analytical chemistry

pubs.acs.org/ac Article

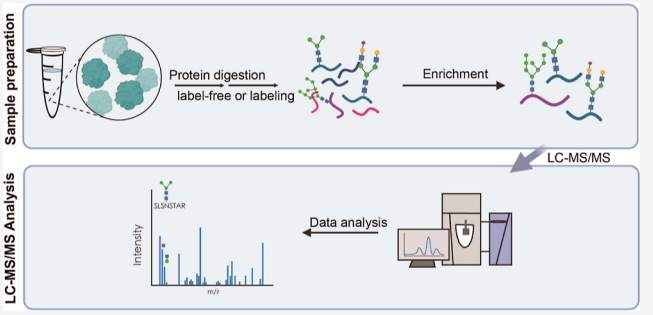
Improving Glycoproteomic Analysis Workflow by Systematic Evaluation of Glycopeptide Enrichment, Quantification, Mass Spectrometry Approach, and Data Analysis Strategies

Zhenyu Sun, T. Mamie Lih, Jongmin Woo, Liyuan Jiao, Yingwei Hu, Yuefan Wang, Hongyi Liu, and Hui Zhang*

Cite This: *Anal. Chem.* 2024, 96, 20481−20490 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Glycosylation is one of the most prevalent and crucial protein modifications. Quantitative site-specific characterization of glycosylation usually requires sophisticated intact glycopeptide analysis using glycoproteomics. Recent efforts have focused on the interrogation of intact glycopeptide analyses using tandem mass spectrometry. However, a systematic evaluation of the quantitative glycoproteomic workflow is still lacking. This study compared different strategies for glycopeptide enrichment alongside glycopeptide quantitation, as well as mass spectrometry and data analysis strategies, providing a comprehensive assessment of their efficacy. The ZIC-HILIC enrichment method demonstrated superior performance, representing a 26% improvement in identified glycopeptiudes compared to the MAX enrichment method. Quantification using TMT provided high precision and throughput with an average CV of 8%. Through systematic evaluation, this study established that the ZIC-HILIC enrichment method, quantification with TMT, and collision energies of 25, 35, and 45 using tandem mass spectrometry are the optimal workflow for higher-energy collisional dissociation (HCD) fragmentation, significantly enhancing the analysis of intact glycopeptides. Precise energy adjustment is crucial for the identification of certain glycans. Intact glycopeptides were analyzed using different software tools to investigate the identification and quantification of glycopeptides. By applying optimal settings, 5514 unique intact glycopeptides were in luminal and basal patient-derived xenograft (PDX) characterized models, highlighting distinct glycosylation profiles that may influence tumor behavior. This study offers a systematic approach to evaluate glycoproteomic analysis workflow.



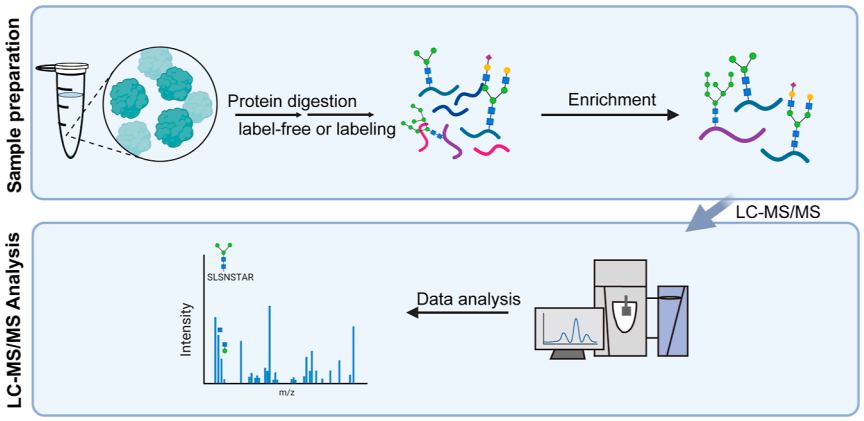
INTRODUCTION

Glycosylation is one of the most prevalent and crucial protein modifications and plays pivotal roles in various biological processes, including protein folding, trafficking, and stability.[1,2] The synthesis of glycans is template-independent, resulting in micro- and macroheterogeneity, posing significant challenges to glycosylation study.[3,4] Mass spectrometry (MS) has emerged as a powerful tool for protein glycosylation analysis of intact glycopeptides for site-specific characterization of micro- and macroheterogeneity.[5,6] Recently, a range of MS-based methods for comprehensive intact N-glycopeptide analysis have been developed.[7,8] MS analysis can simultaneously provide information about glycosites and the types of glycans for intact glycopeptides. Additionally, quantitative analysis of intact glycopeptides can be performed, providing protein glycosylation information. Currently, the workflow for analyzing intact glycopeptides involves enzymatic digestion of proteins, followed by enrichment of glycopeptides and subsequent analysis using liquid chromatography−tandem mass spectrometry (LC−MS/MS) analysis.[9] In peptide mixtures resulting from trypsin or other proteolytic enzyme treatments, intact glycopeptides typically represent in low abundance (approximately 1% of total peptide content).[10] Consequently, the enrichment of intact glycopeptides before MS analysis is crucial for complex biological or clinical samples. This step enhances the signal intensity of intact glycopeptides and minimizes interference from nonglycosylated peptides from the complex samples. Various enrichment strategies for intact glycopeptides have been developed, significantly improving their coverage in different types of samples.[11] Hydrophilic interaction liquid chromatography (HILIC) and mixed-mode strong anion exchange (MAX)

Received: August 21, 2024
Revised: November 25, 2024
Accepted: December 3, 2024
Published: December 16, 2024

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

Scheme 1. General Workflow of Glycoproteomics Analysis



enrichment strategies were the most used in glycoproteomics, which can separate intact glycopeptides via the hydrophilic properties of glycans.[12,13] There are several HILIC materials, and ZIC-HILIC, based on zwitterionic stationary phases, is a common type of HILIC. ZIC-HILIC can selectively enrich glycopeptides based on their hydrophilicity and charge state, offering higher enrichment specificity and sensitivity compared to traditional HILIC methods.[14] Currently, a systematic evaluation of MAX and ZIC-HILIC enrichment methods is still lacking.

