in complement dependent cytotoxicity (Peschke et al., 2017; Aoyama et al., 2019; Markina et al., 2020). Core fucosylation impacts protein activity, antibody-dependent cell-mediated cytotoxicity (ADCC) and antibody-dependent cell-mediated phagocytosis (Chang et al., 2019). Therefore, the glycopatterns on therapeutic proteins are an important quality attribute and the control and monitoring of glycosylation is necessary for the pharmaceutical industry to maintain lot-to-lot consistency.

Given its non-template-driven nature, various parameters, intracellularly and extracellularly, can affect the final glycoforms on glycoproteins in mammalian cells. There are two strategies to control the glycosylation of therapeutic proteins, manipulation of culture conditions and gene editing. Process factors such as pH, ammonia level, dissolved O_2, temperature and bioreactor mode can influence the glycosylation in CHO cells (Wong et al., 2010). Precise gene editing provides a nearly unlimited playground for producing tailored glycans on therapeutic proteins. Moreover, gene editing of glycosylation in mammalian cells could contribute tremendously to the current view of biosynthetic regulation of the cellular glycome and highlight important functions of glycans involved in development, health, and disease (Byrne et al., 2018; Narimatsu et al., 2021). Several glycoengineering strategies have emerged to recreate these beneficial profiles on recombinant proteins. Afucosylation presents one of the most popular glycoengineering approaches for biologics (Lalonde and Durocher, 2017). Afucosylation has been achieved by knocking out FUT8 or the other enzymes involved in the biosynthesis of guanine diphosphate-fucose (Cristea et al., 2013; Sun et al., 2015).

The genomics, transcriptomics, proteomics, metabolomics, glycomics and other omics were extensively applied in the therapeutic protein bioproduction of CHO cells to maximize gains from CHO engineering and bioprocess improvements.

Glyco-analytical technologies are essential for the characterization of therapeutic glycoproteins for drug discovery, development, and for quality control of lot and batch variations to ensure safety and efficacy of the biotherapeutic (O'Flaherty et al., 2018). Glycoproteomics is a powerful tool for glycoprotein analysis, which could identify the diverse glycan sites on each glycoprotein as well as characterize the heterogeneous glycoforms at each site (Kolarich et al., 2012; Gutierrez Reyes et al., 2021). To understand the role of FUT8 in the glycosylation of CHO cells, we analyzed the glycoproteome of FUT8KO and wild-type CHO-K1 cells (Sun et al., 2016; Yang et al., 2018). We enriched the glycopeptides by hydrophilic column MAX and fractionated 25 fractions by basic reversed-phase liquid chromatography (bRPLC). The intact glycopeptides were analyzed by high-resolution LC-MS/MS and annotated by GPQuest. From our results, it revealed that the FUT8KO not only influenced the glycosylation of proteins but also changed the relative abundance of some proteins.

2 MATERIALS AND METHODS

2.1 1FUT8 Knock-Out CHO Cell Culture and Protein Digestion

To silence α-1,6-fucosylation on antibodies expressed in CHO cells, we applied CRISPR-Cas9 method to disrupt the α-1,6-fucosyltranferase (FUT8) gene (Wang et al., 2018). The stable FUT8KO and wild-type (WT) CHO cells were adapted to suspension cell culture status in CD-CHO media with 5 mM glutamine. Cells were seeded at 0.4 million/ml and incubated in shaker flasks for 3 days. After 3-day cell culture, the cell pellets were collected and washed twice with phosphate buffered saline (PBS, pH 7.4) buffer. Cells were then lysed directly with 8 M urea/1 M NH_4HCO_3 solution and briefly sonicated until the solutions were clear. Protein concentrations were determined by bicinchoninic acid (BCA) protein assay reagent (Thermo Scientific, Fair Lawn, NJ, United States). Proteins were reduced by 10 mM DTT at 37°C for 1 h and then the cysteines on proteins were alkylated by 20 mM iodoacetamide for 45 min in dark. Then sequencing-grade trypsin (protein enzyme, 40:1, w/w; Promega, Madison, WI, United States) was added to the diluted proteins and incubated at 37°C overnight. The peptides were cleaned by C18 solid-phase extraction.

2.2 Glycopeptide Enrichment and Fractionation

The glycopeptides were enriched by mixed anion exchange (MAX) extraction cartridges and then fractionated by basic reversed-phase liquid chromatography (bRPLC) (Yang et al., 2018). In brief, 1 mg peptides were loaded into one Oasis MAX extraction cartridges (Waters) and about 200 μg glycopeptide eluate was collected together from three MAX cartridges. Then, the collected glycopeptides were desalted by C18 solid-phase extraction and injected into Zorbax Extend-C18 analytical column containing 1.8 μm particles at a flow rate of 0.2 ml/min. The glycopeptides were fractionated and collected

into 96-well plates in a time-based mode from 16 to 112 min. The resulting 96 fractions of tryptic glycopeptide samples were further concentrated into 24 fractions by combining fractions 1, 25, 49, 73; 2, 26, 50, 74; etc. The 24 fraction samples were then dried in a speed-vac and stored at −80°C until LC-MS/MS analysis.

2.3 Nano LC–MS/MS Analysis

The IGPs were subjected to two LC-MS/MS analytical repeats per fraction in a Q-Exactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The samples were re-suspended with 3% ACN and 0.1% FA. 1 μg in 4 μl per sample was first separated in a Dionex Ultimate 3000 RSLC nano system (Thermo Scientific) with a PepMap RSLC C18 column (75 μm 50 cm, 2 μm, Thermo Scientific) protected by an Acclaim PepMap C18 column (100 μm 2 cm, 5 μm, Thermo Scientific). The mobile phase consisted of 0.1% FA and 3% ACN in water (A) and 0.1% FA, 90% ACN (B) using a gradient elution of 0–2% B, 1 min; 2–8% B, 9 min; 8–31% B, 80 min, 31–38% B, 20 min; 38–95% B, 5 min; 95% B, 10 min; 95–2% B, 4 min. The flow rate was kept at 0.3 μl/min. Data-dependent higher-energy collisional dissociation (HCD) fragmentation was performed on the 12 most abundant ions. The spray voltage was set to 1.5 kV. Spectra (AGC target 3×10^6 and maximum injection time 60 ms) were collected from 400–2,000 m/z at a resolution of 70,000 followed by data-dependent HCD MS/MS (at a resolution of 35,000, NCE 32%, intensity threshold of 4.2×10^4, AGC target 2×10^5 and maximum injection time 120 ms) of the ions using an isolation window of 1.4 m/z. Charge state screening was enabled to reject singly charged ions and ions with more than eight charges. A dynamic exclusion time of 30 s was used to discriminate against previously selected ions.

2.4 MS Data Analysis

The MS data were searched by an in-house developed glycopeptide analysis software called GPQuest 2.0[20]. The database of glycosites was the glycosite-containing peptide data from our previous de-glycopeptide methods (Yang et al., 2018). The Cricetulus griseus glycan database was from a previous CHO cell glycomics profiling study (North et al., 2010). The databases contain 57,653 predicted glycosites and 343 glycan structure entries. Prior to database search, MSConvertGUI was used to convert the.RAW files to.mzML files with the "pick picking" option selected. The parameters for mass tolerance of MS1 and MS2 were 10 and 20 ppm, respectively. The spectra containing an oxonium ion m/z 204.09 were chosen for further searching. Results were filtered based on the following criteria: 1) false discovery rate (FDR) less than 1%, 2) ≥ 2 PSMs for each peptide were required, and 3) all the PSMs should be annotated by at least one N-linked glycans on one glycosites. All the core-fucosylated intact glycopeptides were verified by the peptide + GlcNAc2Fuc1 or peptide + GlcNAc1Fuc1 fragments in the MS/MS spectra.

2.5 Quantification of Glycoproteins and Intact Glycopeptides

For the glycoprotein quantitative analysis, the relative abundances of glycoproteins in each cell type were determined by the summed PSM value of each protein's constituent

glycopeptides from 24 fraction analyses in two replications together. This label-free quantification method was also applied for glycan quantitative analysis. The medium log2 ratio value of an identical glycoprotein or glycan was calculated by the ratio of the relative abundance of the identical glycoprotein or glycan in FUT8KO cells to that in WT cells. As in **Figure 3B**, the ratio of core-fucosylated glycoproteins in fucosylated glycoproteins was determined by their corresponding relative abundances for each glycoprotein.

2.6 Glycosylation-Related Enzymes and Their Glycosylation Analysis

The level of glyco-related enzymes, such as glycosyltransferases and glycosidases can influence the expression of some glycoproteins intracellularly and extracellularly. The identified glycoproteins were blasted with glycosylation-related genes data from RNA-seq analysis and classified into two major categories as glycosyltransferases and glycan degradation enyzmes and 24 sub-catergries like mannosyltransferases, galactosyltransferases and mannosidases, etc. (Xu et al., 2011). The illustration of glycosylated proteins was created using Tbtools (Chen et al., 2020), which relied on the relative abundance of each glycan composition at each glycosite. The schematic representation of the glycoprotein was generated from Pfam 34.0[24].

