# Mass spectrometry-based structure-specific N-glycoproteomics and biomedical applications

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Acta Biochimica et Biophysica Sinica</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>ABBS-2024-278.R3</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Review</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>23-Jul-2024</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>bi, ming tian, zhixin; Tongji University,</td>
</tr>
<tr>
<td>Keywords:</td>
<td>N-glycosylation, N-glycoproteomics, mass spectrometry, structure-specific, biomedical application</td>
</tr>
</tbody>
</table>
Review

Mass spectrometry-based structure-specific N-glycoproteomics and biomedical applications
Ming Bi, and Zhixin Tian *

School of Chemical Science and Engineering, Tongji University, Shanghai 200092, China
*Correspondence address. Tel: +86-18601733248; Email: zhixintian@tongji.edu.cn

Received: 25-Apr-2024
Accepted: 18-Jul-2024

Running title: Mass spectrometry-based structure-specific N-glycoproteomics

Copyright: The Author(s) 2024. This is an open access article distributed under the terms of the
Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Abstract
N-linked glycosylation is a common posttranslational modification of proteins that results in macroheterogeneity of the modification site. However, unlike simpler modifications, N-glycosylation introduces an additional layer of complexity with tens of thousands of possible structures arising from various dimensions, including different monosaccharide compositions, sequence structures, linking structures, isomerism, and three-dimensional conformations. This results in additional microheterogeneity of the modification site of N-glycosylation, i.e., the same N-glycosylation site can be modified with different glycans with a certain stoichiometric ratio. N-glycosylation regulates the structure and function of N-glycoproteins in a site- and structure-specific manner, and differential expression of N-glycosylation under disease conditions needs to be characterized through site- and structure-specific quantitative analysis. Numerous advanced methods ranging from sample preparation to mass spectrum analysis have been developed to distinguish N-glycan structures. Chemical derivatization of monosaccharides, online liquid chromatography separation and ion mobility spectrometry enable the physical
For Peer Review

differentiation of samples. Tandem mass spectrometry further analyzes the macro/microheterogeneity of intact N-glycopeptides through the analysis of fragment ions. Moreover, the development of search engines and AI-based software has enhanced our understanding of the dissociation patterns of intact N-glycopeptides and the clinical significance of differentially expressed intact N-glycopeptides. With the help of these modern methods, structure-specific N-glycoproteomics has become an important tool with extensive applications in the biomedical field.

Key words N-glycosylation, N-glycoproteomics, mass spectrometry, structure specific, biomedical application

Introduction

N-linked glycosylation, referred to as N-glycosylation, is a common posttranslational modification of proteins. For the glycosylation site, N-glycosylation occurs at the motif of N-X-S/T/C (X≠P), and approximately 80% of human proteins contain at least one such motif, which is called a putative N-glycoprotein[1, 2]. A protein may contain more than one motif in its sequence, and an N-glycoprotein can be glycosylated by one or more N-glycans[3, 4]. This macroheterogeneity in the N-glycosylation site parallels that of other small molecules, such as methylation, acetylation, and phosphorylation. In terms of the modification structure, unlike other small-molecule modifications, which involve only a single structure, N-glycosylation involves tens of thousands of structures, including monosaccharide compositions, sequence structures, linkage structures, positional isomers, stereo-conformations and other structural dimensions[5-8]. Consequently, N-glycosylation exhibits microheterogeneity, as one glycosite is commonly modified with different N-glycan structures at specific stoichiometric ratios[9, 10]. N-glycosylation regulates the structure and function of N-glycoproteins in a site- and structure-specific manner[11, 12]. For example, glycosylation at N71 on mutated programmed cell death protein 1 (PD-1) can decrease the suppression of chimeric antigen receptor T cells (CAR-T cells) to enhance the cytotoxicity and efficacy of immunotherapy [13]. N-glycosylation of N57 on PD-1 is necessary for its binding to PD-L1 [14]. AFP-L3, a glycosylation variant of alpha fetoprotein (AFP) with a specific core-fucosylated N-glycan structure (YY(F)M(MYLS)MYLS, where Y represents N-acetylglucosamine, M represents mannose,
L represents galactose, F represents fucose and S represents sialic acid), exhibits greater sensitivity than AFP in the early diagnosis of hepatocellular carcinoma (HCC) [15]. The terminal fucose of N-glycan is highly involved in various biological processes, such as virus infection [16], the gut microbiome [17], and intestinal commensal bacteria [18]. Furthermore, terminal sialic acid is strongly associated with anti-inflammatory effects [19], allergies [20], and the gut microbiome [21]. The immunoglobulin G (IgG) modified with α2,6 sialic acid in the Fc fragment is an anti-inflammatory agent for arthritis treatment, whereas its α2,3 sialic acid is not. N-glycosylation under pathological conditions was characterized by site- and structure-specific quantitative analysis.

With the rapid development and widespread application of material preparation, highly efficient liquid chromatography separation, high-throughput tandem mass spectrometry analysis, intelligent bioinformatics and other technologies, omics technology based on mass spectrometry has been widely applied for qualitative and quantitative analysis of site- and structure-specific protein N-glycosylation under both normal physiological and abnormal pathological conditions.