Even though excellent enrichment methods can mitigate the impact of nonglycosylated peptides on intact glycopeptide analysis, the microheterogeneity of intact glycopeptides still makes their quantitative analysis challenging. In the era of precision medicine, the demand for MS analysis of large-scale clinical samples is steadily increasing.[15] This emphasizes the need to develop robust and highly reproducible quantitative methods for analyzing large cohorts of clinical samples. The MS-based quantification glycoproteomic technique consists of label-based and label-free methods. Sample preparation for the label-free method is more convenient and flexible. However, Intact glycopeptides with certain glycans exhibit low abundance, making their identification and quantification challenging. Label-free and labeling-based approaches were used for glycoproteomic quantification.[16] Label-based methods, such as the tandem mass tagging (TMT) strategy, allow multiple sample analyses simultaneously, which can be used for investigating multiple samples at the same time. In intact glycopeptides, glycans tend to dissociate more readily due to glycosidic bonds absorbing most of the collisional energy, thereby restricting the production of reporter ions.[17] Nevertheless, systematically comparing the quantitative abilities of labeled and label-free methods for intact glycopeptides serves as a fundamental cornerstone for subsequent large-scale glycoproteomic investigations based on clinical samples.

Following sample preparation, analyzing glycopeptides using tandem MS to generate effective fragments for successful identification of peptides and glycans remains a significant challenge. Unlike many other protein modifications that result in a fixed mass shift, glycosylation is more complex due to the varied composition of glycans.[18] Optimal MS/MS acquisition can generate comprehensive fragments, both the glycan and the peptide, for each intact glycopeptide, where collision energy plays a critical role that can significantly influence the identification of glycan types due to variations in oxonium ion fragmentation.[19]

The combination of various strategies for glycoproteomic use using glycopeptide enrichment, mass spectrometry techniques, quantification, and data analysis strategies has not been systematically evaluated in previous research. In this study, we conducted a systematic investigation into the glycoproteomic workflow. First, label-free and TMT labeling strategies were utilized for intact N-glycopeptide quantification. A total of 2924 unique intact N-glycopeptides can be quantified after being labeled using the TMT approach. Compared to the label-free method, TMT labeling is used for its high precision with the coefficient of variation (CV) averaging 8% and multiplex analysis capabilities. Subsequently, we compared the capacity of ZIC-HILIC and MAX enrichment methods, revealing that the ZIC-HILIC method demonstrated superior performance in enriching intact N-glycopeptides, which can identify 3004 GPSMs from a single injection of a patient-derived xenograft (PDX) sample. Furthermore, we systematically evaluated different higher-energy collisional dissociation (HCD) collision energies, determining that the stepped collision energy HCD (sceHCD) settings of 25, 35, and 45 were optimal for both the identification and quantification of intact N-glycopeptides. Of note, our findings indicated a relationship between the dissociation of glycans and applied collision energy, emphasizing the necessity of tailored fragmentation energies for comprehensive glycoform analysis. TMTpro-labeled intact glycopeptides were analyzed using three different software tools, pGlyco 3.0, MS-PyCloud, and MSFragger-Glyco, to investigate the identification and quantification of different glycoproteomic analysis tools for glycopeptide identification and quantification. We applied the aforementioned optimal settings to breast cancer PDX samples, and a total of 5514 unique intact N-glycopeptides were identified from luminal and basal subtypes of the PDX samples, revealing distinct glycosylation profiles. These unique profiles may play a significant role in influencing tumor behavior and patient response to therapy. The significance of this study lies in providing effective technical routes and methodological strategies for optimizing the MS analysis of glycosylation, laying an important groundwork for further exploration of the value of glycoproteomics in clinical applications.

EXPERIMENTAL SECTION

Materials and Chemicals. Trypsin was from Promega (Madison, WI); Lys-C endopeptidase was from Wako Chemicals; Sep-Pak C18 1 cm^3 Vac Cartridge was from

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

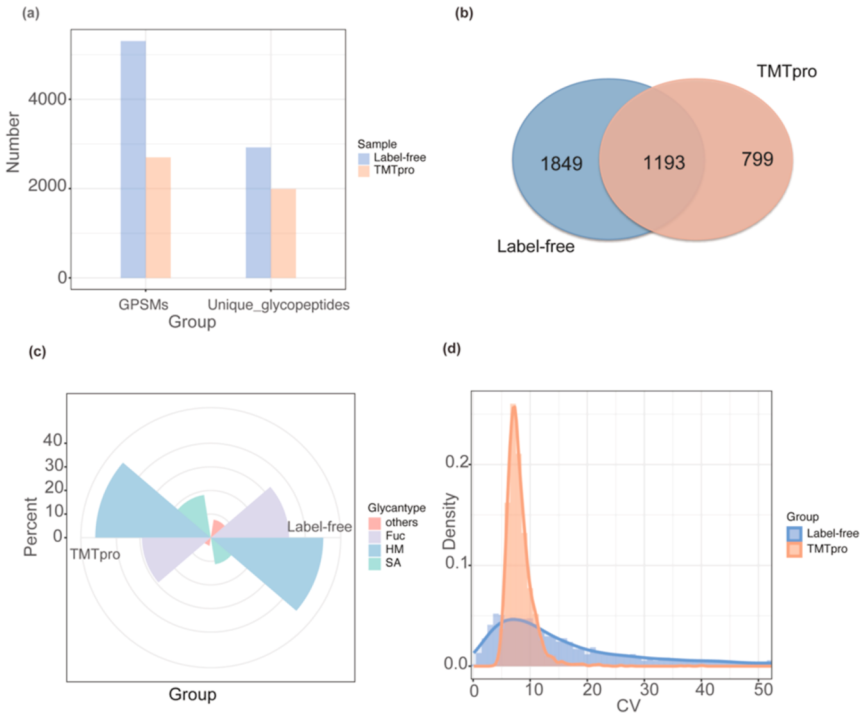


Figure 1. Comparison between label-free and TMTpro labeling of intact N-glycopeptides. (a) Identification number of GPSMs and unique intact N-glycopeptides from label-free and TMTpro labeled glycopeptides. (b) Overlap between label-free and TMTpro labeling strategies. (c) Distribution of identified N-glycan types from unlabeled and TMTpro labeling strategies. (d) CV distribution between label-free and TMTpro labeling strategies.

Waters (Milford, MA); ZIC-HILIC columns were from Sigma-Aldrich (St. Louis, MO); MAX columns were from Waters (Milford, MA); other chemicals such as urea, acetonitrile (ACN), trifluoroacetic acid (TFA), iodoacetamide (IAA), and triethylammonium acetate were purchased from Sigma-Aldrich (St. Louis, MO).

Protein Extraction and Digestion. The PDX tumors used in this study were from previously established basal (WHIM2) and luminal (WHIM16) breast cancers. For protein extraction, 400 μL of 8 M urea lysis buffer were added to 100 mg of tissue for protein extract. The concentrations of proteins were determined using the BCA assay. Samples were incubated in 5 mM dithiothreitol (DTT) for 1 h at room temperature and then alkylated with 10 mM iodoacetamide (IAA) for 45 min at room temperature in the dark. Proteins were then diluted to 2 M urea with gentle shaking for 2 h by Lys-C digestion at an enzyme-to-protein ratio of 1:50 (w/w). And then trypsin was added at an enzyme-to-protein ratio of 1:50 and digested overnight at room temperature, ensuring complete digestion for accurate glycopeptide analysis. Digestion was terminated using 50% formic acid (FA) by adjusting solution pH to about 2−3.[20]

RESULTS AND DISCUSSION

Glycoproteomic workflow for analyzing intact glycopeptides involves enzymatic digestion of proteins, followed by enrichment of glycopeptides and subsequent data acquisition using LC−MS/MS and data analysis (Scheme 1). In this work, we will evaluate each step for setting up the optimal workflow for clinical samples.