3 RESULTS AND DISCUSSION

3.1 Identification of Intact Glycopeptides in the WT and FUT8 KO CHO Cells

To understand the role of FUT8 in the glycosylation of CHO cells, we developed a FUT8 knockout CHO cell line for large-scale glycoproteomic analysis (Wang et al., 2018). To characterize the glycoproteomics of FUT8KO as well as wild-type CHO cells, glycopeptides were enriched using hydrophilic MAX extraction column and fractionated by basic RPLC. The intact glycopeptides were analyzed by Q-Exactive mass spectrometer and identified by GPQuest 2.0. The assigned intact glycopeptides were filtered using peptide spectrum matches (PSMs) with a maximum of 1% false discovery rate (FDR), the morpheus score higher than 6, and the number of PSMs of peptide more than 2. A total of 25,859 intact glycopeptide spectra in WT CHO cells and 21,045 intact glycopeptide spectra in FUT8KO cells were annotated (**Supplementary Tables S1, S2**). In the WT CHO cells, 5,159 intact glycopeptides from 405 glycoproteins containing 837 glycosites and 155 glycan compositions were identified, while 4,607 intact glycopeptides from 362 glycoproteins, 743 glycosites, and 147 glycan compositions were identified in FUT8KO CHO cells. In combination, a total of 442 glycoproteins with 928 glycosites, 181 glycan compositions, and 7,127 unique N-linked glycosite-containing IGPs were identified from FUT8KO and WT CHO cells (**Figure 1A**). Interestingly, we noticed that the number of unique IGPs was decreased in the FUT8KO CHO cells. It showed the same tendency for the high-mannosylated, fucosylated and other complex or hybrid IGPs. Conversely, for sialylated IGPs, its number was increased in the
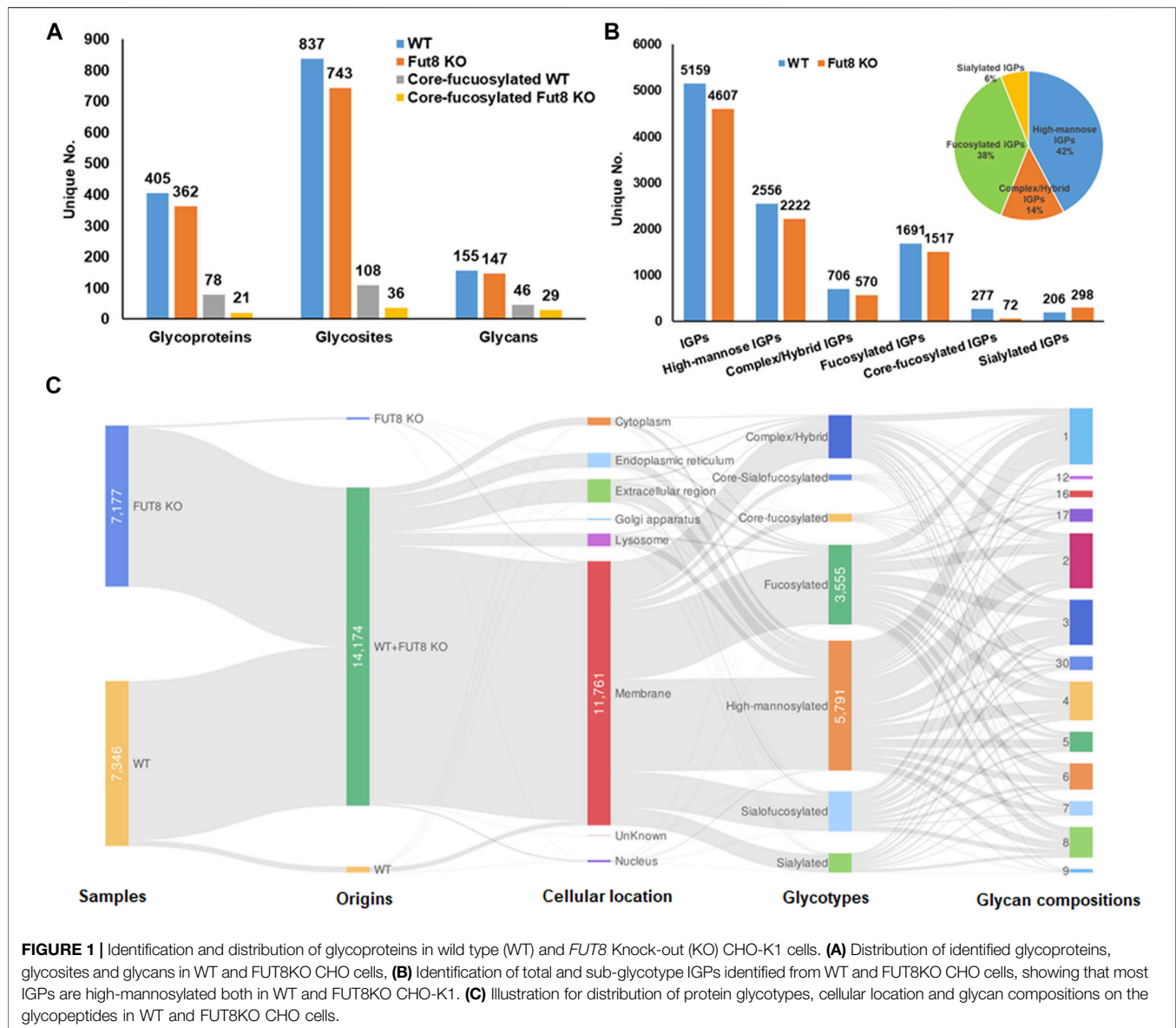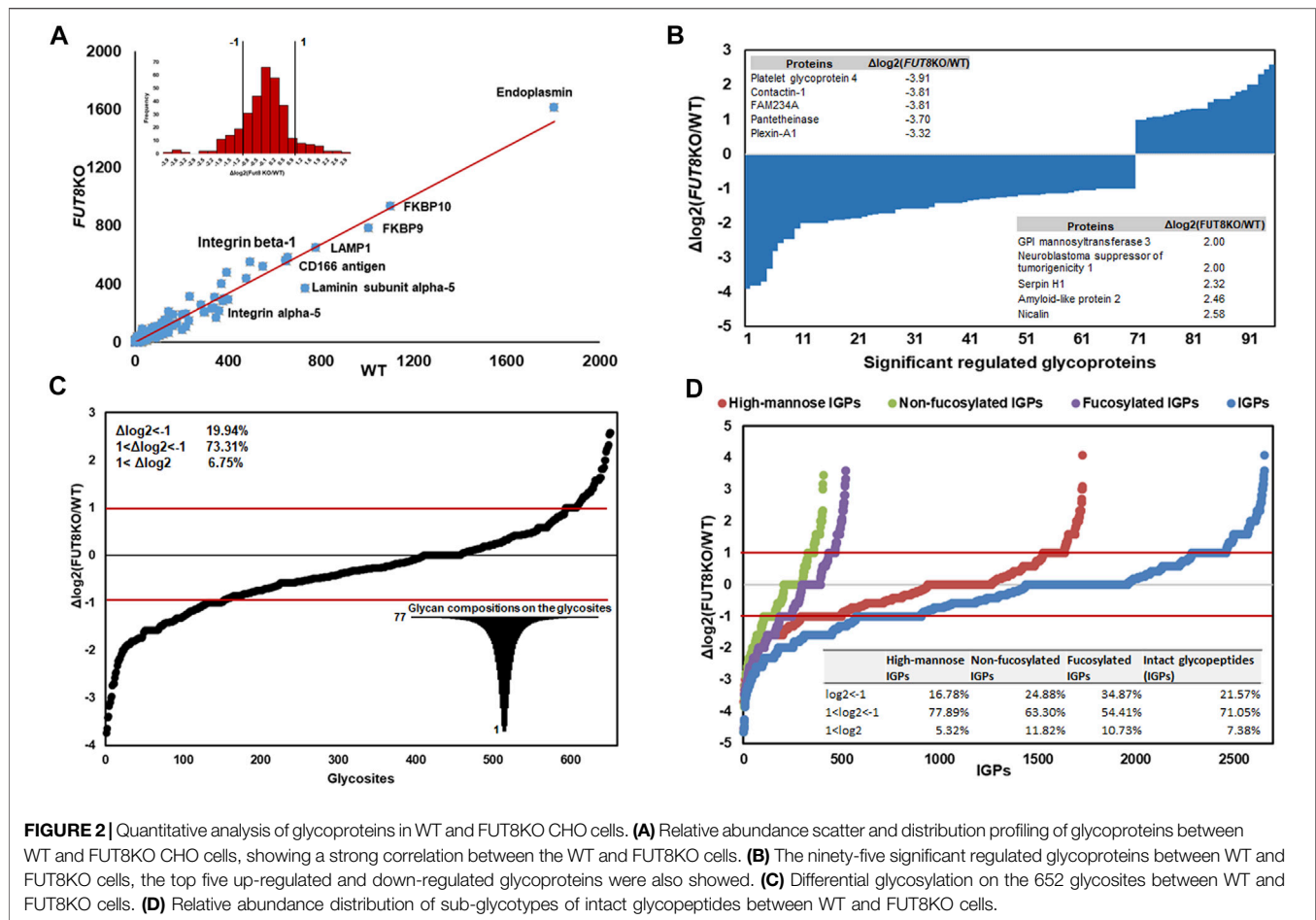
FIGURE 1 | Identification and distribution of glycoproteins in wild type (WT) and *FUT8* Knock-out (KO) CHO-K1 cells. **(A)** Distribution of identified glycoproteins, glycosites and glycans in WT and FUT8KO CHO cells, **(B)** Identification of total and sub-glycotype IGPs identified from WT and FUT8KO CHO cells, showing that most IGPs are high-mannosylated both in WT and FUT8KO CHO-K1. **(C)** Illustration for distribution of protein glycotypes, cellular location and glycan compositions on the glycopeptides in WT and FUT8KO CHO cells.