At the molecular level, N-glycosylation analysis can be conducted on four levels: N-glycoprotein without the use of any enzyme; N-linked glycan obtained with glycosidases (e.g., PNGase F); intact N-glycopeptide obtained with proteases (e.g., trypsin); and peptide containing glycosite obtained with glycosidases and proteases. N-glycosylation analysis of N-linked glycans was reviewed by Tian [22], the Human Glycome Project [23] and Mechref [24]. For N-glycosylation at the level of intact N-glycopeptide, enrichment [25], tandem mass spectrometry dissociation [26], quantitation [27], and biological informatics [28] have been reviewed separately, and Kolarich [29], Scott [30] and Thaysen-Andersen [31] et al. have conducted comprehensive reviews.

For structure-specific analysis, methods and technologies across all stages of the N-glycoproteomic pipeline, including sample preparation, liquid chromatographic separation, tandem mass spectrometry (MS) and informatics, have been extensively developed and are well reviewed in 2021 by Morten Thaysen-Andersen et al. [32].

Here, we provide a selected review of recent progress in mass spectrometry-based structure-specific N-glycoproteomics and biomedical applications. This review highlights the latest developments in N-glycoproteomics, including derivation, separation, ion mobility techniques, tandem mass spectrometry and AI-based informatics. Additionally, biomedical applications of these techniques, including understanding functions, characterizing subtypes, identifying biomarkers and developing therapeutic strategies, were explored (Figure 1).

**Mass Spectrometry-based Methods and Techniques for Structure-specific N-glycosylation Characterization**

This article describes the development of site- and structure-specific N-glycoproteomic methods, including N-glycosite localization based on site-determining fragment ions [33]; the derivatization of selective chemicals for the identification of sialic acid linkages [34]; the separation of hydrophilic chromatographies for the identification of sialic acid linkages and fucose positions; the identification of N-glycan sequence structures based on
structure-diagnostic ions; and the verification of specific sequences and linkage structures (such as core fucose, bisected N-acetylglucosamine, and sialic acid linkage isomers) based on characteristic fragment ions.

**Bracketing of site-determining fragment ions for localization of N-glycosites**

A peptide or protein amino acid sequence commonly contains multiple putative N-glycosines. Considering the high probability of neutral losses of vulnerable modifications such as phosphorylation and glycosylation in tandem mass spectrometry, Tian et al. developed a localization method based on site-determining fragment ion pairs containing modifications (residue or intact molecule) utilizing direct experimental evidence to accurately pinpoint modifications (Figure 2). The phosphorylation localization tool P-Bracket, which utilizes this localization strategy, was used to analyze a dataset of 96 LC-MS/MS spectra containing 101,520 synthesized phosphorylated peptides, and the mistaken localization rate was 0.9% [35]. At the N-glycosite site, accurate localization is based on site-determining b/y ions containing residual N-acetylglucosamine residues (produced by high-energy collision dissociation) and c/z ions containing intact N-glycan molecules (produced by electronic transferred dissociation).

**Derivatization of selective chemicals for the identification of sialic acid linkages**

Alkylamidation derivatization is a chemical reaction common to sialic acid α2,3 and α2,6 linkages in sialic acid. α2,3-linked sialic acid selectively reacts with methylamine, resulting in a mass increase of 13 Da, while α2,6-linked sialic acid selectively reacts with isopropylamine, resulting in a mass increase of 41 Da. Using multiplex tandem mass tag (TMT) labeling, Tian et al. identified 307 intact sialylated N-glycopeptides with linkage-specific sialic acid from differentially expressed sialylated N-glycoproteomes in lung cancer tissues compared to those in adjacent tissues. These corresponded to 84 N-glycans, 232 N-glycosites, 229 peptide backbones and 221 N-glycoproteins. They quantified 84 intact sialylated N-glycopeptides with linkage-specific sialic acid (fold change≥1.5, P<0.05), including 29 with α2,6-linked sialic acids and 55 with α2,3-linked sialic acids. For example, glucose-6-phosphate isomerase (P06744, G6PI_HUMAN), a glycopeptide at N105 with two α2,6-linked sialic acids, increased in cancer tissues by 1.92-fold. An intact N-glycopeptide at N677 on tyrosine-protein kinase (Q9H792, PEAK1_HUMAN) with two α2,3-linked sialic acids was downregulated 0.83-fold (Figure 3).

The Derivatization of Sialylated Glycopeptides (DOSG) method involves alkylamidation derivatization to identify sialic acid linkages and enhance detection sensitivity in mass spectrometry. This approach is based on isopropylamidating carboxylic groups on the peptide moiety, including glutamic acids, aspartic acids, and C-terminal residues[36]. By applying DOSG, a 4.6-fold increase in signal intensity was achieved due to efficient derivatization of the peptide backbone, as demonstrated by the analysis of N-glycopeptides and O-glycopeptides from fetuin. When coupled with LC–MS/MS, DOSG enabled the discovery of 11 N-glycan structures, encompassing 28 isoforms, from human α1-acid glycoprotein (AGP). For instance, a glycopeptide modified with the N-glycan
N5H6F0S3 revealed three distinct linkage structures: biα2,3-linked and mono-α2,6-linked sialic acids; mono-α2,3-linked and biα2,6-linked sialic acids; and tri-α2,6-linked sialic acids.