Quantification Strategies for Glycopeptide Analysis. For quantification strategies, both label-free and label-based methods can be used. Isobaric labeling with TMT enables proteome-wide relative quantification for multiple samples simultaneously in one LC−MS/MS analysis,[22] which could reduce MS analysis time. 18plex TMTpro labeling was included in our workflow; a total of 18 individual samples can be labeled and analyzed in one LC−MS/MS analysis. In this study, the label-free strategy, we identified 5253 glycopeptide spectral matches (GPSMs) and 3042 unique glycopeptides from one LC−MS/MS injection. Following TMT labeling, 2702 GPSMs and 1992 unique glycopeptides were identified (Figure 1a). The overlap between label-free and TMT labeling glycopeptides is 193 (Figure 1b). We also calculated TMT labeling efficiency and achieved an efficiency of 99%.

We further compared the N-glycan types between TMT and the label-free methods. The number of fucose and high-mannose glycans was similar, but 7% more sialic acids were identified after TMTpro labeling, while glycans without sialic acids were increased in the label-free method (Figure 1c). TMTpro labeling converts the primary amines of the lysine residues into moieties containing dimethyl piperidine groups, increasing proton affinity and the number of possible charge states during ionization. Consequently, more sialic acids could be detected compared to unlabeled glycopeptides as sialic acids contain carboxyl groups, which have low intensity in MS. Furthermore, among the identified sialylated glycopeptides, more than 7% of N-glycans contained two sialic acids, and more than 1.6% of N-glycans contained three sialic acids (Figure S1). Moreover, we compared the coefficient of variation (CV) distributions between label-free and TMTpro to assess the quantitative precision of each method (Figure 1d). We observed that the average CV was 21% for the label-free method and 8% for the TMTpro labeling method, indicating higher quantitative reproducibility and reduced

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

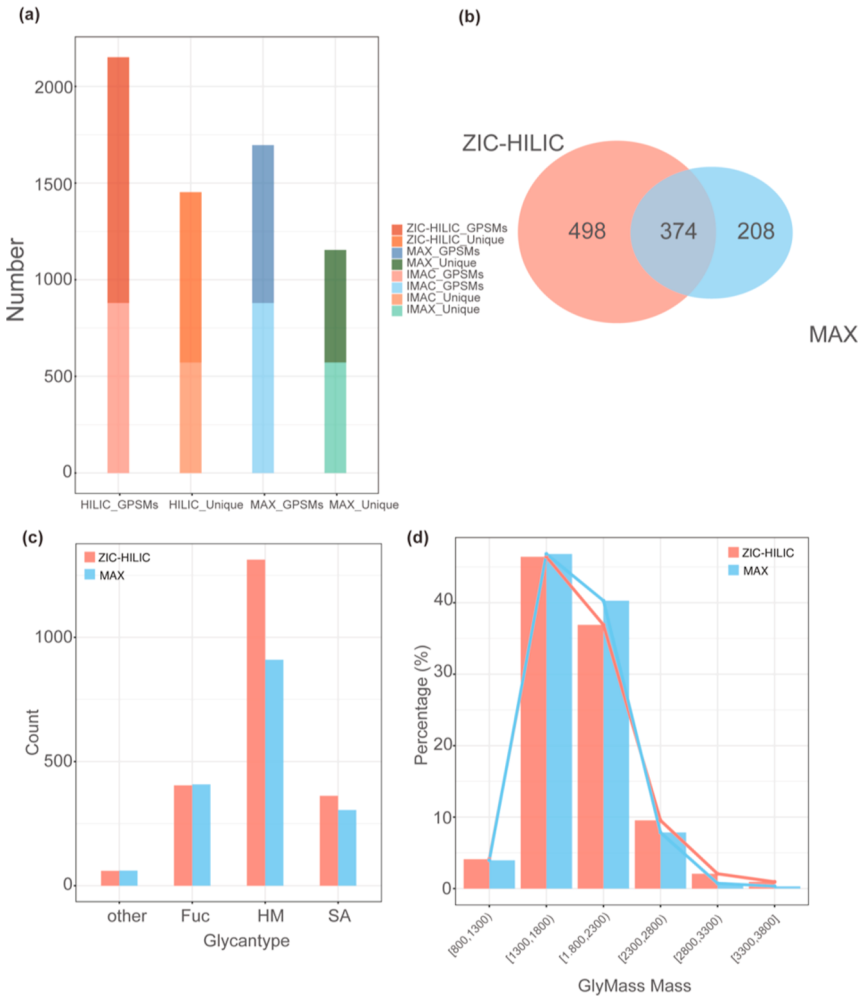


Figure 2. Characteristics of glycopeptides identified by ZIC-HILIC and MAX methods. (a) Numbers of glycopeptide identifications using ZIC-HILIC or MAX. (b) Overlapped and uniquely identified intact N-glycopeptides using ZIC-HILIC and MAX. (c) Distribution of N-glycan types identified using ZIC-HILIC and MAX methods. (d) Distribution of N-glycan mass weight using ZIC-HILIC and MAX methods.

variability with TMTpro labeling. In addition, the TMTpro labeling approach maintained high reproducibility, with the CV for quantified glycopeptides remaining below 15% in more than 98% of glycopeptides. Thus, we selected the TMTpro labeling strategy to decrease the variations among replicates. In addition, stable isotope labeling boosted the intensity of intact N-glycopeptides with a relatively lower abundance in the samples, such as those containing sialic acids.