FUT8KO cells (**Figure 1B**). In the WT parental CHO cells, approximately 5.36% in the total IGPs and 16.38% in all the fucosylated IGPs were core-fucosylated, which were confirmed by the peptide + GlcNAc2Fuc1 or peptide + GlcNAc1Fuc1 fragments in the MS/MS spectra. Meantime, there were about 1.56% in the total IGPs and 4.75% in all the fucosylated IGPs were core-fucosylated in the FUT8KO cells (**Figure 1B**). In the 7,127 IGPs identified from FUT8KO and WT cells together, about 95.92% IGPs were shared by both cell lines, indicating the genetic removal of FUT8 could rarely change the glycan composition of the protein glycosites. Moreover, about 42 and 38% glycosites in all glycosites were modified by high-mannosylated and fucosylated glycan structures, respectively (**Figure 1B**). We also noticed that 25.97% of the IGPs were super-microheterogeneity with six to ten glycan compositions per peptide sequence and 53.65% of the IGPs were hyper-

microheterogeneity with more than ten glycan compositions at one glycosite. The average glycan compositions at one glycosites are about eight. From the cellular location distribution of the glycoproteins, it showed that most of the glycoproteins were membrane proteins (**Figure 1C** and **Supplementary Table S3**).

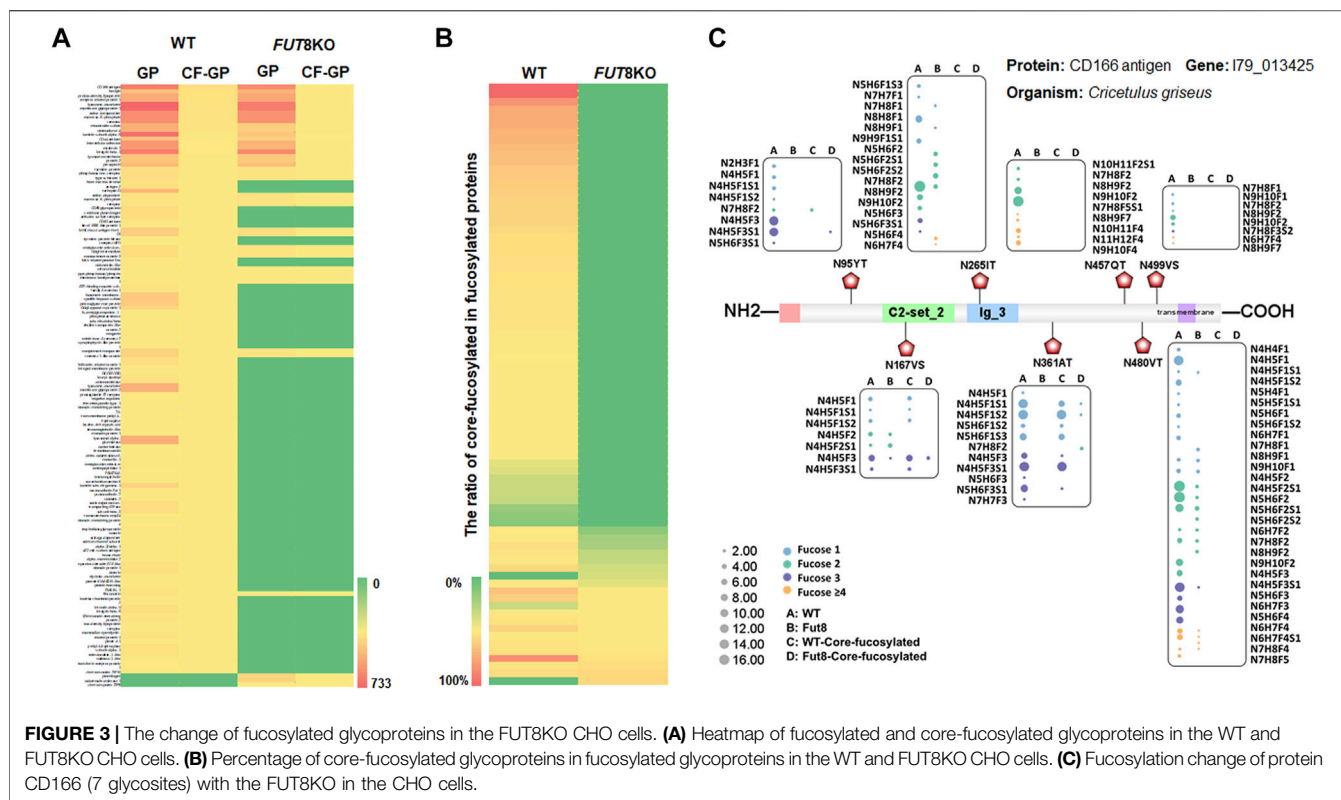## 3.2 Relative Abundance of Glycopeptides in the WT and FUT8KO CHO Cells

In principle, the FUT8 gene is involved in the core-fucosylation of glycoproteins, which catalyzes the transfer of a fucose residue to the innermost GlcNAc residue of N-linked oligosaccharide. To investigate the changing of the glycosylation regulated by FUT8, we performed a comparative analysis using label-free quantification based on spectral counting. To quantitatively

FIGURE 2 | Quantitative analysis of glycoproteins in WT and FUT8KO CHO cells. **(A)** Relative abundance scatter and distribution profiling of glycoproteins between WT and FUT8KO CHO cells, showing a strong correlation between the WT and FUT8KO cells. **(B)** The ninety-five significant regulated glycoproteins between WT and FUT8KO cells, the top five up-regulated and down-regulated glycoproteins were also showed. **(C)** Differential glycosylation on the 652 glycosites between WT and FUT8KO cells. **(D)** Relative abundance distribution of sub-glycotypes of intact glycopeptides between WT and FUT8KO cells.

analyze the glycoproteins, 325 glycoproteins identified in the replication of the two cell lines were analyzed as shown in **Figure 2**. It revealed that the protein Endoplasmin was the highest abundant glycoproteins in the CHO cells, which may participate in the unfolding of cytosolic leaderless cargos. From the relative abundance scatter in **Figure 2A**, it showed that there was no significant expression difference between the WT and FUT8KO CHO cells. The log2 (ratio) of 232 (71.38%) proteins were varied less than two folds. Meantime, 28.62% of the identified glycoproteins were significantly changed, in which 22 (6.77%) and 71 (21.85%) out of the 325 proteins were upregulated and downregulated in the FUT8KO CHO cells. The top five of the upregulated and downregulated proteins were showed in **Figure 2B**, which showed the downregulated proteins were more significantly changed than the upregulated proteins. Especially, the expression of platelet glycoprotein 4 was sharply decreased in the FUT8KO CHO cells compared to WT CHO cells, which was a multifunctional glycoprotein that acts as a receptor for a broad range of ligands. From the comparative analysis of the glycoprotein abundance in the WT and FUT8KO CHO cells, it indicated that FUT8KO CHO cells not only altered the glycan profile by intervening in the glycosylation process in *Golgi* but also regulated the relative abundance of some glycoproteins.

The number of N-glycans is a distinct feature of each glycoprotein sequence and cooperates with the physical properties of the *Golgi* N-glycan-branching pathway to regulate surface glycoprotein levels (Lau et al., 2007). To reveal the influence of FUT8 on the protein N-glycosylation, we also quantitatively analyzed the glycopeptides between the WT and FUT8KO CHO cells. It showed that the diversity of glycosylation was changed due to the knock-out of FUT8 in CHO cells. In this study, 928 glycosites were identified and 652 glycosites were identified both in the WT and FUT8KO CHO cells. In the 652 glycosites identified from two cells, approximately 478 (73.31%) glycosites were changed less than two folds, and 26.69% in the identified glycosites were significantly changed, in which 130 (19.94%) were downregulated and 44 (6.75%) were upregulated. The glycopeptides $N_{457}QT$ from CD166 antigen were most downregulated with the log2 (ratio) as −3.74 in the FUT8KO CHO cells (**Figure 2C**). Interestingly, we also noticed that a great diversity of N-glycosylation on these regulated glycosites. For example, there were 77 glycan compositions on $N_{67}SS$ from protein basigin.

In addition to the glycosites, the glycotype on IGPs was also altered in the FUT8KO CHO cells. Approximately 2,656 IGPs were identified both in the WT and FUT8KO CHO cells and

FIGURE 3 | The change of fucosylated glycoproteins in the FUT8KO CHO cells. **(A)** Heatmap of fucosylated and core-fucosylated glycoproteins in the WT and FUT8KO CHO cells. **(B)** Percentage of core-fucosylated glycoproteins in fucosylated glycoproteins in the WT and FUT8KO CHO cells. **(C)** Fucosylation change of protein CD166 (7 glycosites) with the FUT8KO in the CHO cells.

about 71.05% of IGPs relative abundances were changed less than two folds. It indicated that fucosylated IGPs were altered significantly and about 238 out of 522 (45.60%) fucosylated glycopeptides were upregulated (34.87%) or downregulated (10.73%) in the FUT8KO CHO cells.