Lu et al. [37] devised a method that combines alkylamidation derivatization with stable-isotope coding to establish a stable-isotope-based sequential derivatization technique. This innovative approach optimizes derivatization conditions, enhances ionization efficiency and reduces retention time discrepancies. In their study, they applied this method to analyze IgG extracted from serum samples obtained from HCC patients and healthy volunteers. Through this approach, Lu et al. detected five pairs of sialyl-linkage isomers. Interestingly, among these isomers, the α2,3-linked glycopeptide with the N-glycan N4H4F2S1 exhibited no significant change in HCC serum samples, whereas the α2,6-linked glycopeptide showed a marked decrease.

**Separation of hydrophilic chromatography for identification of sialic acid linkages and fucose positions**

Hydrophilic interaction chromatography (HILIC) is characterized by efficient separation of sialic acid linkage isomers and fucose position isomers and can achieve baseline separation at the level of an intact N-glycopeptide. Analysis of the differentially expressed N-glycopeptide in HepG2 cells via pentaHILIC-MS/MS revealed that the intact N-glycopeptide FLSSPHLPSSYFNASGR_N3H4F0S1 from tripeptidyl-peptidase 1 (O14773, TPP1-HUMNA) with α2,3-linked sialic acid was upregulated in HepG2 cells (6.2±1.2), while the same composition and sequence structure of the N-glycopeptide with α2,6-linked sialic acid was downregulated (0.3±0.1) [38]. Two structures of the intact N-glycopeptide MNITVK_N4H5F1S2 from integrin alpha-3 (P26006, ITA_HUMAN) were identified, and among these, the branch-fucosylated glycopeptide was downregulated (0.4±0.1), while the core-fucosylated glycopeptide was upregulated (1.5±0.1).

The semen prostate-specific antigen (PSA) serves as a biomarker for prostate cancer (PCA) but has limited specificity. However, glycosylated PSA exhibits enhanced specificity. Using HILIC-MRM-MS, sialylated PSA underwent dissociation of its linkage and subsequent quantification[39]. Three chromatographic peaks were observed for tryptic PSA modified with the N-glycan H5N4F1S2. Following digestion by sialidase A, which cleaves α2,3- and α2,6-linked sialic acids, the glycopeptide of PSA exhibited one peak, indicating that the N-glycan structures in the three peaks were the same in addition to terminal sialic acid. Compared with the intensity of three peaks after digestion by sialidase S, which selectively removes only α2,3-linked sialic acid while retaining α2,6-linked sialic acid, N-glycopeptide with α2,3-linked sialic acid eluted slightly earlier than α2,6-linked sialic acid in HILIC, and the middle peak indicated mono α2,3-linked sialic acid and mono α2,6-linked sialic acid.

**Structure-diagnostic fragment ions of tandem MS for identification of sialic acid linkages**

Nilsson and Westerlind et al. [40] found that a dedicated intensity ratio of sialic acid-containing oxonium ions in tandem MS can distinguish α2,3 and α2,6 linkages in both N- and O-linked glycopeptides, which was confirmed with synthetic linkage-specific
For higher-energy collision dissociation (HCD) in Orbitrap MS, at a normalized collision energy (NCE) of 30\%, the Ln/Nn (normalized LacNAc to Neu5Ac) ratio for Neu5Ac $\alpha2,3$ was 0.5-0.7 (Figure 4), while for Neu5Ac $\alpha2,6$, it was 0.9-1.5. At an NCE of 40\%, the Ln/Nn ratio for the Neu5Ac $\alpha2,3$ glycopeptide was 0.5, while the Ln/Nn ratio for the Neu5Ac $\alpha2,6$ glycopeptide was 0.8-1.3. In MS2, Ln/Nn was defined as

$$ \frac{L_n}{N_n} = \frac{I_{204} + I_{366}}{I_{274} + I_{292}} \times \frac{n(\text{Neu5Ac})_{\text{GalGlcNAc}} + 0.5 \times n(\text{Neu5Ac})_{\text{GalGalNAc}}}{n(\text{GlcNAc}) + 0.5 \times n(\text{GalNAc})} $$

The relative intensities of fragment ions derived from the collision energy of N-glycan isomers exhibited notable differences, particularly for branched galactose and terminal sialic acid [41]. Experimental investigations were conducted using standard intact N-glycopeptides modified with the biantenna N-glycan N4H5F0S2, comprising one glycopeptide with two $\alpha2,6$-linked sialic acids and another with two $\alpha2,3$-linked sialic acids. At a collision energy of 20\%, the relative abundances of the oxonium ions B- and B-H2O- ions (m/z 292.1027 and 274.0921, respectively) both accounted for approximately 40\% of the base peak intensity (BPI) for the $\alpha2,3$-linked glycopeptide. In contrast, these values were only 15 and 20\% for the $\alpha2,6$-linked glycopeptide. Similar disparities in BPI were observed across a range of collision energies from 5\% to 50\%, irrespective of the peptide sequence. This observation underscores the potential of fragment ions derived from glycopeptides for discerning N-glycan structures, including linkage isomers. However, further in-depth research is warranted to explore and validate their utility in complex samples.