Enrichment Strategies for Intact Glycopeptides. Hydrophilic interaction liquid chromatography (HILIC) and mixed-mode strong anion exchange (MAX) enrichment strategies were the most used in glycoproteomics, which can separate intact glycopeptides via the hydrophilic properties of glycans.[15,23] ZIC-HILIC, based on zwitterionic stationary phases, is a common type of HILIC. ZIC-HILIC can selectively enrich glycopeptides based on their hydrophilicity and charge state, offering higher enrichment specificity and sensitivity compared to traditional HILIC methods.[24,25] In our previous studies, we employed a sequential enrichment method to simultaneously analyze glycosylation and phosphorylation modifications.[15,26,27] Peptides were labeled with TMTpro. After being labeled, phosphorylation enrichment using IMAC was employed at first since our previous research has already optimized the workflow.[28] The flow-through peptides that were enriched by phosphorylation enrichment were used for glycopeptide enrichment. The ZIC-HILIC method showed superior enrichment efficacy for intact glycopeptide identifications compared to the MAX approach, identifying 880, 1271, and 816 GPSMs through IMAC, ZIC-HILIC, and MAX enrichment methods, respectively. These GPSMs represented 572 (IMAC), 881 (ZIC-HILIC), and 582 (MAX) unique intact glycopeptides that were identified, respectively (Figure 2a). From these, a total of 2151 GPSMs were identified through a combination of IMAC with ZIC-HILIC enrichment from flow-through enrichment, while 1696 GPSMs were identified through a combination of IMAC with MAX. The enrichment specificity was determined by the ratio of oxonium ion-containing MS^2 spectra to total MS^2 spectra. The ratios were 45% for HILIC enrichment and 6% for MAX enrichment. We analyzed the overlap of unique intact glycopeptides identified by the ZIC-HILIC and MAX enrichment methods to assess their comparative performance. The overlapping identification between these two methods was only 374 unique glycopeptides, which indicated that more unique glycopeptides were identified by using the ZIC-HILIC method (Figure 2b). In terms of glycan types, the ZIC-HILIC column preferred to enrich high-mannose N-glycans and sialic acid-containing glycans (Figure 2c). Both methods showed

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

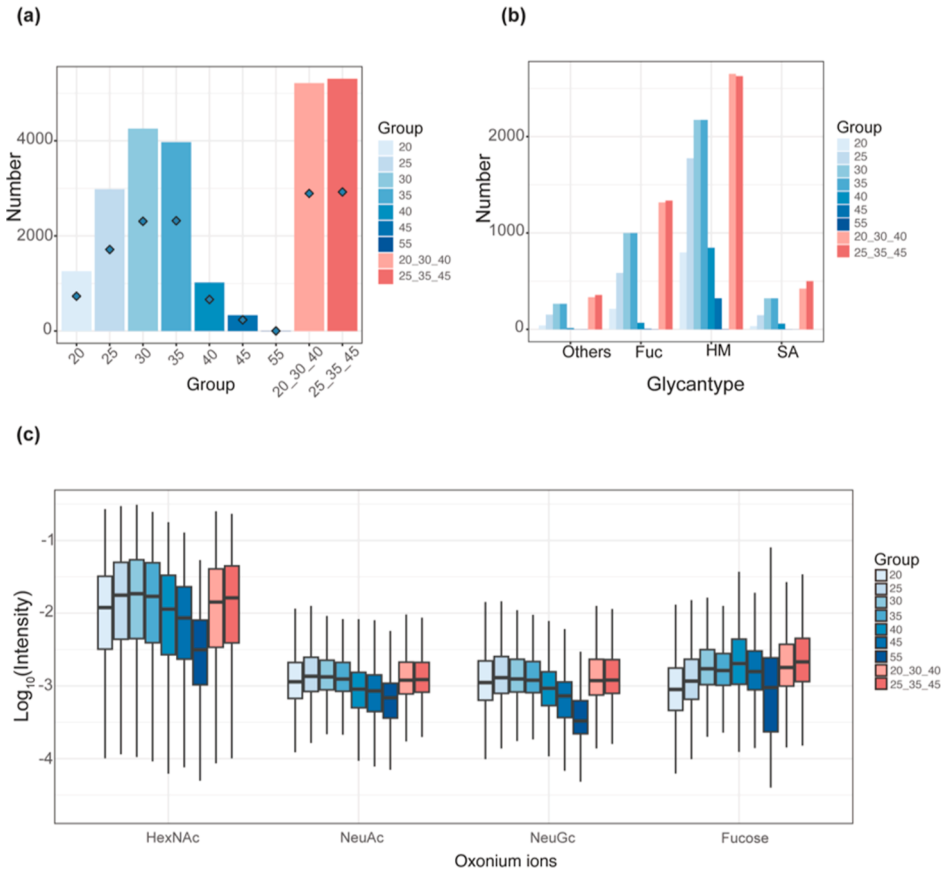


Figure 3. Characterization of label-free intact N-glycopeptides using different collision energy methods. (a) Number of GPSMs and unique intact N-glycopeptides identified from different collision energies. (b) Glycan type distribution from different collision energy methods; "SA" represents sialic acids, "Fuc" represents fucose, and "HM" represents high mannose. (c) MS^2 intensity of oxonium ions from different collision energy methods.

similar efficiency for the enrichment of fucose and other N-glycan types, consistent with the previous study.[12] Furthermore, the MAX method identified 10% more N-glycans than the ZIC-HILIC method in the mass range between 1800 and 2300, while the ZIC-HILIC method identified 4% more N-glycans with mass weight over 2300, which indicated that ZIC-HILIC preferred to enrich glycans with larger mass weight (Figure 2d). An analysis of direct glycopeptide enrichment, without prior phosphopeptide enrichment, yielded consistent findings, demonstrating that the ZIC-HILIC strategy outperformed the MAX enrichment approach (Figure S2). Therefore, we opted to utilize the ZIC-HILIC strategy for the enrichment and analysis of glycopeptides based on its demonstrated superiority in identifying a wider range of N-glycan types.

Optimal Dissociation Conditions for Enhancing Identification of Various Glycan Types in Intact Glycopeptide Analysis. To characterize the site-specific N-glycosylation of proteins from complex samples, it is critical to select an appropriate dissociation method. HCD has been used with success in large-scale glycoproteomic experiments.[29] In this study, we created nine methods to systematically evaluate different HCD collision energies for their performance in characterizing intact N-glycopeptides (Table S1). The average number of GPSMs from each method is summarized in Figure 3a. Stepped collision energy HCD (sceHCD) methods clearly outperform single HCD methods since a much higher number of intact N-glycopeptides were identified using sceHCD (Figure 1a), especially when using sceHCD of 25, 35, and 45. A total of 5305 GPSMs and 2924 unique N-glycopeptides were identified using sceHCD of 25, 35, and 45. Therefore, sceHCD values of 25, 35, and 45 showed the optimal dissociation condition, allowing the identification of various glycan types, thereby improving the reliability of glycopeptide profiling. Moreover, as shown in Figure 3b, high-mannose N-glycans emerged as the most prevalent type across various collision energies. Intriguingly, a direct correlation was observed between increasing collision energy and the identification of sialic acids, with an energy setting of 25_35_45 being optimal for their identification. Raising the collision energy from 20 to 35 resulted in an increase in the proportion of sialic acid from 3.15% to 8.51%, showing a linear relationship between the collision energy and sialic acid fragmentation in this range. However, at the highest energy level of 45, only a small number of sialic acids were detected (0.3%). The identification of intact N-glycopeptides is influenced by both the peptide sequence and the glycan moiety.[19] We selected a peptide sequence, LLNINPNK, of lysosome-associated membrane glycoprotein 1 (LAPM1) as an example to further demonstrate the impact of collision energy on different glycan types (Figure S3). With an energy setting of 35, a total of 12 sialic acids were identified, which was the highest number among the nine methods. By examining the intensities of the representative oxonium ions from all the MS^2 spectra, we found that the intensities of the NeuAc and NeuGc ions were correlated with the identification of sialic acid