3.3 Differential Analysis of the Core-fucosylated Proteins in the FUT8KO and Wild-type CHO Cells

From the relative abundance analysis of intact glycopeptides, we found that fucosylated glycopeptides changed significantly in FUT8KO CHO cells. A total of 309 core-fucosylated IGPs were identified, representing 114 glycosites, 52 glycans and 80 core-fucosylated proteins. 237 fucosylated IGPs from 59 (73.75%) glycoproteins totally lost their core-fucosylation in FUT8KO CHO cells. Moreover, 72 out of 4,607 IGPs from 21 glycoproteins were still identified as core-fucosylated glycopeptides with very low relative abundance in FUT8KO CHO cells. Besides the core-fucosylated proteins, the terminal-fucosylated proteins were also decreased in FUT8KO CHO cells (5 proteins significantly down-regulated and 21 proteins disappeared). Interestingly, 2 proteins (procollagen galactosyltransferase 1 and zinc transporter ZIP6) were not fucosylated in the WT but fucosylated in FUT8KO CHO cells (**Figure 3A**). It indicated that the ratio of core-fucosylated in all fucosylated was sharply decreased from 15.42% in WT cells to 1.14% in FUT8KO cells for glycoproteins and from 72.70% in WT cells to 16.18% in FUT8KO CHO cells for glycopeptides

(**Figure 3B**). As one of the hyper-fucosylated glycoproteins, CD166 antigen has seven fucosylated glycosites identified in which three glycosites were core-fucosylated (**Figure 3C**). The most heterogeneous site $N_{480}VT$ of CD166 antigen contained 31 fucosylated glycan compositions. Moreover, the relative abundances of all the glycan types were decreased in FUT8KO CHO cells. The core-fucosylated glycosylation of the three glycosites on CD166 antigen were almost disappear in the FUT8KO CHO cells (**Figure 3C**).

3.4 Alteration of the N-Linked Glycans and Glycosylation-Related Enzymes in FUT8KO CHO Cells

Furthermore, to investigate the influence of FUT8KO on the N-linked glycans, the glycans from IGPs were analyzed and 181 glycans were identified in total and grouped into five subtypes, high-mannose, fucosylated, sialofucosylated, sialylated and other complex or hybrid N-glycans. About 34 and 26 glycans were only presented in the WT and FUT8KO cells. The relative abundances of 36 glycan compositions were increased and 46 glycan compositions were decreased more than two folds in the FUT8KO CHO cells. From the quantitative analysis of glycan subtypes in **Figure 4A**, it displayed that the core-fucosylation nearly disappeared. The high-mannosylated and fucosylated glycans were generally decreased while sialylated glycans were increased.

The glycosylation of protein is a complex and multistep networking that employs about 200 glycotransferases and

FIGURE 4 | The change of glycosylation and expression of glycosylation related enzymes in FUT8KO CHO cells. **(A)** The change of glycans on the glycoproteins in the WT and FUT8KO CHO cells. **(B)** The expression of glyco-related enzyme in the WT and FUT8KO CHO cells. **(C)** The glycosylation of N-glycan transferase dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A.

glycosidases. These enzymes determine which protein could be glycosylated, the glycosylation site of glycans on the proteins and the glycan structures on the sites. Studies of deficiencies in glycosylation-related enzymes could be facilitated to understand the biological functions of protein glycosylation. From our high-throughput site-specific glycoproteomics, it showed that around 51 glycosylation-related enzymes were also glycosylated, which were classified into 28 glycosyltransferases and 23 glycosidases. It has been reported that 159 glycosylation-related genes were expressed in the CHO cell by RNA-seq data (Xu et al., 2011). It also showed that the relative abundance of 4 glycosyltransferases was obviously increased and 12 glycosylation-related enzymes were obviously decreased in the FUT8KO CHO cells, and all the significantly changed glycan degradation enzymes were decreased (**Figure 4B** and **Supplementary Table S4**). In these glycosylation-related enzymes, lysosomal alpha-glucosidase was the most abundant proteins and with the greatest macroheterogeneity (6 glycosites) and microheterogeneity (17 glycan compositions on $N_{134}LS$). Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A catalyzes the initial transfer of $Glc_3Man_9GlcNAc2$ from the lipid carrier dolichol-pyrophosphate to an asparagine residue in nascent polypeptide chains in the endoplasmic reticulum. It is present in the majority of oligosaccharyltransferase (OST) complexes and mediates N-glycosylation of most sites on target proteins. Three glycosites were identified from the human liver tissue (Chen et al., 2009) and two glycosites were identified from CHO cells either by the de-glycosylated glycopeptide or site-specific intact glycopeptide identification methods in our previous study (Yang et al., 2018). In this study, the two glycosites were identified with 21 glycan compositions on $N_{544}NT$ and 5 glycan compositions on $N_{537}RT$. Interestingly, the glycan compositions on $N_{544}NT$ from

STT3A were increased in the FUT8KO cells, which were modified with 16 glycan compositions in WT and 20 glycan compositions in FUT8KO CHO cells (**Figure 4C**).

3.5 Discussion

In our previous study, we developed a one-step glycopeptide enrichment method and applied this methodology to analyze the protein glycosylation of FUT8KO cells as a proof of concept, from which we identified 2,634 unique IGPs from 243 glycoproteins with 459 glycosites (Yang et al., 2020). Some interesting results indicated that the core-fucosylated glycopeptide was still present in the FUT8KO CHO cells. To comprehensively decipher the protein glycosylation in the FUT8KO CHO cells, we performed a thorough analysis of the site-specific protein glycosylation of the FUT8KO CHO cells using the comprehensive glycoproteome method in this study. 7,127 unique IGPs from 442 glycoproteins with 928 glycosites were identified, which showed more than two-fold IGPs and glycosites identification compared to our previous study (Yang et al., 2020).

To explore the role of core-fucosylation in cells, it is desirable to comprehensively analyze the site-specific fucosylated glycoproteins. In this study, 72 core-fucosylated glycopeptides out of 4,607 IGPs were identified from FUT8KO CHO cells which are from 21 glycoproteins with very low core-fucosylation. The average ratio of core-fucosylated proteins in the glycoproteins was decreased from 0.13 in WT to 0.03 in FUT8KO CHO cells. For these 21 core-fucosylated glycoproteins, most of them were distributed on the plasma membrane and lysosomal membrane. It indicated that the core-fucosylation of these proteins was related to receptor activity, cell adhesion and migration.

As a non-template-mediated post-translational modification (PTM), glycans are not directly regulated by the genetic code. Instead, the glycosylation of proteins was controlled by the

expression levels of glycosyltransferase and glycosidase enzymes. The observation that the N-glycan abundance is decreased in the FUT8KO CHO cells compared with WT CHO cells suggests that there is dysregulation of other enzymes responsible for N-glycan biosynthesis. OST complex catalyzes the transfer of the oligosaccharide from Dol-P-P to Asn-X-Ser/Thr in newly synthesized regions of proteins (Stanley et al., 2017). It is composed of couples of subunits contained STT3A/STT3B, OST4, RPN1, RPN2, etc. The RPN1 and RPN2 were considered to provide binding sites for translating ribosomes, associate with translocon and enable co-translocational glycosylation of entering polypeptides (Mohorko et al., 2011). The significant down-regulation of OST complex subunit RPN1 and RPN2 depressed the expression of all N-linked glycoproteins and the glycosylation level of the glycoproteins. Hence, targeting biosynthetic steps in glycosylation generally results in predictable global outcomes, but can also causes some minor unpredictable glycosylation effects. Therefore, it is desirable to monitor the glycosylation of the proteins in glycoengineered cells.

In this study, we performed an integrated and comprehensive glycoproteomic analysis of FUT8KO and parental wild-type CHO-K1 cells and compared the site-specific protein glycosylation and glyco-related enzymes difference between these two cell lines. By the site-specific glycosylation analysis method, we investigated the influence of FUT8 gene knock-out on proteins glycosylation and their site specificity. Moreover, the whole glycosylation biosynthetic pathways of CHO cells were also changed in the FUT8KO CHO cells. The glycoproteomic data will improve our understanding of the roles of α1,6-fucosyltransferase and the resulting core-fucosylation in CHO cells.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the **Supplementary Tables S1 and S2**.

AUTHOR CONTRIBUTIONS

HZ, GY, and MB conceived the study GY designed and performed the IGP enrichment and MS analysis and conceived the bio-informatic analysis QW provided the Fut8 mutant cell lines and performed cell culture experiments. GY and LC performed and supervised glycoproteomics experiments. GY, HZ, and QW wrote and revised the articles.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.755238/full#supplementary-material

REFERENCES

Aoyama, M., Hashii, N., Tsukimura, W., Osumi, K., Harazono, A., Tada, M., et al. (2019). Effects of Terminal Galactose Residues in Mannose α1-6 Arm of Fc-Glycan on the Effector Functions of Therapeutic Monoclonal Antibodies. *MAbs* 11 (5), 826–836. doi:10.1080/19420862.2019.1608143

Byrne, G., O'Rourke, S. M., Alexander, D. L., Yu, B., Doran, R. C., Wright, M., et al. (2018). CRISPR/Cas9 Gene Editing for the Creation of an MGAT1-Deficient CHO Cell Line to Control HIV-1 Vaccine Glycosylation. *Plos Biol.* 16 (8), e2005817. doi:10.1371/journal.pbio.2005817

Chang, M. M., Gaidukov, L., Jung, G., Tseng, W. A., Scarcelli, J. J., Cornell, R., et al. (2019). Small-molecule Control of Antibody N-Glycosylation in Engineered Mammalian Cells. *Nat. Chem. Biol.* doi:10.1038/s41589-019-0288-4

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009

Chen, R., Jiang, X., Sun, D., Han, G., Wang, F., Ye, M., et al. (2009). Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry. *J. Proteome Res.* 8, 651–661. doi:10.1021/pr8008012

Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M. C., Urnov, F. D., Gregory, P. D., et al. (2013). *In Vivo* cleavage of Transgene Donors Promotes Nuclease-Mediated Targeted Integration. *Biotechnol. Bioeng.* 110 (3), 871–880. doi:10.1002/bit.24733