**Collision cross section of ion-mobility spectrometry for identification of N-glycan 3D structure and conformations**

Ion mobility (IM) is the fourth dimension of mass spectrometry analysis, offering the ability to separate structural isomers based on drift time. This technology relies on the interaction of the analytical sample with buffer gas, where the size and structure of the sample influence the drift time, enabling the separation of isomers. IM-MS has found applications in various fields, including proteomics, glycomics, and glycoproteomics. Hinneburg et al. [42] successfully applied IM-MS to separate sialic acid linkage isomers. The two synthesized N-glycopeptides, EVFVHPNYSK_N4H4F0S1, with different sialic acid linkages were not separated at the level of the intact N-glycopeptide, but the common fragment ions N-H-S (B3-1+) were significantly different in drift time (Figure 5). The collision cross section in nitrogen drift ($^{\text{TW}}$CCS$_{\text{N2}}$) gas of the $\alpha2,6$-linked fragment ion was 236 Å$^2$, while that of the $\alpha2,3$-linked fragment ion was 246 Å$^2$.

By utilizing CCSs of diagnostic fragment ions, linkage-specific N-glycopeptides can be efficiently distinguished without the need for additional derivatization. This IM-MS method has been successfully applied to clinical samples. Lu et al. [43] employed this method for the quantitative analysis of sialylated haptoglobin (Hp) in HCC serum. They calculated the relative abundance of $\alpha2,3$- and $\alpha2,6$-linked sialylated N-glycopeptides using the area under the arrival time distributions (ATDs) of B3 fragments (m/z 657) in IM. This relative
quantification method demonstrated high linearity (R2 > 0.99), good reproducibility (coefficient of variation (CV) < 10%), and a low limit of quantification (LOQ, 35 nM). The ratios of α2,3- and α2,6-B3 ions from N-glycopeptides of the same composition, such as the intact N-glycopeptide MVSHHNLTGTATLINE_H6N5F0S3, were calculated to be 0.23, while those of NLFLNHSE_H6N5F0S3 and VVLHPNYSQVD_H6N5F0S3 were 0.50 and 0.64, respectively. After more HCC patient verification, the ratios of α2,3- and α2,6-linked N-glycopeptides were found to be different; MVSHHNLTGTATLINE_H5N4F1S2 was significantly altered, with a ratio of 0.06±0.01, and the area under the ROC curve (AUC) was 0.824, indicating promising diagnostic performance.

**Structure-diagnostic fragment ions of tandem MS for validation of bisecting motif and monosaccharide sequence isomers**

One N-glycan composition commonly corresponds to numerous sequence structures, and different sequence structures can be distinguished and confirmed by not only false discovery rate (FDR) control based on target-decoy database searches at the spectral level but also structure-diagnostic fragment ions. Tian et al. defined the Glycoform score (briefly called the G score) as the number of structure-diagnostic fragment ions capable of independently discriminating different sequence structures composed of the same molecular composition, including the same monosaccharide composition [38, 44]. The two N-glycan structures, the branch fucose isomer and core fucose isomer, which are composed of the monosaccharide N4H5F1S2, were identified and verified by four (3,5Al4-1+, YI3-2+, YI2-2+, YI1-2+) and three (YI1-1+, ZI4-2+, YI3-2+) structure-diagnostic ions separately dissociated by HCD (Figure 6A-D) [45].

Characteristic sequence structures such as bisecting and core fucose can be specifically verified by the corresponding characteristic monosaccharide sequences. The characteristic monosaccharide sequence of the bisecting motif is N-N-H-N or N(F)-N-H-N (containing core fucose). For example, in the identification of the structure of an intact N-glycopeptide with the N-glycan N5H9F3S1 (Figure 6E), the bisecting structure was verified by observing the two-charge fragment ion of N-N-H-N. At the intact N-glycopeptide level, low-energy fragmentation under HCD tends to generate more characteristic structures. It has been demonstrated that using 20% HCD energy produces the highest intensity fragment ion of the peptide backbone with N-N-H-N [46]. This method has been effectively applied in various studies. For example, in kidney tissue samples from rats, 183 intact N-glycopeptides were identified, including 47 containing bisecting N-glycans, by matching the characteristic bisecting fragment ions in low-energy collision experiments.