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

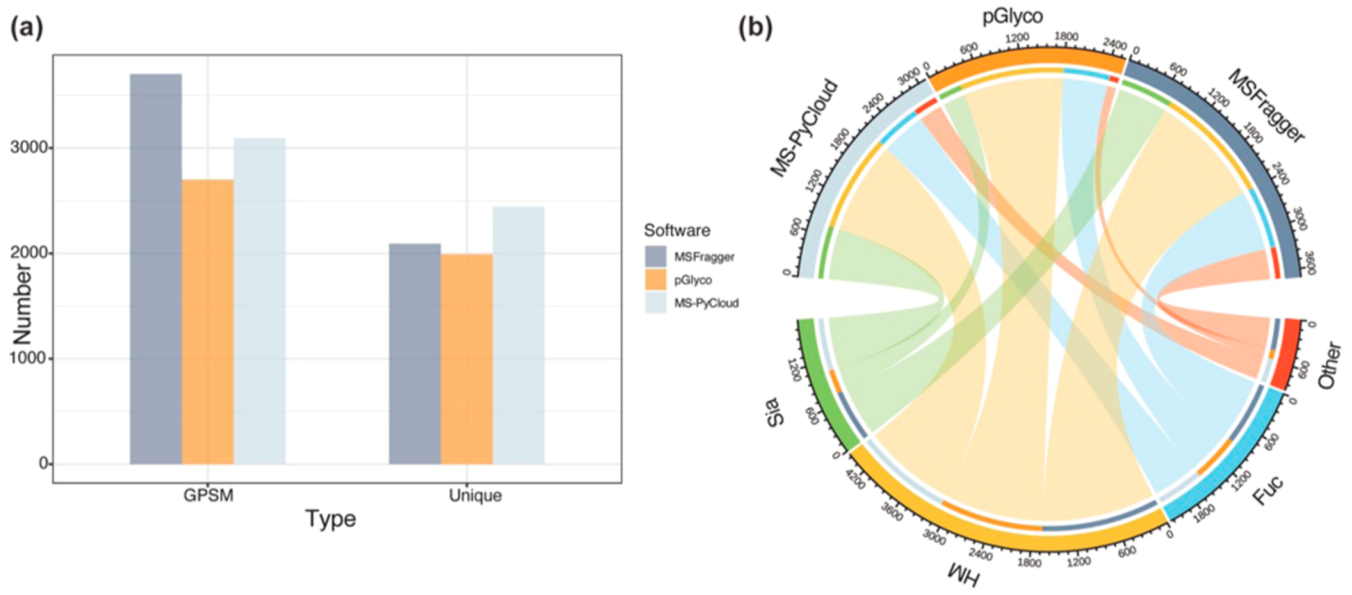


Figure 4. Characterization of TMTpro-labeled intact N-glycopeptides using different software tools. (a) Number of GPSMs and unique intact N-glycopeptides identified from different software tools. (b) Glycan type distribution from different software tools; "SA" represents sialic acids, "Fuc" represents fucose, and "HM" represents high mannose.

(Figure 3c). The intensities of these oxonium ions progressively increased with the rise in normalized collision energy (NCE) levels. When the collision energy exceeds 40, the intensity of HexNAc, NeuAc, and NeuGc significantly decreased. However, the oxonium ions of fucose decreased when the energy was increased to 45. Under collision energy ranging from 20 to 35, the ability to detect fucose remains consistent. Increasing the collision energy to 40 resulted in a significant reduction in the number of fucose identifications, although the intensity of related oxonium ion remains notably high. The results show that glycan dissociation is sensitive to the collision energies used, suggesting that higher energies may cause excessive fragmentation or reduce their detection. These findings highlighted the necessity of selecting specific fragmentation energies for targeted glycoform analysis. In summary, precise energy adjustment was crucial for accurate identification of glycans, particularly for sialic acids. Setting appropriate collision energy would enhance the overall efficacy and specificity of glycoproteomic investigations. Since TMT labeling needs additional collision energies to form reporter ions, we also set different collision energy conditions for MS analysis (Table S2).[16] The sceHCD setting at 25, 35, and 45 was still the best condition for the TMT-labeled intact N-glycopeptide analysis (Figure S4). Therefore, the sceHCD setting of 25, 35, and 45 was selected for the later analysis.

Evaluation of Software Tools for Intact Glycopeptide Analysis. Compared with proteomics analysis, intact glycopeptide data analysis presents significant challenges. An intact glycopeptide comprises both a glycan and a peptide backbone, leading to the generation of multiple ion types during mass spectrometry fragmentation. This substantially increases the complexity and information density of the spectra. Unlike other protein PTMs, glycosylation does not add a fixed mass. With over 1000 glycan structures reported in glycan databases,[30] the diversity of glycans adds complexity to intact glycopeptide analysis.

Recently, advancements have been made in MS data analysis for intact glycopeptide analysis. Three different software tools, MS-PyCloud, pGlyco 3.0, and MSFragger-Glyco, were used to analyze ZIC-HILIC-enriched and TMTpro-labeled intact glycopeptides in this study.[31−33] The analysis identified 2701, 3095, and 3704 GPSMs using pGlyco 3.0, MS-PyCloud, and MSFragger-Glyco, respectively. The number of unique intact glycopeptides identified was 1992 for pGlyco 3.0, 2444 for MS-PyCloud, and 2093 for MSFragger-Glyco (Figure 4a). The number of identified glycoproteins, glycosites, and site-specific glycans is summarized in Figure S5. A total of 403, 560, and 432 glycoproteins were identified by using pGlyco 3.0, MS-PyCloud, and MSFragger-Glyco, respectively. For N-glycans, 119, 262, and 291 were identified using pGlyco 3.0, MS-PyCloud, and MSFragger, respectively. Additionally, 617, 798, and 705 N-glycosites were identified with pGlyco 3.0, MS-PyCloud, and MSFragger-Glyco, respectively. Across these three software tools, a total of 593 glycans, 1163 peptide backbones, and 626 proteins were identified. All three analyses used the same N-glycan database, and the glycan types identified are summarized in Figure 4b. High-mannose glycans were the most prevalent in all three data sets. Notably, 23.9% of the intact glycopeptides identified by MS-PyCloud contained sialic acids, compared to 12.8% and 19.6% in the results from pGlyco 3.0 and MSFragger-Glyco, respectively. We also summarized the peptides and proteins identified by the three software tools, which illustrates the distribution of glycans, peptides, and proteins across different software combinations, highlighting both overlaps and unique identifications in each (Figure S6). Each software tool identified unique components, with the overlap between MS-PyCloud and MSFragger-Glyco being the most; 181 glycans, 501 peptides, and 349 proteins were commonly identified by both tools. Each software provides reliable results, and users can select the most suitable software based on their sample type and specific needs for glycans for data analysis. pGlyco 3.0 was selected for further data analysis in this study.

Investigating Protein Glycosylation in Luminal and Basal Breast Cancer Subtypes Using PDX. Glycosylation of cell surface proteins plays an important role in the regulation

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

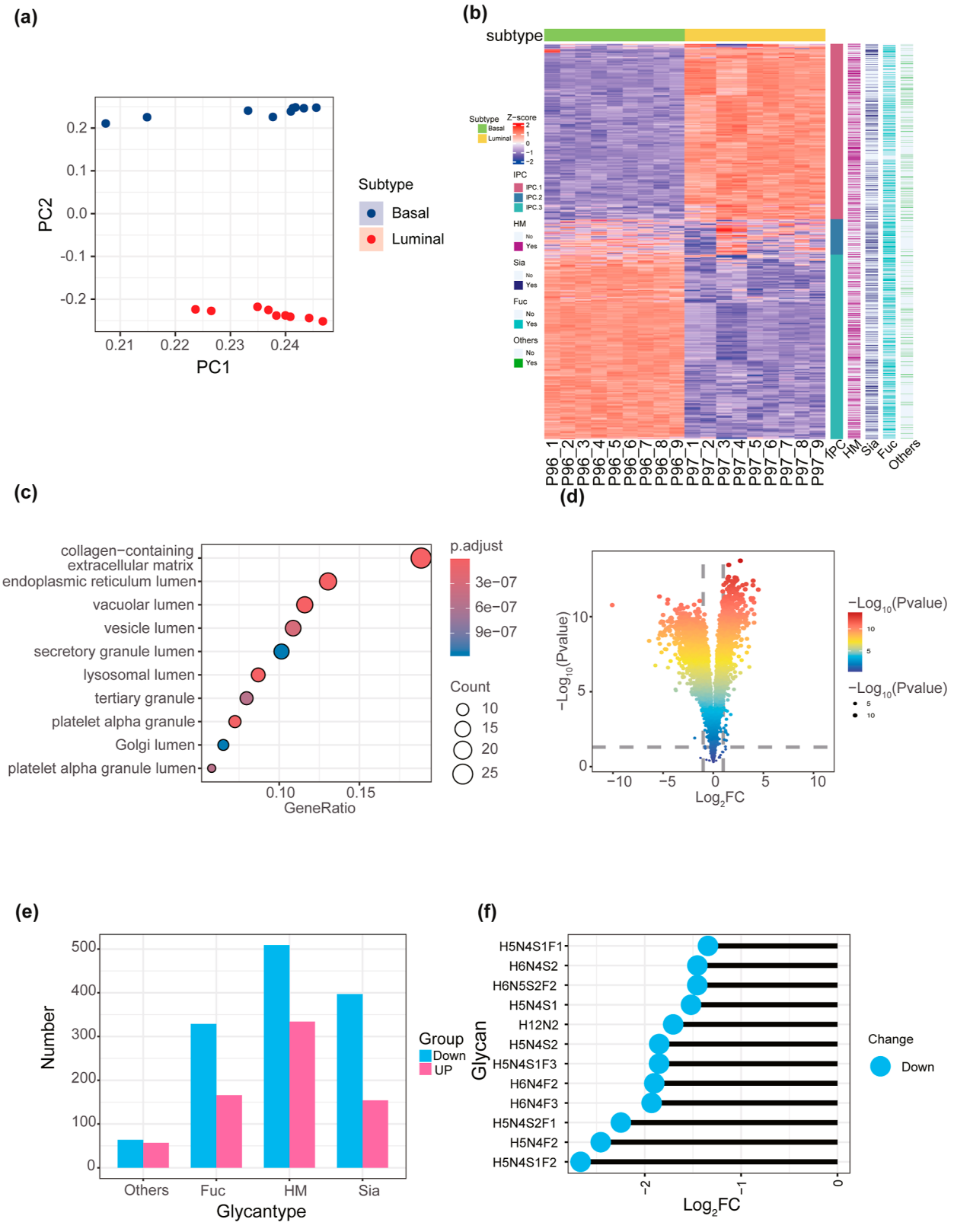


Figure 5. Analysis of altered N-glycosylation in luminal and basal breast cancer subtypes of PDX. (a) Principal component analysis between luminal and basal breast cancer subtypes; (b) heatmap shows the intact N-glycopeptide abundance and N-glycans between luminal and basal breast cancer subtypes; (c) GO analysis of identified intact N-glycopeptides in terms of cellular component; (d) differential analysis between luminal and basal breast cancer subtypes. Significantly altered intact glycopeptides were defined as >2-fold changes with a p-value < 0.05. (e) Distribution of altered N-glycans between luminal and basal breast cancer subtypes; (f) distribution of downregulated intact N-glycopeptides of EpCAM in the luminal subtype compared to the basal subtype.

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490

of apoptosis.[34] Aberrant glycosylation remodeling and metabolism are associated with the epithelial−mesenchymal transition (EMT) and metastasis in breast cancer.[35] The epithelium of the lactiferous ducts in the breast is composed of luminal epithelial cells and underlying basal myoepithelial cells.[36] We applied the optimal method to investigate the differences in N-glycosylation of luminal and basal subtypes using PDX mouse models for better understanding the function of N-glycosylation in breast cancer.

Proteins extracted from the PDX samples were digested to peptides, followed by TMTpro labeling and offline high-pH research phase separation to separate peptides into 12 fractions to increase the depth of identification. Each fraction underwent phosphorylation enrichment prior to the enrichment of intact N-glycopeptides using the follow-through as aforementioned.

A total of 5514 unique intact N-glycopeptides (FDR < 0.01), mapping to 686 glycoproteins, were quantified from basal and luminal samples. Principal component analysis (PCA) showed a clear separation between luminal and basal subtype samples based on the abundances of intact glycopeptides (Figure 5a). There were notable differences in the expression of intact N-glycopeptides between the two subtypes (Figure 5b). The majority of the intact N-glycopeptides were from the proteins in the extracellular matrix, cell surface, lysosome, and plasma membrane regions (Figure 5c). Differential analysis between basal and luminal subtypes revealed a total of 711 intact glycopeptides mapping to 179 proteins with significant upregulation of at least 2-fold changes (p-value < 0.05), and 1299 intact glycopeptides mapping to 242 proteins were downregulated more than 2-fold changes (p-value < 0.05) in the luminal subtype compared to the basal subtype (Figure 5d and Table S3). The differential expression of intact N-glycopeptides was statistically analyzed by glycan types, revealing a marked decrease in the expression of fucose, high mannose, and sialic acids in luminal samples (Figure 5e). EpCAM (CD326) is involved in various functions related to cell adhesion and proliferation and is notably over expressed in primary breast carcinomas (PBCs).[37] A total of 12 intact N-glycopeptides of EpCAM were found with reduced expression in luminal samples. Among these 12 intact N-glycopeptides, 4 were exclusively fucosylated, while the remainder were both fucosylated and sialylated (Figure 5f). The decreased expression of sialylated glycopeptides in luminal samples may align with findings reported in previous studies.[38,39] These results validated the successful application of our optimized workflow for intact N-glycopeptide analysis, demonstrating its potential for large-scale clinical sample analysis. Taken together, our established workflow not only enhanced our understanding of N-glycosylation differences between breast cancer subtypes but also provided strong backing for future research and treatment approaches.

CONCLUSIONS

In conclusion, we selected the TMTpro labeling strategy for quantification due to its high accuracy and high throughput, with the CV averaging 8% and the ability to simultaneously quantify 18 samples in a single run. The ZIC-HILIC enrichment method is superior in terms of identification capacity and unbiased detection of N-glycan types. A total of 2151 GPSMs were identified, representing a 26% improvement compared to the MAX method. We systematically optimized the HCD fragmentation process to improve the identification of intact N-glycopeptides, focusing on glycoform identification and the identification capacity. Precise energy adjustment is crucial for the accurate identification of the glycans. Our results indicate that glycan dissociation is highly sensitive to collision energy, with higher energies potentially causing excessive fragmentation or reducing detection. The optimal HCD collision energy was set as stepping energies of 25, 35, and 45 for subsequent analyses. By comparing label-free and TMTpro labeling quantitative methods, we found that the TMTpro labeling quantification strategy had a higher reproducibility. TMTpro-labeled intact glycopeptides were analyzed using three different software tools, pGlyco 3.0, MS-PyCloud, and MSFragger-Glyco, respectively. The number of unique intact glycopeptides identified was 1992 for pGlyco, 2444 for MS-PyCloud, and 2093 for MSFragger-Glyco. Additionally, the TMTpro labeling strategy reduced the MS analysis time, enabling more robust analysis of large-scale clinical samples. Using the optimal method, we investigated differences in N-glycosylation between luminal and basal subtypes of breast cancer PDX mouse models, identifying a total of 5514 unique intact N-glycopeptides. Further analysis suggested potential distinctions in N-glycosylation patterns associated with breast cancer subtypes. Our established optimal intact N-glycopeptide analysis workflow has provided a crucial reference for future glycoproteomic studies on clinical samples, laying the foundation for further clinical applications.

ASSOCIATED CONTENT

Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD054874.[21]

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.4c04466.

Additional experimental details, including sepPak C18 desalting of digested peptides, TMTpro labeling of peptides, enrichment of glycopeptides by ZIC-HILIC enrichment, enrichment of glycopeptides by MAX enrichment, basic reverse phase fractionation, LC−MS/MS analysis, and data analysis; HCD collision energy for label-free intact glycopeptides and TMTpro-labeled intact glycopeptides; identification number of sialic acid-containing intact glycopeptides from unlabeled and TMTpro-labeled intact glycopeptides; comparative performance of ZIC-HILIC and MAX enrichment strategies for direct glycopeptide enrichment; N-glycans of LLNINPNK.T (LAPM1) from different collision energies for mass spectrometry analysis; characterization of TMT pro-labeling intact N-glycopeptides using different collision energies; number of glycoproteins, glycosites, and N-glycans identified from different software tools; and distribution of glycans, peptide backbones, and proteins in TMTpro-labeled intact glycopeptides analyzed with different software tools (PDF)

Differentially quantified intact glycopeptides between basal and luminal subtypes (XLSX)

AUTHOR INFORMATION

Corresponding Author

Hui Zhang − *Department of Pathology, Oncology, Urology, and Chemical and Biomolecular Engineering, Johns Hopkins*

University School of Medicine, Baltimore, Maryland 21231, United States; orcid.org/0000-0001-8726-7098; Email: huizhang@jhu.edu

Authors

Zhenyu Sun − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States;* orcid.org/0009-0002-5004-5904

T. Mamie Lih − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States*

Jongmin Woo − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States*

Liyuan Jiao − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States*

Yingwei Hu − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States;* orcid.org/0000-0002-4629-0985

Yuefan Wang − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States;* orcid.org/0000-0001-5731-6143

Hongyi Liu − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, United States;* orcid.org/0000-0002-9444-3632

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.4c04466

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, National Cancer Institute, Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA271079), Early Detection Research Network (EDRN, U2CCA271895), and Pancreatic Cancer Detection Consortium (PCDC, U01CA274514).

■ REFERENCES

(1) Gemmer, M.; Chaillet, M. L.; van Loenhout, J.; Cuevas Arenas, R.; Vismpas, D.; Gröllers-Mulderij, M.; Koh, F. A.; Albanese, P.; Scheltema, R. A.; Howes, S. C.; Kotecha, A.; Fedry, J.; Förster, F. *Nature* **2023**, *614*, 160−167.

(2) Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattinger, A.; Durnberger, G.; Dusberger, F.; Elling, U.; Mach, L.; Mechtler, K.; Penninger, J. M. *Nature* **2017**, *549*, 538−542.

(3) Higel, F.; Seidl, A.; Sörgel, F.; Friess, W. *Eur. J. Pharm. Biopharm.* **2016**, *100*, 94−100.

(4) Gudelj, I.; Lame, G.; Pezer, M. *Cell. Immunol.* **2018**, *333*, 65−79.

(5) Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Hoti, N.; Zhang, Z.; Chan, D. W.; Zhang, H. *Nat. Biotechnol.* **2016**, *34*, 84−88.

(6) de Haan, N.; Yang, S.; Cipollo, J.; Wuhrer, M. *Nat. Rev. Chem.* **2020**, *4*, 229−242.

(7) Sun, Z. Y.; Fu, B.; Wang, G. L.; Zhang, L.; Xu, R. F.; Zhang, Y.; Lu, H. J. *Natl. Sci. Rev.* **2023**, *10*, nwac059.

(8) Sun, Z. Y.; Ji, G. H.; Wang, G. L.; Wei, L.; Zhang, Y.; Lu, H. J. *Chem. Commun.* **2021**, *57*, 4154−4157.

(9) Fang, P.; Ji, Y. L.; Oellerich, T.; Urlaub, H.; Pan, K. T. *Int. J. Mol. Sci.* **2022**, *23*, 1609.

(10) Xiao, H. P.; Chen, W. X.; Smeekens, J. M.; Wu, R. H. *Nat. Commun.* **2018**, *9*, 1692.

(11) Suttapitugsakul, S.; Sun, F. X.; Wu, R. H. *Anal. Chem.* **2020**, *92*, 267−291.

(12) Dong, W. B.; Chen, L.; Jia, L.; Chen, Z. X.; Shen, J. C.; Li, P. F.; Sun, S. S. *Anal. Bioanal. Chem.* **2023**, *415*, 6431−6439.

(13) Yang, W. M.; Shah, P.; Hu, Y. W.; Toghi Eshghi, S.; Sun, S. S.; Liu, Y.; Zhang, H. *Anal. Chem.* **2017**, *89*, 11193−11197.

(14) Xia, C. S.; Jiao, F. L.; Gao, F. Y.; Wang, H. P.; Lv, Y. Y.; Shen, Y. H.; Zhang, Y. J.; Qian, X. H. *Anal. Chem.* **2018**, *90*, 6651−6659.

(15) Cao, L.; Huang, C.; Cui Zhou, D.; Hu, Y.; Lih, T. M.; Savage, S. R.; Krug, K.; Clark, D. J.; Schnaubelt, M.; Chen, L.; et al. *Cell* **2021**, *184*, 5031−5052.e26.

(16) Fang, P.; Ji, Y.; Silbern, I.; Doebele, C.; Ninov, M.; Lenz, C.; Oellerich, T.; Pan, K. T.; Urlaub, H. *Nat. Commun.* **2020**, *11*, 5268.

(17) Lee, H. J.; Cha, H. J.; Lim, J. S.; Lee, S. H.; Song, S. Y.; Kim, H.; Hancock, W. S.; Yoo, J. S.; Paik, Y. K. *J. Proteome Res.* **2014**, *13*, 2328−2338.

(18) Fang, Z.; Qin, H. Q.; Mao, J. W.; Wang, Z. Y.; Zhang, N.; Wang, Y.; Liu, L. Y.; Nie, Y. Z.; Dong, M. M.; Ye, M. L. *Nat. Commun.* **2022**, *13*, 1900.

(19) Halim, A.; Westerlind, U.; Pett, C.; Schorlemer, M.; Rüetschi, U.; Brinkmalm, G.; Sihlbom, C.; Lengqvist, J.; Larson, G.; Nilsson, J. *J. Proteome Res.* **2014**, *13*, 6024−6032.

(20) Mertins, P.; Tang, L. C.; Krug, K.; Clark, D. J.; Gritsenko, M. A.; Chen, L. J.; Clauser, K. R.; Clauss, T. R.; Shah, P.; Gillette, M. A.; et al. *Nat. Protoc.* **2018**, *13*, 1632−1661.

(21) Perez-Riverol, Y.; Bai, J. W.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S. B.; Brazma, A.; Vizcaíno, J. A. *Nucleic Acids Res.* **2022**, *50*, D543−D552.

(22) Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgarden, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. *Nat. Methods* **2020**, *17*, 399−404.

(23) Qing, G. Y.; Yan, J. Y.; He, X. N.; Li, X. L.; Liang, X. M. *TrAC, Trends Anal. Chem.* **2020**, *124*, 115570.

(24) Neue, K.; Mormann, M.; Peter-Katalinic, J.; Pohlentz, G. *J. Proteome Res.* **2011**, *10*, 2248−2260.

(25) Yeh, C. H.; Chen, S. H.; Li, D. T.; Lin, H. P.; Huang, H. J.; Chang, C. I.; Shih, W. L.; Chern, C. L.; Shi, F. K.; Hsu, J. L. *J. Chromatogr., A* **2012**, *1224*, 70−78.

(26) Hu, Y.; Shah, P.; Clark, D. J.; Ao, M.; Zhang, H. *Anal. Chem.* **2018**, *90*, 8065−8071.

(27) Lih, T. M.; Cho, K. C.; Schnaubelt, M.; Hu, Y.; Zhang, H. *Cell Rep.* **2023**, *42*, 112409.

(28) Cho, K. C.; Chen, L. J.; Hu, Y. W.; Schnaubelt, M.; Zhang, H. *ACS Chem. Biol.* **2019**, *14*, 58−66.

(29) Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R. *J. Proteome Res.* **2020**, *19*, 3286−3301.

(30) Tiemeyer, M.; Aoki, K.; Paulson, J.; Cummings, R. D.; York, W. S.; Karlsson, N. G.; Lisacek, F.; Packer, N. H.; Campbell, M. P.; Aoki, N. P.; et al. *Glycobiology* **2017**, *27*, 915−919.

(31) Zeng, W. F.; Cao, W. Q.; Liu, M. Q.; He, S. M.; Yang, P. Y. *Nat. Methods* **2021**, *18*, 1515−1523.

(32) Polasky, D. A.; Yu, F. C.; Teo, G. C.; Nesvizhskii, A. I. *Nat. Methods* **2020**, *17*, 1125−1136.

(33) Toghi Eshghi, S.; Shah, P.; Yang, W. M.; Li, X. D.; Zhang, H. *Anal. Chem.* **2015**, *87*, 5181−5188.

(34) Hamouda, H.; Kaup, M.; Ullah, M.; Berger, M.; Sandig, V.; Tauber, R.; Blanchard, V. *J. Proteome Res.* **2014**, *13*, 6144−6151.

(35) Yu, R.; Longo, J.; van Leeuwen, J. E.; Zhang, C. J.; Branchard, E.; Elbaz, M.; Cescon, D. W.; Drake, R. R.; Dennis, J. W.; Penn, L. Z. *Cancer Res.* **2021**, *81*, 2625−2635.

(36) Gudjonsson, T.; Adriance, M. C.; Sternlicht, M. D.; Petersen, O. W.; Bissell, M. J. *J. Mammary Gland Biol. Neoplasia* **2005**, *10*, 261−272.

(37) Cimino, A.; Halushka, M.; Illei, P.; Wu, X.; Sukumar, S.; Argani, P. *Breast Cancer Res. Treat.* **2010**, *123*, 701−708.

(38) Sethi, M. K.; Kim, H.; Park, C. K.; Baker, M. S.; Paik, Y. K.; Packer, N. H.; Hancock, W. S.; Fanayan, S.; Thaysen-Andersen, M. *Glycobiology* **2015**, *25*, 1064−1078.

(39) Liu, X.; Gao, J.; Sun, Y.; Zhang, D.; Liu, T.; Yan, Q.; Yang, X. *Biol. Chem.* **2017**, *398*, 1119−1126.

https://doi.org/10.1021/acs.analchem.4c04466
Anal. Chem. 2024, 96, 20481−20490