Gutierrez Reyes, C. D., Jiang, P., Donohoo, K., Atashi, M., and Mechref, Y. S. (2021). Glycomics and Glycoproteomics: Approaches to Address Isomeric Separation of Glycans and Glycopeptides. *J. Sep. Sci.* 44 (1), 403–425. doi:10.1002/jssc.202000878

Hu, Y., Shah, P., Clark, D. J., Ao, M., and Zhang, H. (2018). Reanalysis of Global Proteomic and Phosphoproteomic Data Identified a Large Number of Glycopeptides. *Anal. Chem.* 90, 8065–8071. doi:10.1021/acs.analchem.8b01137

Kolarich, D., Lepenies, B., and Seeberger, P. H. (2012). Glycomics, Glycoproteomics and the Immune System. *Curr. Opin. Chem. Biol.* 16 (1-2), 214–220. doi:10.1016/j.cbpa.2011.12.006

Lalonde, M.-E., and Durocher, Y. (2017). Therapeutic Glycoprotein Production in Mammalian Cells. *J. Biotechnol.* 251, 128–140. doi:10.1016/j.jbiotec.2017.04.028

Lau, K. S., Partridge, E. A., Grigorian, A., Silvescu, C. I., Reinhold, V. N., Demetriou, M., et al. (2007). Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell* 129 (1), 123–134. doi:10.1016/j.cell.2007.01.049

Liu, L. (2015). Antibody Glycosylation and its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins. *J. Pharm. Sci.* 104 (6), 1866–1884. doi:10.1002/jps.24444

Markina, Y. V., Gerasimova, E. V., Markin, A. M., Glanz, V. Y., Wu, W.-K., Sobenin, I. A., et al. (2020). Sialylated Immunoglobulins for the Treatment of Immuno-Inflammatory Diseases. *Ijms* 21 (15), 5472. doi:10.3390/ijms21155472

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* 49 (D1), D412–D419. doi:10.1093/nar/gkaa913

Mizukami, A., Caron, A. L., Picanço-Castro, V., and Swiech, K. (2018). Platforms for Recombinant Therapeutic Glycoprotein Production. *Methods Mol. Biol.* 1674, 1–14. doi:10.1007/978-1-4939-7312-5_1

Mohorko, E., Glockshuber, R., and Aebi, M. (2011). Oligosaccharyltransferase: the central Enzyme of N-Linked Protein Glycosylation. *J. Inherit. Metab. Dis.* 34 (4), 869–878. doi:10.1007/s10545-011-9337-1

Narimatsu, Y., Büll, C., Chen, Y. H., Wandall, H. H., Yang, Z., and Clausen, H. (2021). Genetic Glycoengineering in Mammalian Cells. *J. Biol. Chem.* 296, 100448. doi:10.1016/j.jbc.2021.100448

North, S. J., Huang, H.-H., Sundaram, S., Jang-Lee, J., Etienne, A. T., Trollope, A., et al. (2010). Glycomics Profiling of Chinese Hamster Ovary Cell Glycosylation Mutants Reveals N-Glycans of a Novel Size and Complexity. *J. Biol. Chem.* 285 (8), 5759–5775. doi:10.1074/jbc.m109.068353

O'Flaherty, R., Trbojević-Akmačić, I., Greville, G., Rudd, P. M., and Lauc, G. (2018). The Sweet Spot for Biologics: Recent Advances in Characterization of Biotherapeutic Glycoproteins. *Expert Rev. Proteomics* 15 (1), 13–29. doi:10.1080/14789450.2018.1404907

Peschke, B., Keller, C. W., Weber, P., Quast, I., and Lünemann, J. D. (2017). Fc-Galactosylation of Human Immunoglobulin Gamma Isotypes Improves C1q Binding and Enhances Complement-dependent Cytotoxicity. *Front. Immunol.* 8, 646. doi:10.3389/fimmu.2017.00646

Sethuraman, N., and Stadheim, T. A. (2006). Challenges in Therapeutic Glycoprotein Production. *Curr. Opin. Biotechnol.* 17 (4), 341–346. doi:10.1016/j.copbio.2006.06.010

Stanley, P., Taniguchi, N., and Aebi, M. (2017). Chapter 9 N-Glycans. *Cold Spring Harbor Lab. Press: Cold Spring Harbor (Ny)*, 3.

Sun, S., Shah, P., Eshghi, S. T., Yang, W., Trikannad, N., Yang, S., et al. (2016). Comprehensive Analysis of Protein Glycosylation by Solid-phase Extraction of N-Linked Glycans and Glycosite-Containing Peptides. *Nat. Biotechnol.* 34 (1), 84–88. doi:10.1038/nbt.3403

Sun, T., Li, C., Han, L., Jiang, H., Xie, Y., Zhang, B., et al. (2015). Functional Knockout of FUT8 in Chinese Hamster Ovary Cells Using CRISPR/Cas9 to Produce a Defucosylated Antibody. *Eng. Life Sci.* 15 (6), 660–666. doi:10.1002/elsc.201400218

Wang, Q., Chung, C.-Y., Rosenberg, J. N., Yu, G., and Betenbaugh, M. J. (2018). Application of the CRISPR/Cas9 Gene Editing Method for Modulating Antibody Fucosylation in CHO Cells. *Methods Mol. Biol.* 1850, 237–257. doi:10.1007/978-1-4939-8730-6_16

Wong, N. S. C., Wati, L., Nissom, P. M., Feng, H. T., Lee, M. M., and Yap, M. G. S. (2010). An Investigation of Intracellular Glycosylation Activities in CHO Cells: Effects of Nucleotide Sugar Precursor Feeding. *Biotechnol. Bioeng.* 107 (2), 321–336. doi:10.1002/bit.22812

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., et al. (2011). The Genomic Sequence of the Chinese Hamster Ovary (CHO)-K1 Cell Line. *Nat. Biotechnol.* 29 (8), 735–741. doi:10.1038/nbt.1932

Yang, G., Höti, N., Chen, S.-Y., Zhou, Y., Wang, Q., Betenbaugh, M., et al. (2020). One-Step Enrichment of Intact Glycopeptides from Glycoengineered Chinese Hamster Ovary Cells. *Front. Chem.* 8, 240. doi:10.3389/fchem.2020.00240

Yang, G., Hu, Y., Sun, S., Ouyang, C., Yang, W., Wang, Q., et al. (2018). Comprehensive Glycoproteomic Analysis of Chinese Hamster Ovary Cells. *Anal. Chem.* 90 (24), 14294–14302. doi:10.1021/acs.analchem.8b03520

Mass Spectrometry for O-GlcNAcylation

Ruoting Yin, Xin Wang, Cheng Li, Yuhan Gou, Xuecheng Ma, Yongzhao Liu, Jianfang Peng, Chao Wang and Ying Zhang*

Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an, China

O-linked β-N-acetylglucosamine modification (O-GlcNAcylation) at proteins with low-abundance expression level and species diversity, shows important roles in plenty of biological processes. O-GlcNAcylations with abnormal expression levels are associated with many diseases. Systematically profiling of O-GlcNAcylation at qualitative or quantitative level is vital for their function understanding. Recently, the combination of affinity enrichment, metabolic labeling or chemical tagging with mass spectrometry (MS) have made significant contributions to structure-function mechanism elucidating of O-GlcNAcylations in organisms. Herein, this review provides a comprehensive update of MS-based methodologies for quali-quantitative characterization of O-GlcNAcylation.

Keywords: mass spectrometry, O-GlcNAcylation, O-GlcNAc, O-GlcNAcylated proteins, quali-quantitative charactering, isotope labeling

INTRODUCTION

O-GlcNAcylation, a ubiquitous post-translational modification (PTM) on nuclear, and cytoplasmic proteins (Hart et al., 2007), takes charge of numerous cardinal biological processes, such as signal transduction, transcriptional regulation, stress response, etc. Abnormal expression of O-GlcNAcylation is associated with some diseases, such as alzheimer (Yuzwa and Vocadlo, 2014), diabetes mellitus (Yang et al., 2008) and cancer (Nie et al., 2020). Therefore, the qualitative and quantitative study of glycosylation pattern of O-GlcNAcylated proteins is significant to understand the biological roles of O-GlcNAcylation during a pathological process.

Due to biological importance of O-GlcNAcylation, systematical characterization of O-GlcNAcylation has received increasing attention. However, O-GlcNAcylated proteins with multifarious types are often expressed at low level in organism, such as transcription factor CREB (Rexach et al., 2012) and protein kinase (Dias et al., 2009). Thus, systematically profiling of overall O-GlcNAcylation still faces challenges.

MS with advantages of high sensitivity and traces sample consumption has been widely used in the structural profiling of O-GlcNAcylation (Ma and Hart, 2017). Due to the low abundance and structural diversity of glycosylation, direct MS analysis of O-GlcNAcylation faces challenges. Usually, an enrichment step is necessary for MS-based profiling of O-GlcNAcylation. With the development of stable isotope tagging, quali-quantitative profiling of O-GlcNAcylation has made remarkable progress, accelerating the structure-function mechanism elucidation of O-GlcNAcylated proteins. We summarize the recent research progress in MS-based quali-quantitative analysis of O-GlcNAcylated proteins.