**Bioinformatics and AI in glycan structure characterization**

StrucGP [47] identifies the structure of an intact N-glycopeptide from three distinct components of the core structure, the glycan subtype and the branch structures based on B and Y ions in the tandem mass spectrum from the glycan moiety, where an independent database is adopted to enhance structure identification. In addition, characteristic B and Y ions are used for structural diagnosis and isomer distinction (such as fucose position isomers and bisecting N-acetylglucosamine).
With the development of computer science, artificial intelligence (AI) has become indispensable in glycoproteomics. Machine learning tools have been employed to enhance the analysis and interpretation of N-glycoproteomic data. Using the GlycoProteome Analyzer (IQ-GPA) tool for identification and quantification, core-fucosylated IgG and outer-fucosylated alpha-1-acid glycoprotein (AGP) were analyzed by LC–MS/MS as training models. The relative abundances of B/Y fragment ions were calculated via deep neural network (DNN) and support vector machine (SVM) methods, which output fucose types, including “none”, “core”, “outer”, and “dual”[48]. From the training model, the area under the curve (AUC) and FDR were calculated and applied in the fucosylation classification of human plasma. DeepGlyco[49] employs a deep learning model to predict the fragment spectra of intact N-glycopeptides, demonstrating its partial ability to discriminate glycan structural isomers. GlycanFinder[50] is designed for de novo sequencing, enabling the identification of potentially novel N-glycans in glycopeptide mass spectra. The machine learning model was trained to reconstruct glycan structures without any predefined rules. When a spectrum was not matched to any N-glycan in the given database, the intact N-glycopeptide was constructed from the peptide backbone to the N-glycan moiety by iterative addition of monosaccharides predicted by the learning model. A transformer neural network for graphs captures the structures, and a second neural network captures the Y and B ions of the matched glycopeptide between the candidate structures and the spectrum. The top-scoring glycan structures are then selected for further iteration, which refines the identification process and improves the accuracy of N-glycopeptide characterization.

Biomedical application of mass spectrometry-based structure-specific N-glycoproteomics
The aforementioned structure-specific N-glycoproteomic methods have been applied in the N-glycosylation analysis of various systems, such as tumor cell models (cancer cells [38, 51], cancer stem cells [52], and drug-resistant cancer cells [53-57]), paired clinical cancer tissues and adjacent tissues, cancer serum [58], human body fluids (urine [45]), and large molecule drugs (COVID-19 recombinant vaccines [59] and monoclonal antibodies [60]).

Identification of putative N-glycoprotein biomarkers
In site- and structure-specific quantitative N-glycoproteomic studies in cell lines, aberrant site and structure N-glycosylation has been observed. In HepG2 cells compared to LO2 cells, a total of 720 differentially expressed N-glycopeptides were identified, with 447 showing decreased expression and 273 showing increased expression. Notably, various sialic acid linkage isomers were separated by HILIC and exhibited different fold changes [38]. For example, the 3-linked isomer on the N-glycosite N403 of ubiquitin-associated domain-containing protein 1 was upregulated (fold change $2.6 \pm 0.3$), while the corresponding 6-linked isomer was downregulated (fold change $0.2 \pm 0.0$). The intact N-glycopeptide NCTSISGDHLHIPVAFR in epidermal growth factor receptor (EGFR, P00533), with N-glycan compositions of N2H7F0S0 and N2H8F0S0, were both
upregulated. Notably, EGFR was quantified in a proteomics study of HCC cancer tissues, where it was consistently upregulated. EGFR has been implicated in various cancer types, including liver cancer, where it serves as a receptor for hepatitis C virus (HCV) and plays a role in viral entry and membrane fusion processes. In the context of breast cancer, 144 differentially expressed intact N-glycopeptides were observed in MCF-7 cancer stem cells (CSCs) compared to non-CSCs. Notably, several putative N-glycoprotein biomarkers, including zinc finger protein GLI1 (N344_N2H9F0S0, upregulated), CD63 (N130_N2H6F0S0, downregulated), CD49F (N332_N2H8F0S0, upregulated), CD151 (N159_N2H6F0S0 and N159_N2H7F0S0, downregulated), and CD97 (N453_N4H5F1S0, upregulated), were identified [52]. In MCF-7/ADR CSCs, several drug resistance markers, such as adenosine triphosphate-binding cassette (ABC) drug transporters, which serve as drug efflux pumps, were identified [61]. Three ABC proteins, ABCC5, ABCA4 and ABCB9, were quantified as intact N-glycopeptides modified with the N-glycan N2H8F0S0 in MCF-7/ADR CSCs, in which glycosylated ABCC5 decreased by 0.46-fold, glycosylated ABCA4 decreased by 0.22-fold, and glycosylated ABCB9 increased by 1.89-fold. The dysregulation of ABC transporters in chemoresistant cancer is not negligible.

Furthermore, N-glycan size-dependent differential expression was also observed. In the N-glycoproteomic analysis of MCF-7/ADR CSCs compared to MCF-7/ADR CSCs, high-mannose N-glycans exhibited size-dependent characteristics [61] (Figure 7). At N-glycosite 120 of serpin H1 (P50454, SERPH_HUMAN), the modified N-glycans were N2H5F0S0, N2H6F0S0, N2H7F0S0, N2H8F0S0 and N2H9F0S0, with ratios of 3.80±0.05, 1.57±0.06, 1.30±0.11, 0.64±0.08 and 0.50±0.14, respectively. As the number of mannoses increased, the fold change in the level of the intact N-glycopeptide decreased. Interestingly, in the quantitative N-glycoproteomic analysis of MCF-7/ADR cells relative to MCF-7 cells, a size-dependent effect of SERPH_N120 was also observed, but as the number of mannoses increased, the ratio increased [54]. Serpin H1 was found to be upregulated in fourteen cancers, including breast cancer. The upregulation of serpin H1 has been associated with reduced overall survival (OS), disease-specific survival and progression-free intervals [62]. The N-glycosylation of serpin H1 deserves further exploration in other cancers.