QUALITATIVE CHARACTERIZATION OF O-GLCNACYLATION BY MS

Direct MS

Earlier, collision-induced dissociation (CID), quadrupole time-of-flight (Q-TOF), electron-capture dissociation (ECD) and electron-transfer dissociation (ETD) MS have been used in O-GlcNAcylation analysis. O-GlcNAc shows easier dissociation character over other glycosylation at proteins during ionization procedure, enabling direct MS profiling of O-GlcNAc (Chalkley and Burlingame, 2001). However, the obtained GlcNAc fragment, oxonium ion, often afforded at low yield, leading to signal loss of the O-GlcNAcylation, which might be not suitable for detecting of O-GlcNAcylated proteins expressed at low levels in organism.

Lectin Enrichment for MS

Due to the low expression level of O-GlcNAcylation, an enrichment procedure is usually needed before MS identification of the O-GlcNAcylated proteins. Lectins with feature of bonding GlcNAc have been used in enrichment of the O-GlcNAcylated proteins.

After enriching O-GlcNAcylated proteins by *Ricinus comminis agglutinin* I (RCAI) and Wheat germ agglutinin (WGA) affinity chromatography, O-GlcNAcylated proteins have been well determined by LC-ES/MS (Hayes et al., 1995; Cieniewski-Bernard et al., 2004). Succinylated wheat germ agglutinin (sWGA) and *Agrocybe aegerita lectin* 2 (AAL2), which show better binding specificity over WGA, have been used for O-GlcNAcylated proteins enrichment for subsequent MS profiling (Kupferschmid et al., 2017; Liu et al., 2018).

However, the non-specific binding of lectin to other glycan (N-glycosylated GlcNAc terminal) might decrease the detection accuracy of glycosylation. Thus, a PNGase F digestion is needed before lectin enrichment.

Antibody Enrichment for MS

Pan-specific antibody, CTD110.6 that could bind to O-GlcNAc has been employed to enrich the O-GlcNAcylated proteins to improve MS characterization (Wells et al., 2002). To improve the enrichment of proteins, the combined utilization of three O-GlcNAc-specific IgG monoclonal antibodies [18B10.C7(3), 9D1.E4(10) and 1F5.D6(14)] to immunoprecipitate the O-GlcNAcylated proteins for subsequent O-GlcNAc-omics analysis by MS (Teo et al., 2010).

Given the importance of antibodies enrichment, the low bonding efficiency of antibodies to O-GlcNAcylated proteins and certain peptide dependence might reduce the detection accuracy.

Metabolic Engineering and Solid Phase Enrichment for MS

With the development of metabolic oligosaccharides engineering (MOE), the O-GlcNAcylated proteins could be labeled with the reactive groups (such as alkynyl, azide, etc.)

for subsequent enrichment, as shown in **Figure 1**. Generally, cells were cultured with metabolic chemical reporters (MCRs) such as $Ac_4GlcNAz$ (Sprung et al., 2005), $Ac_4GlcNAlk$ (Zaro et al., 2011), $Ac_36AzGlcNAc$ (Chuh et al., 2014), $Ac_3dGlcNAz$ (Li et al., 2016), $Ac_36AlkGlcNAc$ (Chuh et al., 2017), $Ac_46AzGlc$ (Darabedian et al., 2018), $Ac_36AzGalNAc$ (Guo et al., 2019) and $1,3\text{-}Pr_2GalNAz$ (Hao et al., 2019), etc., to synthesize O-GlcNAcylated proteins with active reactive groups. Then, the biotin probes with corresponding reactive groups (**Figure 1B**) were introduced to tag the labeled O-GlcNAcylations through staudinger linkage, copper-catalyzed azido-alkyne cycloaddition (CuAAC) or strain-promoted azide-alkyne cycloaddition (SPAAC). Finally, the characterization of O-GlcNAcylated proteins could be achieved by MS profiling after the biotin-avidin enrichment.

The combination of MOE and solid phase enrichment for MS profiling has made great contribution in charactering of O-GlcNAcylation. Nevertheless, some unspecific labeling to other glycosylation such as S-glycoylation was observed (Qin et al., 2020).

Chemoenzymatic Labeling and Solid Phase Enrichment for MS

As shown in **Supplementary Figure S1**, GalT Y289L could transfer UDP-galactose analogues with reactive groups (ketone, alkynyl or azide) to C4-position of the O-GlcNAc at proteins. Then the labeled O-GlcNAcylated proteins could be captured by biotins with reactive groups through the orthogonal reactions such as ammoxidatin reaction (Tai et al., 2004) or click chemical reaction (Ma et al., 2019) for MS profiling. Since O-GlcNAc transferase (OGT) can recognize GlcNAc at other glycan terminals, the N-glycosylation interference should be eliminated by a PNGase F digestion before chemoenzymatic labeling.

Nevertheless, the enriched O-GlcNAcylated proteins might be difficult to elute for subsequent MS profiling. The developments of cleavable biotin linkers such as disulfide linker (Tsai et al., 2010), photocleavable linker (Li et al., 2019), acid cleavable linker (Szychowski et al., 2010), and diazobenzene linker (Yang et al., 2010), or affinity column with hydrazide cleavable linker (Nishikaze et al., 2013) to improve the dissociation efficiency of the enriched molecules have enabled more effectively profiling of O-GlcNAcylation by MS.

QUALI-QUANTITATIVE CHARACTERIZATION OF O-GLCNACYLATION BY MS

MS-Based Quali-Quantitative Characterization of O-GlcNAcylation Using $^0D/^6D$-BEMAD Strategy

The glycosylation site of O-GlcNAcylation could be labeled with a nucleophile tag [dithiotreitol (DTT)] by β-elimination followed
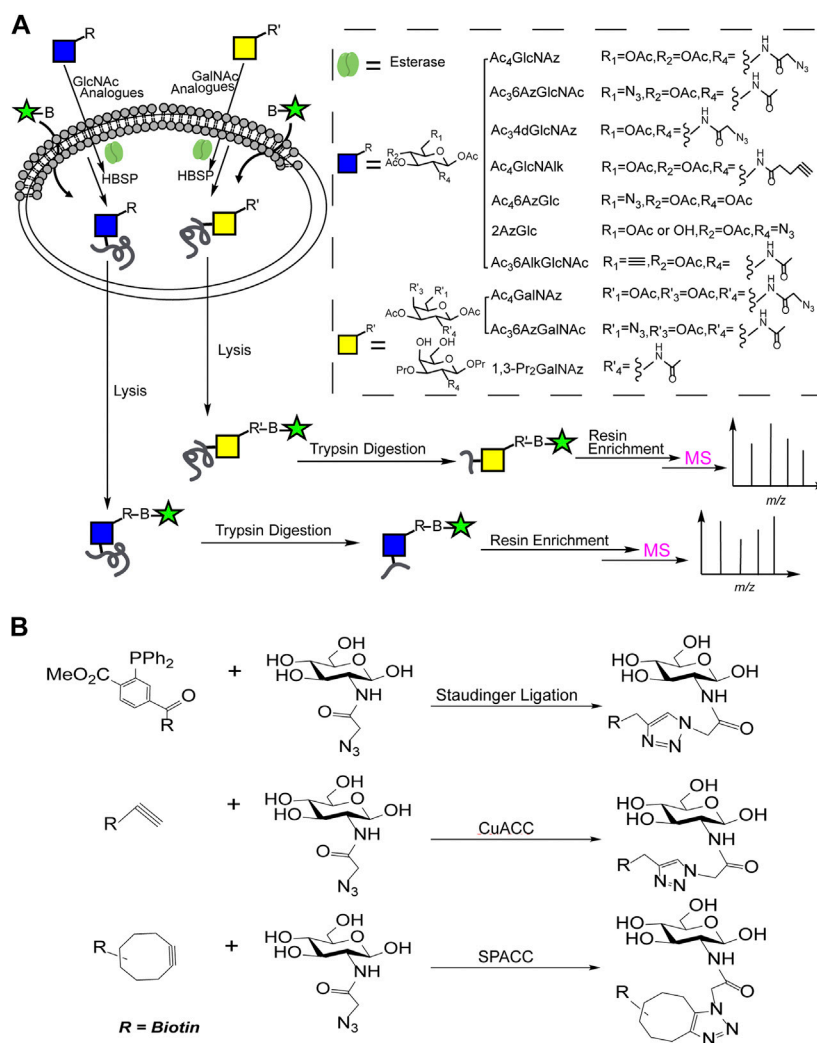
FIGURE 1 | (A) The combination of MOE and solid phase enrichment for MS profiling of O-GlcNAcylated proteins; **(B)** Reactions involved in solid phase enrichment of O-GlcNAcylation.

by Michael addition (BEMAD). As a result, the unstable O-GlcNAc glycosidic bond has been converted to be a stable derivative, enabling characterization of the O-GlcNAcylation by MS (Hédou et al., 2009). However, BEMAD strategy may not be suitable for distinguishing phosphorylation from O-GlcNAcylation.

When involving ^{0}D/^{6}D-DTT in BEMAD strategy, MS-based quali-quantitative characterization of O-GlcNAcylation could be achieved. Two samples, respectively, digested by PNGase F and trypsin digestion were subjected to ^{0}D/^{6}D-BEMAD, as shown in **Supplementary Figure S2**. Then the labeled glycopeptides captured through a mercaptans affinity chromatography and equally mixed were subjected to MS-based quali-quantitative characterization (Vosseller et al., 2005).

To improve detection efficiency and accuracy of O-GlcNAcylation by MS, an enrichment step for

O-GlcNAcylated proteins has been involved (such as lectin, chemoenzyme labeling, etc.) before BEMAD.

MS-Based Quali-Quantitative Characterization of O-GlcNAcylation by Metabolic Labeling of Stable Isotope Labels

By feeding cells with ^{12}C and ^{13}C glucose successively, the O-GlcNAcylated proteins could be labeled through the hexosamine biosynthetic pathway. Then the dynamic changes of O-GlcNAcylated proteins during biological procedure were determined by MS, as shown in **Supplementary Figure S3A** (Wang et al., 2016).

As shown in **Supplementary Figure S3B**, feeding cells in the presence of normal (light) or isotopically enriched (heavy) amino acid could produce normally or isotopically labeled proteins by SILAC (stable isotope labeling with amino acids). After 1:1 mixing,
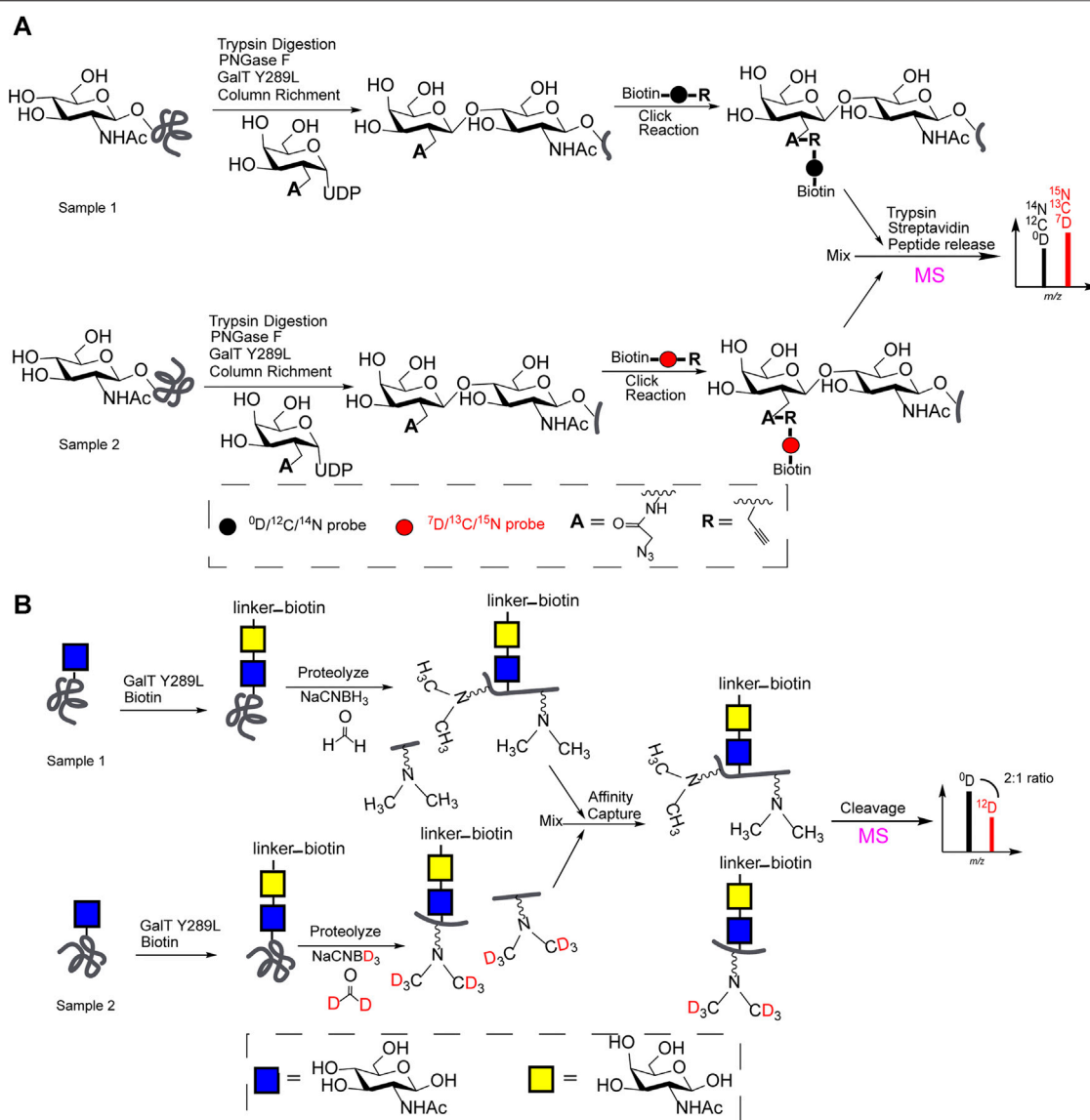
FIGURE 2 | (A) Chemoenzymatic and stable isotope labeling ("light" and "heavy" biotin linker) for MS-based quali-quantitative characterization of O-GlcNAcylation; (B) QUIC-Tag for MS-based quali-quantitative profiling of O-GlcNAcylation.

the mixture subjected to trypsin digestion and enriched by affinity chromatography (antibodies, lectin, etc.) to capture O-GlcNAc modified peptides, were later assigned by MS-based quali-quantitative characterization (Wang et al., 2007).

MS-Based Quali-Quantitative Characterization of O-GlcNAcylation by Chemoenzymatic and Stable Isotope Labeling

As shown in **Figure 2A**, GalT Y289L transfers UDP-galactose analogues with reactive groups (acetylene or azide) to the C4-position of the O-GlcNAc. The "light" (^{0}D, ^{12}C, or ^{14}N probe) and "heavy" (isotope-labeled, ^{7}D, ^{13}C, or ^{15}N probe) biotin linker,

respectively, were used to label each O-GlcNAcylation via biological orthogonal reaction. Then MS-based quali-quantitative characterization of O-GlcNAcylation between two samples was achieved after equally mixing (Qin et al., 2018; Li et al., 2019).

Meanwhile, quantitative isotopic and chemoenzymatic tagging (QUIC-Tag) for MS-based quali-quantitative profiling of O-GlcNAcylation, was illustrated in **Figure 2B**. Generally, samples were enriched by avidin-biotin affinity chromatography after a chemoenzymatic labeling of O-GlcNAc. After a trypsin digestion, two samples (peptides) respectively were treated with formaldehyde/NaCNBH$_3$ or deuterated formaldehyde/NaCNBD$_3$ via reductive amination reaction for subsequent MS profiling. The expression levels of O-GlcNAcylation at proteins involved in the

regulation of transcription has been quantitatively characterized (Khidekel et al., 2007).

CONCLUSION

O-GlcNAcylation plays an important role in plenty of biological activities, abnormal changes of O-GlcNAcylation are closely associated with the development of kinds of diseases. MS with advantages of quali-quantitatively profiling structural details of glycan compositions, glycosidic linkages and glycosylation sites, has accelerated understanding the O-GlcNAcylations.

Due to the low abundance and structural diversity of O-GlcNAc modified proteins, the combination of MOE or chemoenzymatic labeling, isotopic tagging or affinity chromatography enrichment with MS-based quali-quantitative profiling, have played important roles in understanding the biological roles of O-GlcNAcylation. However, some non-specific bonding (lectin), low bonding capacity (antibody) and unspecific labeling (S-glycosylation) occur, as summarized in **Supplementary Tables S1, S2**. Still, attentions should be paid to the development of specific enrichment strategy for selectively capturing the O-GlcNAcylated proteins.

Even enriching the O-GlcNAcylated proteins by specific affinity chromatography, the phosphorylation at peptide would produce the false positive signal, bringing inevitable interference in signal assignation. Efforts should be focused on developing MS-based technique combined with chemical releasing strategy to distinguish O-GlcNAcylations from O-phosphorylation in future.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.737093/full#supplementary-material

REFERENCES

Chalkley, R. J., and Burlingame, A. L. (2001). Identification of GlcNAcylation Sites of Peptides and α-crystallin Using Q-TOF Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* 12, 1106–1113. doi:10.1016/s1044-0305(01)00295-1

Chuh, K. N., Batt, A. R., Zaro, B. W., Darabedian, N., Marotta, N. P., Brennan, C. K., et al. (2017). The New Chemical Reporter 6-Alkynyl-6-Deoxy-GlcNAc Reveals O-GlcNAc Modification of the Apoptotic Caspases that Can Block the Cleavage/activation of Caspase-8. *J. Am. Chem. Soc.* 139, 7872–7885. doi:10.1021/jacs.7b02213

Chuh, K. N., Zaro, B. W., Piller, F., Piller, V., and Pratt, M. R. (2014). Changes in Metabolic Chemical Reporter Structure Yield a Selective Probe of O-GlcNAc Modification. *J. Am. Chem. Soc.* 136, 12283–12295. doi:10.1021/ja504063c

Cieniewski-Bernard, C., Bastide, B., Lefebvre, T., Lemoine, J., Mounier, Y., and Michalski, J.-C. (2004). Identification of O-Linked N-Acetylglucosamine Proteins in Rat Skeletal Muscle Using Two-Dimensional Gel Electrophoresis and Mass Spectrometry. *Mol. Cell Proteomics* 3, 577–585. doi:10.1074/mcp.M400024-MCP200

Darabedian, N., Gao, J., Chuh, K. N., Woo, C. M., and Pratt, M. R. (2018). The Metabolic Chemical Reporter 6-Azido-6-Deoxy-Glucose Further Reveals the Substrate Promiscuity of O-GlcNAc Transferase and Catalyzes the Discovery of Intracellular Protein Modification by O-Glucose. *J. Am. Chem. Soc.* 140, 7092–7100. doi:10.1021/jacs.7b13488

Dias, W. B., Cheung, W. D., Wang, Z., and Hart, G. W. (2009). Regulation of Calcium/calmodulin-dependent Kinase IV by O-GlcNAc Modification. *J. Biol. Chem.* 284, 21327–21337. doi:10.1074/jbc.M109.007310