Tissue specimens are more suitable for clinical research in disease systems, including cancers. In gastric cancer, 644 site- and structure-specific N-glycopeptides were quantified that satisfied the criteria of a 1.5-fold change and a p value <0.05 [63]. The N-glycan N4H5F0S2, which is attached to the N-glycosite N271 of N-glycoprotein alpha-1-antitrypsin (A1AT_HUMAN, P01009), was downregulated (0.29 ± 0.02) in gastric cancer tissue relative to adjacent tissue. The AIAT was identified and validated as a potential diagnostic biomarker for gastric cancer proteomics [64, 65]. Lu et al analyzed N-glycoproteomic data from pancreatic cancer tissue from four patients and found that 24 intact N-glycopeptides were differentially expressed in at least three patients, including 19 intact N-glycopeptides whose expression was upregulated and 5 whose expression was downregulated [66]. One of the differentially expressed glycoproteins, a putative biomarker, thrombospondin-1 (TSP1_P07996), was identified as an N-glycopeptide with N-glycans N4H5F0S0, N4H5F0S1 and N5H4F0S0 and exhibited a fold change of 2.58 to
Thrombospondin-1 is a natural angiogenesis inhibitor, and when combined with different proteins, it participates in various functions in cancers, such as the promotion of colorectal cancer and the inhibition of ovarian cancer [67]. In an investigation of lung tumor tissues and adjacent nontumor tissues, 521 intact N-glycopeptides were identified, including 214 upregulated and 307 downregulated [34]. Among these glycoproteins, 7 glycoproteins were enriched in the N-glycan biosynthesis signaling pathway. According to quantitative N-glycoproteomic analysis of tumor and para-cancerous tissues from hepatocellular carcinoma patients, alpha-1-antitrypsin (A1AT) has been shown to exhibit macro- and microheterogeneity [68]. The concentration of A1AT was shown to have a sensitivity of 99% and a specificity of 79% for distinguishing HCC patients from healthy individuals [69]. Eleven unique N-glycopeptides from A1AT were quantified, including 2 glycosites, N107 and N271. The two glycopeptides N107_N4H5F0S2 and N271_N4H5F0S2 were modified with the same N-glycan structure but exhibited different expression patterns, in which N107_N4H5F0S2 increased while N271_N4H5F0S2 decreased. The same N-glycopeptides, N107_N4H5F0S2 and N271_N4H5F0S2, were also identified by Lubman’s group [70] using LC-HCD-PRM-MS/MS in liver disease patient serum, and the ratio of different charge states (2+/3+) of N271_N4H5F0S2 of A1AT can significantly distinguish early-stage HCC from cirrhosis, with an area under the receiver operating characteristic curve (AUC) of 0.9.

Compared with traditional protein biomarkers, N-glycoprotein biomarkers offer enhanced sensitivity and specificity due to the presence of site- and structure-specific N-glycan moieties. Alpha-fetoprotein (AFP) is a common biomarker for diagnosing HCC, whereas fucosylated AFP (AFP-L3) demonstrates increased specificity and sensitivity[71]. In addition to diagnosis, N-glycosylation-specific subtyping has also been widely reported for clinical tumors. Zhang et al.[72] collected 83 high-grade serous ovarian carcinoma (HGSC) tissues and utilized glycoproteomics to classify tumors into three distinct clusters that could be correlated with the clinical phenotypes of tumor cellularity and anatomic site. The glycoproteins in cluster 1 were modified by complex glycans containing fucose and/or sialic acid glycans from the complement and coagulation cascade pathways. The glycoproteins in cluster 2 were predominantly modified by high-mannose or fucosylated glycans from the ECM-receptor interaction pathway. In N-glycoproteomic analysis of clear cell renal cell carcinoma, three distinct glycosylation types were identified[73]; one subtype was characterized by high-grade and high-stage tumors, while the other two had lower grades and stages.

Characterization of N-glycoprotein drugs
In addition to comprehensive clinical samples, site- and structure-specific N-glycoproteomics has also been employed in the analysis of single N-glycoprotein micro- and macroheterogeneity. One N-glycan structure can be modified at different glycosites, and one N-glycan monosaccharide tends to have multiple corresponding structures. For example, the recombinant SARS-CoV-2 S protein receptor binding region contains three N-glycosites, N331, N334 and N343 [59], where 12, 17, and 19 N-glycan structures, respectively, each containing three sets of sequence isomers, were identified.
Tian et al. applied site- and structure-specific N-glycoproteomics to the National Institute of Standards and Technology (NIST), which provided reference material (RM) 8671, a standard humanized IgG with an N-glycosite in the Fc domain, and identified 59 N-glycan structures. Among these N-glycans, the N-glycan composition of N4H3F1S0 was verified to have four structures utilizing structure-diagnostic ions (Figure 8). In biantenna N4H3F1S0, the fragment ion XII3 discriminated the fucose locations, and the other two tri-antenna structures, 01Y(61F)41Y41M(31 M)61 M(21Y)61Y and 01Y41Y41M(31 M)61 M(21Y31F)61Y, were confirmed by two (0,1AI2 (m/z = 324.13, z = 1) and YI3 (m/z = 1085.53, z = 4)) and two (CI2 (m/z = 367.14, z = 2) and YI3 (m/z = 801.70, z = 3)) structure-diagnostic ions.

**Perspectives**

Site- and structure-specific quantitative N-glycoproteomic methods based on the latest advancements in mass spectrometry have been developed for qualitative and quantitative protein N-glycosylation analysis. At the intact N-glycopeptide molecule level, highly efficient liquid chromatography has achieved high separation and peak capacity for intact N-glycopeptide complex mixtures compared with N-linked glycans and N-glycoproteins, guaranteeing the depth of qualitative and quantitative analysis. Isotope labeling and isobaric labeling of the peptide backbone can achieve relative quantification for benchmark and large-scale samples and accurate quantification of differentially expressed intact N-glycopeptides under disease conditions. In terms of biological informatics analysis, compared with traditional mass matching, the isotopic envelope fingerprinting algorithm can more accurately recognize parent ions in MS and fragment ions in MS/MS and can more efficiently analyze overlapping spectra; additionally, the maximum matching of experimental fragment ions directly improves the identification accuracy of amino acids, monosaccharide sequence structures and N-glycosites. Accurate determination of N-glycosites relies on the presence of partial (single N-acetylg glucosamine monosaccharides remaining on peptide backbones under high-energy collision-induced dissociation) or intact (intact N-glycan portions remaining after electron-based dissociation) glycosite site-determining fragment ions, which is direct experimental evidence. For the monosaccharide sequence structure, potential sequence structures corresponding to a single molecule composition, including the N-glycan composition, can be discriminated and confirmed by structure-diagnostic ions and characteristic fragment ions; for sialic acid linkages, linkage-specific chemical derivations (such as alkylamidation), hydrophilic interaction chromatography and tandem mass spectrometry can be used to effectively discriminate and confirm these structures. Promotion of site- and structure-specific quantitative N-glycoproteomics is beneficial for the development of highly efficient N-glycoprotein early diagnostic biomarkers, research on drug and drug targets and precision medicine.

**Funding**
This research was financially supported by the grants from the National Natural Science Foundation of China (Nos. 22074105 and 21775110) and the Shanghai Science and Technology Commission (No. 14DZ2261100).

Conflict of Interest
The authors declare no conflict of interest.

References


5 Trefulka M, Palecek E. Distinguishing glycan isomers by voltammetry. Modification of 2,3-sialyllactose and 2,6-sialyllactose by osmium(VI) complexes. Electrochemistry Communications 2017, 85: 19-22


8 She YM, Tam RY, Li XG, Rosu-Myles M, Sauvé S. Resolving Isomeric Structures of Native Glycans by

Acta Biochimica et Biophysica Sinica


26 Reiding KR, Bondt A, Franc V, Heck AJR. The benefits of hybrid fragmentation methods for

27 Delafield DG, Li LJ. Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. Molecular & Cellular Proteomics 2021, 20

28 Polasky DA, Nesvizhskii AI. Recent advances in computational algorithms and software for large-scale glycoproteomics. Current Opinion in Chemical Biology 2023, 72


34 Yang HL, Tian ZX. Sialic acid linkage-specific quantitative N-glycoproteomics using selective alkylamidation and multiplex TMT-labeling. Analytica Chimica Acta 2022, 1230


37 Peng Y, Gu B, Sun ZY, Li YY, Zhang Y, Lu HJ. Linkage-selective derivatization for glycosylation site-
and glycoform-specific characterization of sialic acid isomers using mass spectrometry. Chemical Communications 2021, 57: 9590-9593


39 van der Burgt YEM, Siliakus KM, Cobbaert CM, Ruhaak LR. HILIC-MRM-MS for Linkage-Specific Separation of Sialylated Glycopeptides to Quantify Prostate-Specific Antigen Proteoforms. Journal of Proteome Research 2020, 19: 2708-2716


41 Maliepaard JCL, Damen JMA, Boons G, Reiding KR. Glycoproteomics-Compatible MS/MS-Based Quantification of Glycopeptide Isomers. Analytical Chemistry 2023, 95: 9605-9614


45 Shen Y, Xiao KJ, Tian ZX. Site- and structure-specific characterization of the human urinary N-
glycoproteome with site-determining and structure-diagnostic product ions. Rapid Communications in Mass Spectrometry 2021, 35


49 Yang Y, Fang Q. Prediction of glycopeptide fragment mass spectra by deep learning. Nature Communications 2024, 15


53 Qin SD, Tian ZX. Gain-of-glycosylation in breast multidrug-resistant MCF-7 adenocarcinoma cells and cancer stem cells characterized by site- and structure-specific N-glycoproteomics. Analytica Chimica Acta
2023, 1252


68 Bi M, Tian ZX. High-throughput N-glycoproteomics with fast liquid chromatographic separation. Analytica Chimica Acta 2024, 1288


Figure 1. Overview of structure-specific N-glycoproteomic technology and its applications.

Figure 2. Illustration of PTM locations based on site-determining fragment ions. (A) Site-determining ion pairs containing phosphate groups for phosphorylated site location (adapted from the literature [35]). (B) Site-determining b/y ion pairs containing N-acetylglucosamine residues (generated by HCD) and c/z ion pairs containing intact N-linked glycans (generated by ETD) at the level of intact N-glycopeptide.

Figure 3. Intact N-glycopeptides identified by sialic acid derivatization (cited from the literature [34]). (A, B) Illustration of quantitative intact N-glycopeptides of α2,6-linked sialylation AVLHVALRNR_N4H7F1S2 with a fold change of 1.92 (N105, P06744, G6PI_HUMAN) and α2,3-linked sialylation VPDNTTSK_N4H7F0S2 with a fold change of 0.83 (N677, Q9H792, PEAK1_HUMAN). (C-F) Illustration of the intact N-glycopeptide TKPREEQYNSTYR_N3H6F1S1, including a graphical fragmentation map of the N-glycan moiety, an isotopic envelope fingerprinting map of the precursor ion and an annotated MS/MS spectrum with the matched fragment ions.

Figure 4. (A) Workflow of semisynthetic α2,3-/α2,6-linked sialylated intact N-glycopeptide. (B) Ln/Nn values of semisynthetic N-glycopeptides at NCEs of 25 to 45% (cited from the literature [40]).

Figure 5. Analysis of the intact N-glycopeptide EVFVHPNYSK_N4H4F0S1 by CID followed by IM-MS (cited from the literature [42]). (A) Annotated MS2 spectrum of two sialic acid linkage isomers of the intact N-glycopeptide. (B) Drift time of intact N-glycopeptide and B3 fragment ions with α2,3- and α2,6-linked sialic acid.

Figure 6. Discrimination of N-glycan isomers based on structure-diagnostic ions. (A-D) annotated MS/MS spectra and graphical dissociation of branch and core fucose isomers. (E) Annotated MS/MS spectrum matching the peptide backbone and N-linked glycan portions. The left inset shows the observed characteristic sequence N-N-H-N corresponding to the +2 charged fragment ions, while the right inset displays the graphical dissociation spectrum of the N-linked glycan portion.

Figure 7. Fold changes of the intact N-glycopeptide series SLSNSTAR_N2HxF0S0 (x = 5, 6, 7, 8, 9; from left to right: A–E) (adapted from the literature [61]).

Figure 8. Graphical fragmentation maps of four structural isomers of the N-glycan N4H3F1S0 identified in NIST8671 (cited from the literature [60]).
Structure-specific N-glycoproteomics

- Functions
- Subtypes
- Biomarkers
- Therapeutics
A  Phospho-site  P-brackets
(A) b1 Ac-W N T* Q S* T Y S E A y1
(B) b1 Ac-W N T* Q S T* Y S E A y1
(C) b1 Ac-W N T* Q S T Y S* E A y1
(D) b1 Ac-W N T* Q S T Y S* E A y1
(E) b1 Ac-W N T Q S* T* Y S E A y1
(F) b1 Ac-W N T Q S T Y S A y1
(G) b1 Ac-W N T Q S T Y S* E A y1
(H) b1 Ac-W N T Q S T Y S E A y1
(I) b1 Ac-W N T Q S T Y S* E A y1
(J) b1 Ac-W N T Q S T Y S* E A y1

B

82x79mm (300 x 300 DPI)
130x181mm (220 x 220 DPI)
A

B

166x106mm (100 x 100 DPI)
For Peer Review

A

B

165x88mm (96 x 96 DPI)
164x114mm (220 x 220 DPI)
Highlights

Completely different from small-molecule post-translational modifications with single structures, N-glycosylation has multi-dimensional structures of monosaccharide sequence, linkage, anomer and stereoisomer. This article provides an updated review of recent method and technology developments in structure-specific N-glycoproteomics as well as applications.

(1) Structure-specific methods including chemical derivatization, liquid chromatography, ion mobility, and tandem mass spectrometry are used to analyze macro-/micro-heterogeneity of intact N-glycopeptides.

(2) The development of search engines and AI-based software has improved the understanding of dissociation patterns of intact N-glycopeptides and their clinical significance.

(3) Structure-specific N-glycoproteomics has become an important tool with extensive applications in the biomedical field.
Structure-specific N-glycoproteomics

- Functions
- Subtypes
- Biomarkers
- Therapeutics

150x73mm (300 x 300 DPI)