Guo, J., Zhang, G., Ma, J., Zhao, C., Xue, Q., Wang, J., et al. (2019). Detection and Identification of O-GlcNAc-Modified Proteins Using 6-Azido-6-Deoxy-N-Acetyl-Galactosamine. *Org. Biomol. Chem.* 17, 4326–4334. doi:10.1039/c9ob00516a

Hao, Y., Fan, X., Shi, Y., Zhang, C., Sun, D.-e., Qin, K., et al. (2019). Next-generation Unnatural Monosaccharides Reveal that ESRRB O-GlcNAcylation Regulates Pluripotency of Mouse Embryonic Stem Cells. *Nat. Commun.* 10, 4065. doi:10.1038/s41467-019-11942-y

Hart, G. W., Housley, M. P., and Slawson, C. (2007). Cycling of O-Linked β-N-acetylglucosamine on Nucleocytoplasmic Proteins. *Nature* 446, 1017–1022. doi:10.1038/nature05815

Hayes, B. K., Greis, K. D., and Hart, G. W. (1995). Specific Isolation of O-Linked N-Acetylglucosamine Glycopeptides from Complex Mixtures. *Anal. Biochem.* 228, 115–122. doi:10.1006/abio.1995.1322

Hédou, J., Bastide, B., Page, A., Michalski, J.-C., and Morelle, W. (2009). Mapping of O-Linkedβ-N-Acetylglucosamine Modification Sites in Key Contractile Proteins of Rat Skeletal Muscle. *Proteomics* 9, 2139–2148. doi:10.1002/pmic.200800617

Khidekel, N., Ficarro, S. B., Clark, P. M., Bryan, M. C., Swaney, D. L., Rexach, J. E., et al. (2007). Probing the Dynamics of O-GlcNAc Glycosylation in the Brain Using Quantitative Proteomics. *Nat. Chem. Biol.* 3, 339–348. doi:10.1038/nchembio881

Kupferschmid, M., Aquino-Gil, M. O., Shams-Eldin, H., Schmidt, J., Yamakawa, N., Krzewinski, F., et al. (2017). Identification of O-GlcNAcylated Proteins in Plasmodium Falciparum. *Malar. J.* 16, 485. doi:10.1186/s12936-017-2131-2

Li, J., Li, Z., Duan, X., Qin, K., Dang, L., Sun, S., et al. (2019). An Isotope-Coded Photocleavable Probe for Quantitative Profiling of Protein O-GlcNAcylation. *ACS Chem. Biol.* 14, 4–10. doi:10.1021/acschembio.8b01052

Li, J., Wang, J., Wen, L., Zhu, H., Li, S., Huang, K., et al. (2016). An OGA-Resistant Probe Allows Specific Visualization and Accurate Identification of O-GlcNAc-Modified Proteins in Cells. *ACS Chem. Biol.* 11, 3002–3006. doi:10.1021/acschembio.6b00678

Liu, W., Han, G., Yin, Y., Jiang, S., Yu, G., Yang, Q., et al. (2018). AANL (Agrocybe Aegerita Lectin 2) Is a New Facile Tool to Probe for O-GlcNAcylation. *Glycobiology* 28, 363–373. doi:10.1093/glycob/cwy029

Ma, J., and Hart, G. W. (2017). Analysis of Protein O-GlcNAcylation by Mass Spectrometry. *Curr. Protoc. Protein Sci.* 87, 24. doi:10.110.1610.1002/cpps.2410.1002/cpps.24

Ma, J., Wang, W.-H., Li, Z., Shabanowitz, J., Hunt, D. F., and Hart, G. W. (2019). O-GlcNAc Site Mapping by Using a Combination of Chemoenzymatic

Labeling, Copper-free Click Chemistry, Reductive Cleavage, and Electron-Transfer Dissociation Mass Spectrometry. *Anal. Chem.* 91, 2620–2625. doi:10.1021/acs.analchem.8b05688

Nie, H., Ju, H., Fan, J., Shi, X., Cheng, Y., Cang, X., et al. (2020). O-GlcNAcylation of PGK1 Coordinates Glycolysis and TCA Cycle to Promote Tumor Growth. *Nat. Commun.* 11, 1–14. doi:10.1038/s41467-019-13601-8

Nishikaze, T., Kawabata, S.-i., Iwamoto, S., and Tanaka, K. (2013). Reversible Hydrazide Chemistry-Based Enrichment for O-GlcNAc-Modified Peptides and Glycopeptides Having Non-reducing GlcNAc Residues. *Analyst* 138, 7224–7232. doi:10.1039/c3an00880k

Qin, K., Zhang, H., Zhao, Z., and Chen, X. (2020). Protein S-Glyco-Modification through an Elimination-Addition Mechanism. *J. Am. Chem. Soc.* 142, 9382–9388. doi:10.1021/jacs.0c02110

Qin, K., Zhu, Y., Qin, W., Gao, J., Shao, X., Wang, Y.-l., et al. (2018). Quantitative Profiling of Protein O-GlcNAcylation Sites by an Isotope-Tagged Cleavable Linker. *ACS Chem. Biol.* 13, 1983–1989. doi:10.1021/acschembio.8b00414

Rexach, J. E., Clark, P. M., Mason, D. E., Neve, R. L., Peters, E. C., and Hsieh-Wilson, L. C. (2012). Dynamic O-GlcNAc Modification Regulates CREB-Mediated Gene Expression and Memory Formation. *Nat. Chem. Biol.* 8, 253–261. doi:10.1038/nchembio.770

Sprung, R., Nandi, A., Chen, Y., Kim, S. C., Barma, D., Falck, J. R., et al. (2005). Tagging-via-substrate Strategy for Probing O-GlcNAc Modified Proteins. *J. Proteome Res.* 4, 950–957. doi:10.1021/pr050033j

Szychowski, J., Mahdavi, A., Hodas, J. J. L., Bagert, J. D., Ngo, J. T., Landgraf, P., et al. (2010). Cleavable Biotin Probes for Labeling of Biomolecules via Azide–Alkyne Cycloaddition. *J. Am. Chem. Soc.* 132, 18351–18360. doi:10.1021/ja1083909

Tai, H.-C., Khidekel, N., Ficarro, S. B., Peters, E. C., and Hsieh-Wilson, L. C. (2004). Parallel Identification of O-GlcNAc-Modified Proteins from Cell Lysates. *J. Am. Chem. Soc.* 126, 10500–10501. doi:10.1021/ja047872b

Teo, C. F., Ingale, S., Wolfert, M. A., Elsayed, G. A., Nöt, L. G., Chatham, J. C., et al. (2010). Glycopeptide-specific Monoclonal Antibodies Suggest New Roles for O-GlcNAc. *Nat. Chem. Biol.* 6, 338–343. doi:10.1038/nchembio.338

Tsai, C.-S., Liu, P.-Y., Yen, H.-Y., Hsu, T.-L., and Wong, C.-H. (2010). Development of Trifunctional Probes for Glycoproteomic Analysis. *Chem. Commun.* 46, 5575–5577. doi:10.1039/c0cc00345j

Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W., et al. (2005). Quantitative Analysis of Both Protein Expression and Serine?/?threonine post-translational Modifications through Stable Isotope Labeling with Dithiothreitol. *Proteomics* 5, 388–398. doi:10.1002/pmic.200401066

Wang, X., Yuan, Z.-F., Fan, J., Karch, K. R., Ball, L. E., Denu, J. M., et al. (2016). A Novel Quantitative Mass Spectrometry Platform for Determining Protein O-GlcNAcylation Dynamics. *Mol. Cell Proteomics* 15, 2462–2475. doi:10.1074/mcp.O115.049627

Wang, Z., Pandey, A., and Hart, G. W. (2007). Dynamic Interplay between O-Linked N-Acetylglucosaminylation and Glycogen Synthase Kinase-3-dependent Phosphorylation. *Mol. Cell Proteomics* 6, 1365–1379. doi:10.1074/mcp.M600453-MCP200

Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002). Mapping Sites of O-GlcNAc Modification Using Affinity Tags for Serine and Threonine post-translational Modifications. *Mol. Cell Proteomics* 1, 791–804. doi:10.1074/mcp.M200048-MCP200

Yang, X., Ongusaha, P. P., Miles, P. D., Havstad, J. C., Zhang, F., So, W. V., et al. (2008). Phosphoinositide Signalling Links O-GlcNAc Transferase to Insulin Resistance. *Nature* 451, 964–969. doi:10.1038/nature06668

Yang, Y.-Y., Grammel, M., Raghavan, A. S., Charron, G., and Hang, H. C. (2010). Comparative Analysis of Cleavable Azobenzene-Based Affinity Tags for Bioorthogonal Chemical Proteomics. *Chem. Biol.* 17, 1212–1222. doi:10.1016/j.chembiol.2010.09.012

Yuzwa, S. A., and Vocadlo, D. J. (2014). O-GlcNAc and Neurodegeneration: Biochemical Mechanisms and Potential Roles in Alzheimer's Disease and beyond. *Chem. Soc. Rev.* 43, 6839–6858. doi:10.1039/C4CS00038B

Zaro, B. W., Yang, Y.-Y., Hang, H. C., and Pratt, M. R. (2011). Chemical Reporters for Fluorescent Detection and Identification of O-GlcNAc-Modified Proteins Reveal Glycosylation of the Ubiquitin Ligase NEDD4-1. *Proc. Natl. Acad. Sci.* 108, 8146–8151. doi:10.1073/pnas.1102458108

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership