

Recent Advances in Software Tools for More Generic and Precise Intact Glycopeptide Analysis

Authors

Weiqian Cao, Mingqi Liu, Siyuan Kong, Mengxi Wu, Yang Zhang, and Pengyuan Yang

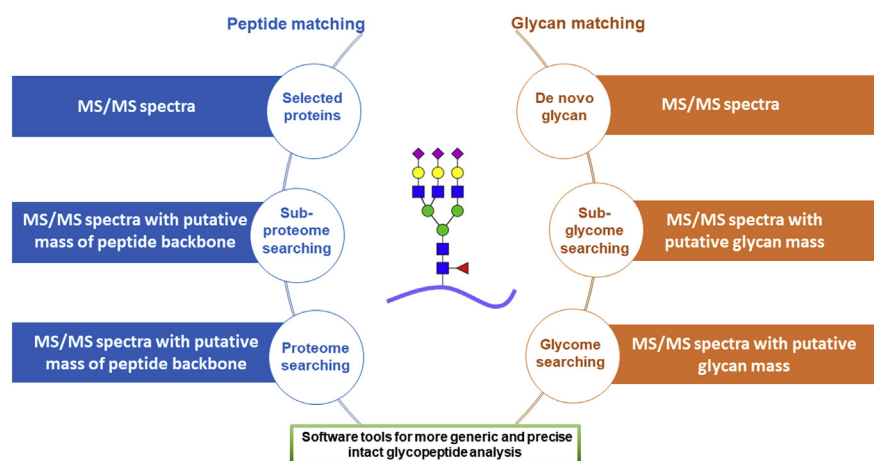
Correspondence

wqcao@fudan.edu.cn;
pyyang@fudan.edu.cn

In Brief

This article provides a systematic review of the most recent MS-based strategies and corresponding software tools for the analysis of intact glycopeptides, particularly intact N-glycopeptides, reported in the last decade, including the process of identifying N-glycopeptides from MS data, the existing methods of MS data acquisition and interpretation, the quality control methods, the display of results, and the software applications.

Graphical Abstract



Highlights

- Summarize the development of software tools for intact glycopeptide analysis.
- Survey the MS acquisition strategies for glycopeptide spectrum acquisition.
- Conclude the quality control methods for glycopeptide-spectrum matching.
- Show the application of intact glycopeptide analysis in glycoproteomic research.



Recent Advances in Software Tools for More Generic and Precise Intact Glycopeptide Analysis

Weiqian Cao^{1,2,3,*}, Mingqi Liu¹, Siyuan Kong¹, Mengxi Wu^{1,4}, Yang Zhang^{1,3}, and Pengyuan Yang^{1,2,3,4,*}

Intact glycopeptide identification has long been known as a key and challenging barrier to the comprehensive and accurate understanding the role of glycosylation in an organism. Intact glycopeptide analysis is a blossoming field that has received increasing attention in recent years. MS-based strategies and relative software tools are major drivers that have greatly facilitated the analysis of intact glycopeptides, particularly intact N-glycopeptides. This article provides a systematic review of the intact glycopeptide-identification process using MS data generated in shotgun proteomic experiments, which typically focus on N-glycopeptide analysis. Particular attention is paid to the software tools that have been recently developed in the last decade for the interpretation and quality control of glycopeptide spectra acquired using different MS strategies. The review also provides information about the characteristics and applications of these software tools, discusses their advantages and disadvantages, and concludes with a discussion of outstanding tools.

Protein glycosylation has long been known as a heterogeneous post-translational modification generating greater protein diversities than other post-translational modifications (1). These diversities alter the functions of proteins and subsequently exert a profound effect on various biological processes (2, 3). The ability to perform an in-depth and precise identification of glycoproteins is the key to obtaining a comprehensive and accurate understanding of the role of glycosylation in an organism.

MS has been recognized as an ideal tool for glycoprotein analysis since the first report in 1978 (4). MS-based glycoproteomics, which emerged with the goal of defining the glycoproteome of a biological system, has made strides in deglycosylation-centric glycoproteomic studies in the past few decades (5–9). Previously, enriched glycopeptides were usually deglycosylated to generate an identifiable mass tag on

glycosylation sites before the MS analysis that can be directly searched using routine proteomic software. Although the use of the deglycosylation-centric strategy as a compromise greatly simplified the MS identification and expanded our knowledge of glycoproteins, the site-specific glycan information was lost during the removal of glycans from glycopeptides. Subsequently, a more challenging issue in the high-throughput and precise identification of intact glycopeptides (also known as site-specific glycan analysis) was identified to better assess the biological attributes of the glycosylation.

Since then, an analysis of intact glycopeptides has been a challenge for several reasons. First, the inherently low abundance and ionization efficiency of glycopeptides make them difficult to identify using MS. Second, MS dissociation energy for glycans and peptides differ, hampering the acquisition of informative glycopeptide spectra using the single MS fragmentation method. Third, glycopeptides, which comprise a group of various types of monosaccharides linked to amino acid residues with different compositions and linkages, are so complicated that routine proteomic search engines are not qualified for an analysis of such enormous numbers of variable modifications. Many robust and mature glycopeptide enrichment methods developed in the past decades have dramatically improved the abundance of glycopeptides in MS analysis and considerably reduced the signal suppression of glycopeptide in the mass spectra by nonglycopeptides. Subsequently, the urgent needs for intact glycopeptide analysis are the ability to obtain informative glycopeptide spectra data and reliably decipher those data with expert software.

In recent years, substantial progress has been achieved in addressing the immediate needs for an analysis of intact glycopeptides, particularly N-glycopeptides (10). Some generic analytical pipelines and useful software became available and enabled a high-throughput and precise analysis of intact glycopeptides. Through intact glycopeptide analyses, glycopeptide sequences with attached glycan composition

From the ¹The Fifth People's Hospital of Fudan University and Institutes of Biomedical Sciences, ²NHC Key Laboratory of Glycoconjugates Research, ³The Shanghai Key Laboratory of Medical Epigenetics and the International Co-Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, ⁴Department of Chemistry, Fudan University, Shanghai, China

*For correspondence: Weiqian Cao, wqcao@fudan.edu.cn; Pengyuan Yang, pyyang@fudan.edu.cn.

can be conceivably gained. This review will specify the most recent MS-based strategies and corresponding software tools for the analysis of intact glycopeptides, particularly intact N-glycopeptides, reported in the last decade, while reviewing the process of identifying N-glycopeptides from MS data, the existing methods of MS data acquisition and interpretation, the quality control methods, the display of results, and the software applications (Table 1).

MS FRAGMENTATION METHODS FOR GLYCOPEPTIDES

In the past 10 years, software tools for the automatic identification of intact glycopeptides using different LC-MS/MS strategies were developed. Regardless of the type of strategy used, sufficient fragments from both peptide backbones and the attached glycans are required. Informative glycopeptide mass spectra are difficult to acquire because the fragmentation varies widely between the two components because of the differences in the physical and chemical properties of the glycosidic bonds and peptide bonds. Dissociation methods, including resonance activation (ion trap) collision-induced dissociation (CID), beam-type (quadrupole-time of flight [Q-TOF]) CID, higher-energy collision dissociation (HCD), and electron-transfer dissociation (ETD), are the most commonly used approaches in glycopeptide identification.

CID Fragmentation

Mass spectra generated by ion-trap CID are dominated by Y-ions but contain a few B-ions and a limited number of b/y-ions because of the single-bond cleavages of precursor ions, low dissociation energy, and the “1/3” cutoff in the low m/z region (11), which can be used to identify the conjugated glycans but not for confident identification of the location of the glycosylation site and peptide sequences, particularly in complex samples. Thus, this technique has only been applied to analyze glycopeptides in simple samples in some previous studies (12–15).

HCD Fragmentation

In contrast, HCD tends to produce abundant diagnostic oxonium ions, such as m/z 138.055, 204.087 and 366.140 fragments. It also yields B- and Y-ions for glycan identification and partial b/y-ion series for peptide identification. Studies showed an increase in the b/y-ion series as the HCD collision energy increased (16). Different collision energies in HCD-MS/MS produce complementary fragments of the glycan and peptides (17, 18). GlycoFinder used HCD collision energies of 30% and 50% to analyze glycopeptide in simple samples (19). The fragmentation patterns produced by the beam-type CID in the Q-TOF spectra resemble the patterns produced by HCD in C-trap on Orbitrap mass spectrometers.

Many studies have investigated the behaviors of glycopeptide fragmentation using different dissociation methods

(20–25). In 2016, we used various MS/MS collision parameters, including CID and HCD, each with nine different energies, as well as electron-transfer/collision-induced dissociation/electron-transfer/higher-energy collision dissociation. Stepped collision energies (SCEs)–HCD–MS/MS at 20–30–40% energies generated the most informative and abundant fragment ions for both the glycan and peptide of a glycopeptide in a single spectrum (26). Since then, SCE–HCD–MS/MS has been widely used in high-throughput identification of intact glycopeptides (27, 28).

ETD Fragmentation

ETD primarily breaks N-C_α bonds and generates c/z-ions. Ideally, ETD generates abundant c/z-ions from peptide backbone fragments, which are very useful for the identification of glycosylation sites and peptide sequences. However, ETD often suffers from incomplete fragmentation, leading to a large amount of residual precursor ions (29). Thus, ETD is rarely used alone and is frequently combined with HCD or CID methods for glycopeptide identification. With the development of the ion-trigger technique in MS, HCD–product-dependent (PD)–ETD increases the efficiency of combining fragments because ion-trigger technology generally avoids the reaction time of ETD wasted on nonglycopeptide dissociation. Hybrid dissociation techniques developed for an Orbitrap Fusion mass spectrometer, such as electron-transfer/collision-induced dissociation (30, 31), electron-transfer/higher-energy collision dissociation (29, 32, 33), and activated ion electron transfer dissociation (34, 35), where a supplementary energy source is applied to all fragment ions formed by ETD, have shown great potential in N-glycopeptide identification (11, 35–39).

BOTTOM-UP EXPERIMENTAL AND MS ACQUISITION STRATEGIES

In a typical bottom-up intact glycopeptide-identification strategy, glycopeptides are first enriched to reduce the complexity and increase the detectability of glycopeptides. Then, enriched intact glycopeptides are analyzed using different LC-MS/MS strategies and interpreted with different software tools. After surveying the development of software tools for MS-based glycopeptide-identification strategies in last 10 years, two major types of software have been used according to the experimental strategy with or without a deglycosylation procedure: one is the software combining the results from two LC-MS/MS analyses (one with deglycosylated peptides and another with native intact glycopeptide) and the other is the software that directly identifies mass spectra from an intact glycopeptide analysis.

Experimental and MS Acquisition Strategy With the Deglycosylation Procedure

Generally, in the experimental strategy with the deglycosylation step, two experimental workflows, deglycosylated

TABLE 1
A summary table of the most recent MS-based strategies and corresponding software tools for the analysis of intact N-glycopeptides

Pipeline	Spectral collection				Scoring/matching				Quality control		
	MS/MS		MS ³ /de- glycopeptide	Peptide		Glycan		Diagnostic ions	Peptide	Glycan	Intact glycopeptide
	HCD	CID		ETD	Method	Database	Method				
GlycoFragwork, etc. (Tang's laboratory)	BY-C	Y-a	c/z		Probability matches	Selected proteins	<i>De novo</i>		Glycan classification	Decoy sequence	
GlycoPep Detector, etc. (Desaire's laboratory)		Y-a	c/z		Probability matches	Selected proteins	Probability matches	Manual input		Decoy sequence	
GlycoPeptide Search		Y-a	b/y		Existing algorithm	Selected proteins	Core Y-ions matches	GlycomeDB	Spectral filtration	Decoy sequence	
GP Finder	Y-a				Probability matches	Selected proteins	Probability matches	Glycomic analysis	Spectral filtration		Decoy precursor mass
Sweet-Heart, etc. (Khoo's laboratory)	BY-c	Y-a	c/z	b/y from MS ³	Existing algorithm	Proteome	Probability matches	Semi de-novo	Product-trigger	Decoy sequence	
ArMone	Y-c			De-glyco needed ^b	Mass compare	Sub-Proteome ^e	Core Y-ions matches	GlycomeDB			
GRIP	Y-c	Y-a		De-glyco needed ^b	Mass compare	Sub-Proteome ^e	Probability matches	GlycomeDB			
GPQuest	Y-a			De-glyco needed ^c	Spectral library	Sub-Proteome ^e	Mass compare	Glycomic analysis	Spectral filtration	Decoy sequence	
GlycoFinder	BY-a				Mass compare	Selected proteins	Core Y-ions matches	Manual input	Spectral filtration		
GlycoMaster DB	Y-a		c/z		Existing algorithm	Proteome	Probability matches	GlycomeDB	Spectral filtration	Decoy sequence	
MAGIC	B b/y	Y-a			Existing algorithm ^d	Proteome	Core Y-ions matches	Semi de-novo	Spectral filtration	Decoy sequence	
I-GPA, etc. (Yoo's laboratory)	B b/y	Y-a			Probability matches	Sub-Proteome ^f	Probability matches	GlycomeDB	Glycan classification		Decoy glycopeptide
GlycoPAT	b/y	Y-a	c/z		Probability matches	Selected proteins	Probability matches	Semi de-novo	Spectral filtration		Decoy glycopeptide
SugarQB	B b/y	Y-a			Existing algorithm ^d	Proteome	Core Y-ions matches	Manual input	Spectral filtration	Decoy sequence	
glyXtoolMS	B b/y	Y-a			Probability matches	Selected proteins	Core Y-ions matches	GlyTouCan	Glycan classification		
pMatchGlyco	B b/y	Y-a		De-glyco needed ^c	Spectral library	Sub-Proteome ^e	Mass compare	GlycomeDB			Decoy glycopeptide
GPSeeker	b/y	Y-a			Existing algorithm	Proteome	Probability matches	Glycomic analysis	Spectral filtration	Decoy sequence	Decoy fragment ⁹

TABLE 1—Continued

Pipeline	Spectral collection			Scoring/matching				Quality control		
	MS/MS		MS ³ /de- glycopeptide	Peptide		Glycan		Peptide	Glycan	Intact glycopeptide
	HCD	CID		ETD	Method	Database	Method			
O-Search	BY-c			Existing algorithm ^d	Proteome	Core Y-ions matches	Semi de-novo	Spectral filtration	Decoy sequence	
pGlyco	BY-c	Y-a	b/y from MS ³	Existing algorithm	Proteome	Probability matches	GlycomeDB	Product-trigger	Decoy sequence	Decoy fragment
pGlyco 2.0	B b/y	Y-a	Analyzed ^a	Existing algorithm	Proteome	Probability matches	GlycomeDB	Spectral filtration	Decoy sequence	Decoy fragment

CID, collision-induced dissociation; ETD, electron-transfer dissociation; HCD, higher-energy collision dissociation; Y-a, all Y ions; Y-c, Y ions from N-glycan core.

^aDifferent collisional energies.

^bPregenerated de-glycopeptide database.

^cPregenerated de-glycopeptide spectra.

^dSubstitution of precursor mass.

^eCandidates from de-glyco experiment.

^fCandidates from the sample-dependent database.

^gThe decoy method from pGlyco.

peptide analysis and intact glycopeptide analysis, are performed. A spectral library of peptides containing glycosylation sites in the samples is first built by analyzing the isolated deglycosylated peptides using HCD/CID LC-MS/MS. Spectra of intact glycopeptides are selected using diagnostic oxonium ions. The glycan composition at a glycosylation site is determined by matching the mass difference between the precursor ion of the intact glycopeptide and the deglycosylated peptide to the glycan database. Three main software tools support this identification strategy: GRIP developed by us in 2014 (40), ArMone 2.0 developed by Cheng *et al.* in 2014 (41), and GPQuest developed by Toghi Eshghi *et al.* in 2015 (42). Although GRIP, ArMone, and GPQuest have been used for high-throughput profiling of intact N-glycopeptides in complex samples, the two workflows make the experimental operation tedious. GlycoSeq developed by Yu *et al.* in 2016 (43), which aims to identify glycopeptides in isolated glycoproteins from their CID spectra alone, also requires a preliminary step for identifying the deglycosylated sites.

Experimental and MS Acquisition Strategy Without the Deglycosylation Procedure

In the experimental strategy without the deglycosylation step, software tools were developed to directly identify intact glycopeptides in mass spectra. The software tools were designed either for the interpretation of MS/MS data from one type of fragmentation method or for MS/MS data from multiple fragmentation methods. In early years, many software tools were expertly designed for a specific dissociation-based LC-MS/MS analysis of glycopeptides. For example, GlycoPeptide Search (14, 15), GP Finder (12), and MAGIC (44) were all developed for a Q-TOF-CID MS/MS analysis, in which GP Finder was specifically designed for a nonspecific analysis of digested glycopeptides using Pronase. ProteinProspector (45, 46) and GlycoPep Detector (47) were designed for an ETD-MS/MS analysis. GlycoPep Grader (13), which was developed by the same laboratory as GlycoPep Detector, was used for CID-MS/MS analyses. Although the experimental procedures of these software-supported analytical strategies are easy to operate, information about both glycan and peptide fragments in a spectrum obtained from a single dissociation method is insufficient and substantially limits the throughput and reliability of the identification. All of these software tools, except MAGIC (44), are only suitable for an analysis of simple samples.

Combined dissociation methods have been applied to analyze glycopeptides and obtain informative glycopeptide mass spectra. HCD, ETD, and ion-trap CID are often combined to dissociate the same precursor ions because of their complementary nature when applied to N-glycopeptide dissociation. Corresponding software tools were developed for deciphering a precursor with multiple fragment types in MS/MS data. Byonic, which is a commercial and most commonly used software (48), supports searching MS/MS

data with multiple fragment types. gFinder (49) and I-GPA (50) developed by Yoo *et al.* were both designed to interpret CID/HCD-MS/MS spectra. GlypID 2.0 improves upon the GlypID tool (51) developed by Tang *et al.* and supports searches for HCD and CID fragmentation ions. GlycoFragwork developed by the same laboratory characterizes glycopeptide in biological proteome samples by scoring complementary fragmentation techniques, including CID, HCD, and ETD. GlycoPAT (52) supports searches of CID/ETD-MS/MS or CID/HCD/ETD-MS/MS data. GlycoMaster DB (53) was designed to decipher HCD/ETD combined with MS/MS data and HCD-MS/MS data alone. Sweet-Heart, a series of software tools developed by Khoo *et al.* substantially improved the analytical procedure of glycopeptide identification from multiple fragmentation mass data. Sweet-Heart was first proposed in 2013 as an integrated suite of computational tools enabling the sequencing and identification of glycopeptides from low-resolution and low accuracy CID-MS² and targeted-CID-MS³ (54). In 2014, taking advantage of the new trihybrid Orbitrap configuration, the authors experimented with adding a parallel ion-trap CID-MS² data acquisition mode to the HCD-PD-ETD function and optimized Sweet-Heart for this analytical procedure (55).

The combination of HCD/CID with ETD is one of the widely used methods in multiple dissociation strategies. However, the long reaction time and low dissociation efficiency of ETD have limited the throughput of glycopeptide analysis. Recently, a preferable strategy using SCE-HCD-MS/MS and a dedicated search engine pGlyco 2.0 (26) to completely utilize the abundant information in SCE-HCD-MS/MS spectra were developed, enabling a high-throughput analysis of N-glycopeptides.

INTERPRETATION OF GLYCOPEPTIDE MASS SPECTRA

The method used to interpret complex tandem mass spectra generated from glycopeptides is the core of a search tool. Generally, processes used to interpret glycopeptide mass spectra include screening glycopeptide mass spectra and matching the glycan moiety and carrier peptide with glycopeptide spectra.

Selection of Glycopeptide Mass Spectra

A screen of glycopeptide mass spectra is preferred to avoid interference from nonglycosylated peptide spectra and improve the reliability of the identification. The abundant glycopeptide-specific oxonium ions obtained upon HCD are normally used to determine whether the spectrum is attributed to a glycopeptide. The HexNAc-derived fragment ions m/z 138.055, m/z 204.087, and m/z 366.140 derived from Hex+-HexNAc are mostly used as diagnostic ions for the selection of glycopeptide spectra. These diagnostic ions have been utilized either “online” in the “HCD-PD” mode to enable a more efficient time of dissociation of glycopeptides instead of nonglycopeptides in the subsequent triggered ETD or CID or

“offline” as a diagnostic ion assistant for the software-based selection of glycopeptide spectra. In addition, fragment ions m/z 292.103/274.092 and m/z 308.098/290.087 for NeuAc/NeuAc-H₂O and NeuGc/NeuGc-H₂O, respectively, are specific for sialic acid-containing glycopeptides. Other diagnostic oxonium ions, such as m/z 109.028, m/z 115.035, m/z 127.039, and 145.050 from fragments of Hex, m/z 126.055, m/z 144.066, m/z 168.066, and m/z 186.076 from fragments of HexNAc, m/z 657.140 from Hex+HexNAc+NeuAc, and m/z 673.230 from Hex+HexNAc+NeuGc, are also used alone or in combination for either the selection of glycopeptide spectra or glycan matching.

Another strategy for selecting glycopeptide spectra is based on precursor ions (56–58). Isotope-targeted glycoproteomics (IsoTag), in which metabolic labeling of glycoproteome is combined with the chemical enrichment and isotopic recoding of glycopeptides with a dibromide motif developed by Woo *et al.*, (58) is the most representative approach for that strategy. The natural abundance of the stable isotopes, ⁷⁹Br and ⁸¹Br (1:1), were used in the study by Palaniappan *et al.* (57) to provide a ready source for isotope recoding and convoluting the mass envelope of a glycopeptide with a triplet signature (1:2:1(M, M + 2, M + 4)), which can be detected computationally using a pattern-recognition algorithm developed in-house.

Matching of the Glycan Moiety and Carrier Peptide

Computational tools tackle the matching of the glycan moiety and carrier peptides using different approaches probably according to the different experimental and MS acquisition strategies they used (Fig. 1). Generally, for an analysis of a glycan moiety, most tools perform probability matches against the glycan database from GlycomeDB (currently merged into GlyTouCan). A few tools developed semi *de novo* glycan profiling algorithms; for carrier peptide analysis, software tools perform probability matching against select protein sequences. When a complete proteome database is used in a large-scale analysis, the incorporation of existing proteomic search engines (*i.e.*, Mascot, Sequest, and pFind) is a routine approach (59, 60).

The approach for the strategy with the deglycosylation step considers the carrier peptide as a predefined mass increment according to their pre-established spectral library of peptides containing glycosylation sites to deduce the glycan moiety from glycopeptide spectra with different glycan matching strategies. The core of GRIP is a network-centric algorithm for deciphering glycan fragmentation CID MS/MS spectra (40). ArMone 2.0 identifies glycans by searching a constructed glycan database with a calculated molecular weight of glycans and MS/MS spectra of the glycopeptide, while only scoring the pentasaccharide core structure observed in the N-glycosylation (Man₃GlcNAc₂) of mammalian proteins. GPQuest assigns the glycan composition by matching the mass difference between the precursor ion of the intact glycopeptide

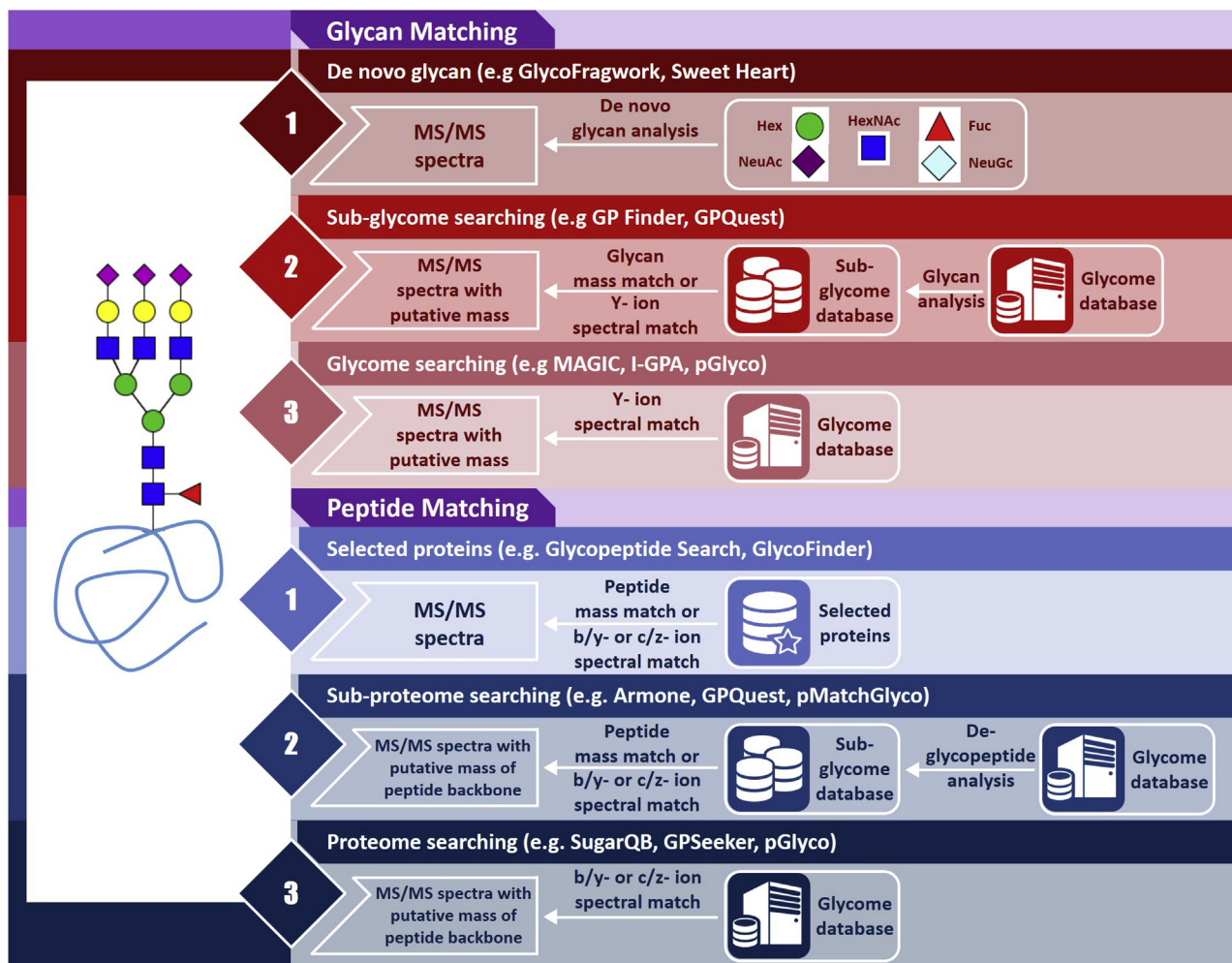


FIG. 1. General approaches for matching of the glycan moiety and carrier peptide.

and the glycosite-containing peptide to a glycan database (42). It is practically impossible to distinguish two glycan compositions with the same molecular weight when deducing the glycan moiety based only on the molecular weight. The same limitation also exists when directly matching the carrier peptide from the glycopeptide spectra with the peptide library, although common retention times have been used to filter between the two experiments (identification of the intact glycopeptide and deglycosylated peptide) in ArMone 2.0.

The interpretation approach for the strategy using combined fragmentation methods without the deglycosylation step generally assigns the glycan composition from acquired CID-MS/MS spectra that are rich in glycan fragment ions and define the carrier peptide sequences by an additional mode of MS² or MS³ analysis (61). The corresponding computational tools mentioned in the “[Experimental and MS Acquisition Strategy Without the Deglycosylation Procedure](#)” section all use this approach to deduce the glycan composition and peptide sequences of glycopeptides. For example, Sweet-Heart identifies the glycan composition in low-resolution and

low-accuracy ion-trap MS² data and putative peptide backbone in targeted MS³ data (54).

The approaches used by software tools to interpret the single fragmentation patterns differ. Some mandatorily detect the glycan and peptide ions in one spectrum. Meanwhile, other search tools, such as MAGIC (44) and SugarQB (62) for HCD/Q-TOF spectra interpretation, detect Y-ion patterns from the fragmentation of the common trimannosyl core of N-glycopeptides, and search the subsequent peptide sequence using common database search engines, such as Mascot, through the generation of *in silico* spectra by overwriting the original precursor with the naked peptide m/z and removing all of the glycan-related ions.

A superior interpretation strategy was used by pGlyco 2.0 (26), which was developed for the interpretation of abundant information in SCE-HCD-MS/MS spectra. It performed an integrated open search of each spectrum: a spectrum was first scored against the glycome database to identify the glycan candidates and then scored against the proteome database to identify the peptide candidates. This process

substantially increased the efficiency and accuracy of the interpretation.

QUALITY CONTROL METHODS FOR GLYCOPEPTIDE-SPECTRUM MATCHING

The false discovery rate (FDR) is an important measurement and a crucial parameter used to define the specificity of the identification and quality control of results. Since this technique was first proposed by Gygi in proteomics (63, 64), the target-decoy approach has been adopted as an acceptable method to calculate the FDRs for glycopeptides.

Before 2016, almost all identification software estimated the FDR of glycopeptide identification only at the peptide level by using a decoy database created based on the target protein sequence database. For example, GPQuest (42) creates a decoy database by randomly arranging amino acid sequences of those glycosite-containing peptides identified in their two-step experimental strategy with the same length as the target database. GlycoPep Evaluator (65) is a software program designed for the FDR analysis of small datasets that generates a decoy database with fixed mass of attached glycan for each glycopeptide in a 1:20 target-to-decoy ratio, according to a user-specified mass tolerance interval.

However, a high FDR would be obtained for glycopeptide identification results even if the peptide-spectrum match score is high because the FDR is only calculated at the peptide level and does not control for glycan identification. For example, Wu *et al.* (55) manually compared the results of HCD spectra identified using Sweet-Heart and Byonic (48) and validated as high as 37% false positives from 551 positive spectrum matches despite a claim of zero FDR using the Byonic criteria because they lack an FDR estimation of the glycan identity. As shown in the study by us, Byonic might substantially underestimate the metric of FDR for glycopeptide identification (26).

Researchers also realized that an FDR evaluation at all three levels of matches to glycans, peptides, and glycopeptides is needed for comprehensive quality control (66). GP Finder (12) was a pioneer program attempting to resolve the problem of glycan decoy database construction. This software mainly focuses on identifying glycopeptides in simple samples. The issues with the mathematical model and algorithm of FDR for glycopeptide identification remained unsolved until pGlyco developed a novel target-decoy approach to estimate the FDR of the glycan identification. The scoring scheme for glycan identification by pGlyco was a revised version of the previously reported algorithm for the CID-MS/MS spectral analysis of glycopeptides (40). pGlyco 2.0 conducts comprehensive quality control, including an FDR analysis of glycans and peptides, and uses a new model for the glycopeptide FDR estimation (Fig. 2A). In addition, a new quantitative analysis pipeline using metabolically labeled glycoproteome samples was established to validate the FDR reported by the pGlyco 2.0 (Fig. 2B) (26).

MASS SPECTRUM ANNOTATION AND RESULT DISPLAY

Automated annotation of glycopeptide spectra and a proper result display greatly facilitate subsequent analyses of intact glycopeptides. However, a range of previously developed software tools, including many well-designed search engines such as Sweet-Heart (54), ArMone (41), and GPQuest (42), do not include these functions. Some subsequently developed software tools, such as pGlyco (61), glyXtoolMS (67), and GPSeeker (68), conduct simple mass spectrum annotation steps. A web server, MAGIC-web (69), which was established by the MAGIC (44) software development team, enables mass spectrum annotation and glycan structure visualization. This function was designed as one of the functional modules on the web server and not integrated into the search engine of MAGIC and MAGIC+, requiring users have to complete additional operation themselves.

The dedicated search engine pGlyco 2.0 (26) considered that issue. In pGlyco 2.0, all identified spectra were automatically annotated and displayed by the software tool gLabel embedded in pGlyco 2.0, which greatly facilitates manual verification and data analysis (Fig. 3A). Another impressive tool for visualizing results was proposed by Riley *et al.* (35). The authors developed several approaches to visualize identified glycopeptides and explore profiles of heterogeneity present at multiple levels of proteomic information, from glycosites to subcellular regions (Fig. 3, B and C), using a dataset identified by an activated ion electron transfer dissociation-enabled method, which proved useful for analyzing intact glycopeptides in future studies. In addition, some specialized glycopeptide mass spectrum annotation and visualization tools, such as PepSweetener (70) and gpAnnotate (71), were also designed to facilitate the identification and analysis of results.

SOFTWARE APPLICATIONS

Advanced and dedicated software tools for high-throughput glycopeptide identification and efficient identification strategies have facilitated the increase in the N-glycoproteome coverage in recent years (48, 62, 72–81). The intact glycopeptide identification scale increased by three orders of magnitude from 2010 to the present. Before 2014, only tens to hundreds intact glycopeptides were identified because almost all identification strategies and software tools were only suitable for glycopeptide analysis in simple or standard glycoprotein sample and the analytical throughput was low. For example, only 45 N-glycopeptides were identified using GlycoPep Grader in 2012 (13). Pompach *et al.* (15) identified 57 glycopeptides in haptoglobin and 14 glycopeptides using GlycoPeptide Search. The development of advanced search tools, more powerful instrumentation, and well-designed MS acquisition methods have increased the scale of identification from approximately 100 (42, 54, 82) to more than 1000 (40, 53, 76, 83). One of the most impressive and representative studies

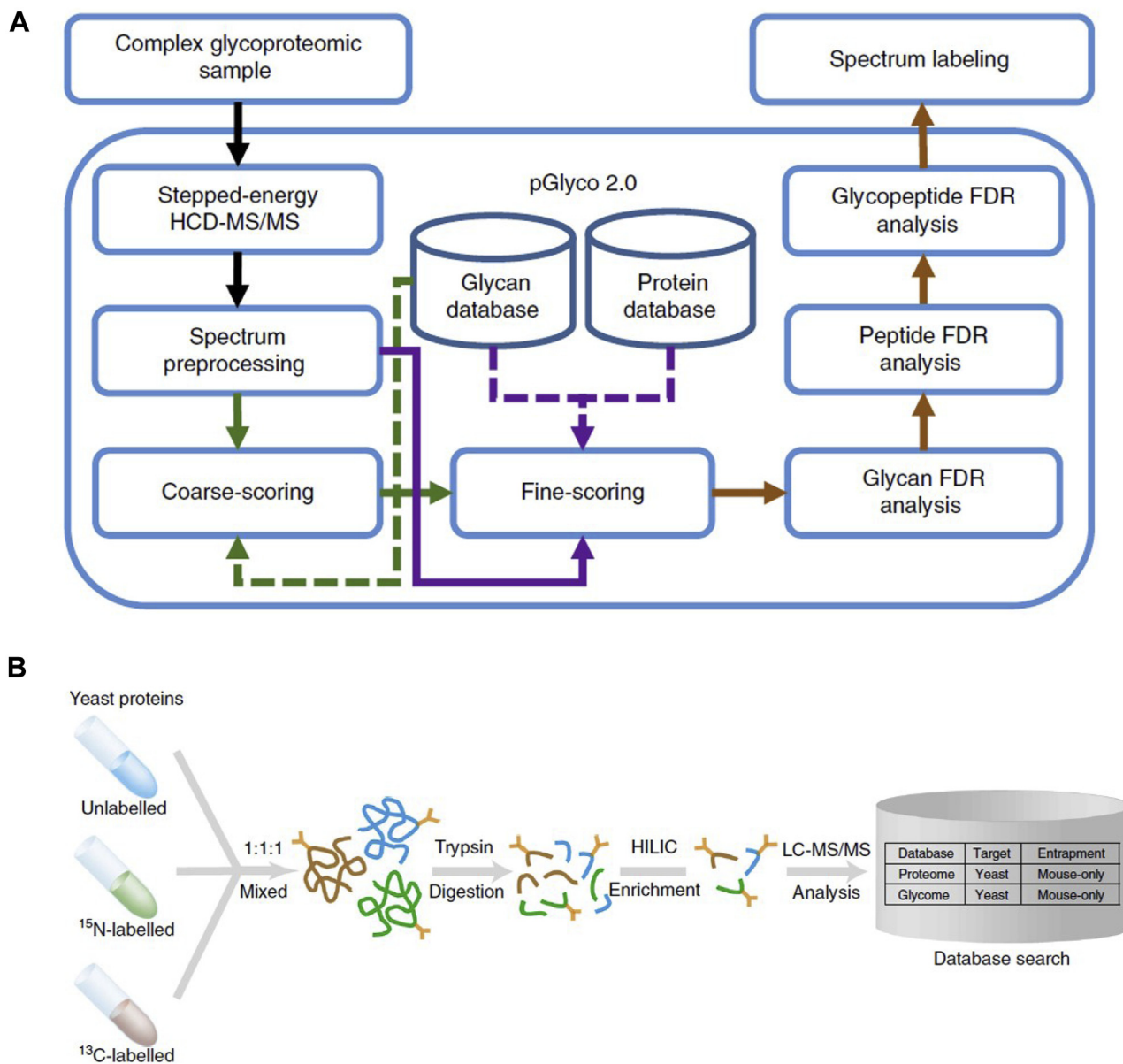


FIG. 2. **A dedicated software pGlyco 2.0 for site-specific glycan interpretation.** A, the flowchart of pGlyco 2.0. B, the FDR validation workflow. Reprinted with permission from Ref (26), under Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>). FDR, false discovery rate.

was published in 2016 by Sun *et al.* (83), who used GPQuest software and described an innovative method called solid-phase extraction of N-linked glycans and glycosite-containing peptides for N-glycopeptide identification. The authors successfully identified 1562 N-glycopeptides in different cell lines using the GPQuest software they developed in 2015 (42), which was a highlight and substantial achievement at that time.

In 2017, a major breakthrough was achieved by pGlyco 2.0 in the global characterization of intact N-glycopeptides (10, 26). The powerful search engine pGlyco 2.0 with

comprehensive quality control and the ingenious experimental strategy identified more than 10,000 glycopeptides and reached the deepest and largest scale ever reported. A total of 10,009 intact N-glycopeptides were identified in five mouse tissues. The generic, precise, and open-access characters (download freely from <http://pfind.ict.ac.cn/software/pGlyco/index.html>) attracted increasing attention to pGlyco 2.0, and this method adopted by an increasing number of studies (28, 84–95). For instance, pGlyco 2.0 was used to identify 3524 intact N-glycopeptides in the APP/PS1 mouse model of Alzheimer's disease and WT mice. Based on the further

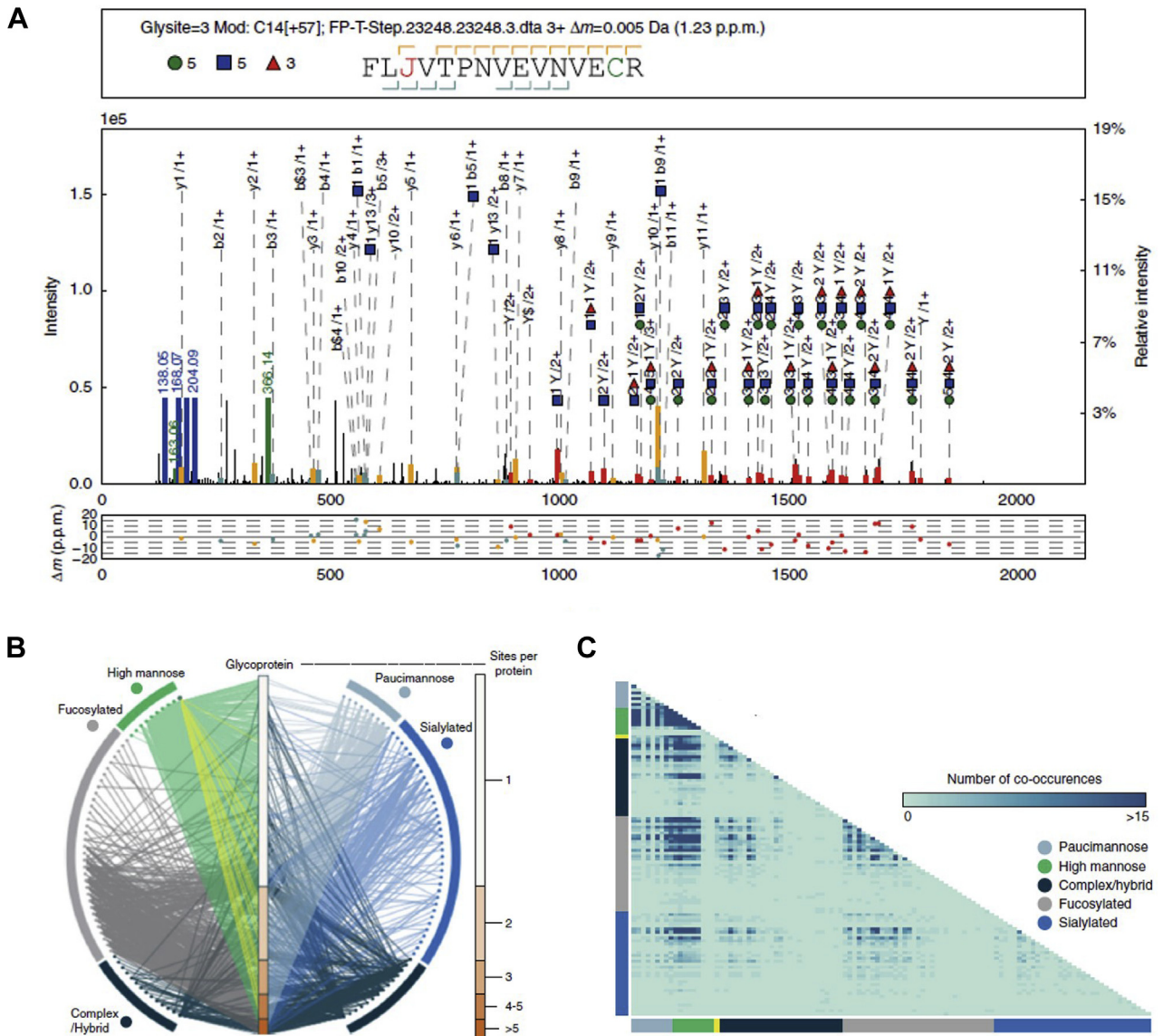


FIG. 3. **Glycopeptide spectrum annotation and result display.** A, automated annotation of glycopeptide spectra by pGlyco 2.0. B, glycoprotein–glycan network maps. C, a glycan co-occurrence heat map represents the number of times glycan pairs appeared together at the same glycosite. Reprinted with permission from Ref (26) and Ref (35), under Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

integration and identification of N-glycopeptides, proteome and lectin microarray data for glycan epitopes and dysfunctions of N-glycoproteins were shown to affect the development of Alzheimer’s disease (27). Using pGlyco 2.0, Zhang *et al.* (92) established an in-depth N-glycopeptide profile of human plasma and identified 1644 intact N-glycopeptides.

Byonic has become commercially available around 2012 (48). Since then, it has been widely used in glycopeptide identification (96–100). In addition, Byonic enables the user to perform glycopeptide searches by specifying glycan masses using Byonic’s usual modification fine-control mechanism, which greatly facilitates many studies (101, 102).

CONCLUSIONS AND FUTURE PERSPECTIVES

Powerful software tools are tremendously advancing the depth and precision of the identification of intact glycopeptides at a rapid rate and will greatly facilitate a comprehensive and accurate understanding of the role of glycosylation in organisms. Here, we reviewed the recent advances in software tools for MS-based identification of intact glycopeptides, typically for intact N-glycopeptides (Table 1). Throughout its development in the last decade, the field of N-glycopeptide analysis recently advanced in the last 5 years, changing the status of software development that lagged behind several years ago (103).

Glycopeptide analysis is moving toward the next stage of more comprehensive and quantitative identification. In our view, however, several problems must be solved in this process. The first is to develop dedicated strategies and search engines for intact O-glycopeptides. The problem of identifying O-glycopeptides are many folds, such as lack of efficient enrichment method, hard to obtain informative O-glycopeptide mass spectra. Although strategies and interpretation tools for O-glycopeptides are emerging (104–106), they are immature and not as generic and precise as software tools for N-glycopeptides. The second issue to be addressed is the accurate site-specific localization of the glycan, particularly for O-glycopeptides, because several glycosylation sites occur adjacently in one digested O-glycopeptides, which presents a substantial challenge for both MS identification and software interpretation. The third is that generic software tools supporting both MS and tandem MS quantitation of intact glycopeptides urgently needed, which is extremely important for explorations of the roles of glycosylation in different biological and pathological stages. Fourth and importantly, because only the composition and a small degree of topology (such as fucose being on chitobiose core, etc.) can be generated by the current glycopeptide analysis approaches, the analyses of glycan structures and glycan isoforms on glycopeptides are additional and important challenges due to the inherent complexity of glycans. Extensive developments in techniques and software tools are needed to achieve more exquisite and superb tools for intact glycopeptide characterization and are anticipated in the next 5 to 10 years.

Acknowledgments—This work was supported by grants from the National Natural Science Foundation of China Project (91853102), the National Key Research and Development Program (2016YFA0501303, 2018YFC0910300, and 2016YFB0201702), and the Innovative Research Team of High-level Local Universities in Shanghai.

Author contributions—W. C. wrote the manuscript and supervised this work. M. L. helped to draft the manuscript. S. K., W. M., and Y. Z. participated in reference collection. P. Y. supervised this work.

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: CID, collision-induced dissociation; ETD, electron-transfer dissociation; FDR, false discovery rate; HCD, higher-energy collision dissociation; PD, product dependent; Q-TOF, quadrupole-time of flight; SCE, stepped collision energy.

Published, MCPRO Papers in Press, February 6, 2021, <https://doi.org/10.1074/mcp.R120.002090>

REFERENCES

- Hart, G. W., and Copeland, R. J. (2010) Glycomics hits the big time. *Cell* **143**, 672–676
- Feizi, T., and Mulloy, B. (2003) Carbohydrates and glycoconjugates. Glycomics: The new era of carbohydrate biology. *Curr. Opin. Struct. Biol.* **13**, 602–604
- Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3–49
- Morris, H. R., Thompson, M. R., Osuga, D. T., Ahmed, A. I., Chan, S. M., Vandenheede, J. R., and Feeney, R. E. (1978) Antifreeze glycoproteins from the blood of an antarctic fish. The structure of the proline-containing glycopeptides. *J. Biol. Chem.* **253**, 5155–5162
- Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* **71**, 4431–4436
- Medzihradsky, K. F., Gillece-Castro, B. L., Settineri, C. A., Townsend, R. R., Masiarz, F. R., and Burlingame, A. L. (1990) Structure determination of O-linked glycopeptides by tandem mass spectrometry. *Biomed. Environ. Mass Spectrom.* **19**, 777–781
- Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907
- Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **21**, 667–672
- Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660–666
- 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**, 11–25
- Hu, H., Khatri, K., Klein, J., Leymarie, N., and Zaia, J. (2016) A review of methods for interpretation of glycopeptide tandem mass spectral data. *Glycoconj. J.* **33**, 285–296
- Strum, J. S., Nwosu, C. C., Hua, S., Kronewitter, S. R., Seipert, R. R., Bachelor, R. J., An, H. J., and Lebrilla, C. B. (2013) Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. *Anal. Chem.* **85**, 5666–5675
- Woodin, C. L., Hua, D., Maxon, M., Rebecchi, K. R., Go, E. P., and Desaire, H. (2012) GlycoPep grader: A web-based utility for assigning the composition of N-linked glycopeptides. *Anal. Chem.* **84**, 4821–4829
- Chandler, K. B., Pompach, P., Goldman, R., and Edwards, N. (2013) Exploring site-specific N-glycosylation microheterogeneity of haptoglobin using glycopeptide CID tandem mass spectra and glycan database search. *J. Proteome Res.* **12**, 3652–3666
- Pompach, P., Chandler, K. B., Lan, R., Edwards, N., and Goldman, R. (2012) Semi-automated identification of N-glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search. *J. Proteome Res.* **11**, 1728–1740
- Khatri, K., Staples, G. O., Leymarie, N., Leon, D. R., Turiak, L., Huang, Y., Yip, S., Hu, H., Heckendorf, C. F., and Zaia, J. (2014) Confident assignment of site-specific glycosylation in complex glycoproteins in a single step. *J. Proteome Res.* **13**, 4347–4355
- Cao, Q., Zhao, X., Zhao, Q., Lv, X., Ma, C., Li, X., Zhao, Y., Peng, B., Ying, W., and Qian, X. (2014) Strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry. *Anal. Chem.* **86**, 6804–6811
- Reiding, K. R., Bondt, A., Franc, V., and Heck, A. J. R. (2018) The benefits of hybrid fragmentation methods for glycoproteomics. *Trends Analyt. Chem.* **108**, 260–268
- Cao, L., Tolic, N., Qu, Y., Meng, D., Zhao, R., Zhang, Q., Moore, R. J., Zink, E. M., Lipton, M. S., Pasa-Tolic, L., and Wu, S. (2014) Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal. Biochem.* **452**, 96–102
- Hinneburg, H., Stavenhagen, K., Schweiger-Hufnagel, U., Pengelley, S., Jabs, W., Seeberger, P. H., Silva, D. V., Wührer, M., and Kolarich, D. (2016) The art of destruction: Optimizing collision energies in

- quadrupole-time of flight (Q-TOF) instruments for glycopeptide-based glycoproteomics. *J. Am. Soc. Mass Spectrom.* **27**, 507–519
21. Hoffmann, E.d., and Stroobant, V. (2007) *Mass Spectrometry: Principles and Applications*. J. Wiley, Chichester, West Sussex, England; Hoboken, NJ
 22. Kolli, V., Roth, H. A., De La Cruz, G., Fernando, G. S., and Dodds, E. D. (2015) The role of proton mobility in determining the energy-resolved vibrational activation/dissociation channels of N-glycopeptide ions. *Anal. Chim. Acta* **896**, 85–92
 23. Zhang, Y., Xie, X., Zhao, X., Tian, F., Lv, J., Ying, W., and Qian, X. (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
 24. Aboufazel, F., Kolli, V., and Dodds, E. D. (2015) A comparison of energy-resolved vibrational activation/dissociation characteristics of protonated and sodiated high mannose N-glycopeptides. *J. Am. Soc. Mass Spectrom.* **26**, 587–595
 25. Kuo, C. W., Guu, S. Y., and Khoo, K. H. (2018) Distinctive and complementary MS(2) fragmentation characteristics for identification of sulfated sialylated N-glycopeptides by nanoLC-MS/MS workflow. *J. Am. Soc. Mass Spectrom.* **29**, 1166–1178
 26. Liu, M. Q., Zeng, W. F., Fang, P., Cao, W. Q., Liu, C., Yan, G. Q., Zhang, Y., Peng, C., Wu, J. Q., Zhang, X. J., Tu, H. J., Chi, H., Sun, R. X., Cao, Y., Dong, M. Q., 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
 27. Fang, P., Xie, J., Sang, S., Zhang, L., Liu, M., Yang, L., Xu, Y., Yan, G., Yao, J., Gao, X., Qian, W., Wang, Z., Zhang, Y., Yang, P., and Shen, H. (2020) Multilayered N-glycoproteome profiling reveals highly heterogeneous and dysregulated protein N-glycosylation related to Alzheimer's disease. *Anal. Chem.* **92**, 867–874
 28. Zhang, Y., Zhao, W., Zhao, Y., Mao, Y., Su, T., Zhong, Y., Wang, S., Zhai, R., Cheng, J., Fang, X., Zhu, J., and Yang, H. (2019) Comparative glycoproteomic profiling of human body fluid between healthy controls and patients with papillary thyroid carcinoma. *J. Proteome Res.* <https://doi.org/10.1021/acs.jproteome.9b00672>
 29. Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**, 9668–9673
 30. Campbell, J. L., Hager, J. W., and Le Blanc, J. C. (2009) On performing simultaneous electron transfer dissociation and collision-induced dissociation on multiply protonated peptides in a linear ion trap. *J. Am. Soc. Mass Spectrom.* **20**, 1672–1683
 31. 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**, 5311–5318
 32. Chen, G., Zhang, Y., Trinidad, J. C., and Dann, C., 3rd (2018) Distinguishing sulfotyrosine containing peptides from their phosphotyrosine counterparts using mass spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 455–462
 33. Li, S., Zhou, Y., Xiao, K., Li, J., and Tian, Z. (2018) Selective fragmentation of the N-glycan moiety and protein backbone of ribonuclease B on an Orbitrap Fusion Lumos Tribrid mass spectrometer. *Rapid Commun. Mass Spectrom.* **32**, 2031–2039
 34. Riley, N. M., Hebert, A. S., Durnberger, G., Stanek, F., Mechtler, K., Westphall, M. S., and Coon, J. J. (2017) Phosphoproteomics with activated ion electron transfer dissociation. *Anal. Chem.* **89**, 6367–6376
 35. Riley, N. M., Hebert, A. S., Westphall, M. S., and Coon, J. J. (2019) Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat. Commun.* **10**, 1311
 36. Caval, T., Zhu, J., and Heck, A. J. R. (2019) Simply extending the mass range in electron transfer higher energy collisional dissociation increases confidence in N-glycopeptide identification. *Anal. Chem.* **91**, 10401–10406
 37. Chen, Z., Yu, Q., Hao, L., Liu, F., Johnson, J., Tian, Z., Kao, W. J., Xu, W., and Li, L. (2018) Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD). *Analyst* **143**, 2508–2519
 38. Parker, B. L., Thaysen-Andersen, M., Fazakerley, D. J., Holliday, M., Packer, N. H., and James, D. E. (2016) Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-alpha-induced insulin resistance in adipocytes. *Mol. Cell. Proteomics* **15**, 141–153
 39. Yu, Q., Wang, B., Chen, Z., Urabe, G., Glover, M. S., Shi, X., Guo, L. W., Kent, K. C., and Li, L. (2017) Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J. Am. Soc. Mass Spectrom.* **28**, 1751–1764
 40. Liu, M., Zhang, Y., Chen, Y., Yan, G., Shen, C., Cao, J., Zhou, X., Liu, X., Zhang, L., Shen, H., Lu, H., He, F., and Yang, P. (2014) Efficient and accurate glycopeptide identification pipeline for high-throughput site-specific N-glycosylation analysis. *J. Proteome Res.* **13**, 3121–3129
 41. Cheng, K., Chen, R., Seebun, D., Ye, M., Figeys, D., and Zou, H. (2014) Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J. Proteomics* **110**, 145–154
 42. 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
 43. Yu, C. Y., Mayampurath, A., Zhu, R., Zacharias, L., Song, E., Wang, L., Mechref, Y., and Tang, H. (2016) Automated glycan sequencing from tandem mass spectra of N-linked glycopeptides. *Anal. Chem.* **88**, 5725–5732
 44. Lynn, K. S., Chen, C. C., Lih, T. M., Cheng, C. W., Su, W. C., Chang, C. H., Cheng, C. Y., Hsu, W. L., Chen, Y. J., and Sung, T. Y. (2015) MAGIC: An automated N-linked glycoprotein identification tool using a Y1-ion pattern matching algorithm and in silico MS(2) approach. *Anal. Chem.* **87**, 2466–2473
 45. Baker, P. R., Trinidad, J. C., and Chalkley, R. J. (2011) Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteomics* **10**, M111.008078
 46. Medzihradsky, K. F., Kaasik, K., and Chalkley, R. J. (2015) Tissue-specific glycosylation at the glycopeptide level. *Mol. Cell. Proteomics* **14**, 2103–2110
 47. Zhu, Z., Hua, D., Clark, D. F., Go, E. P., and Desaire, H. (2013) GlycoPep detector: A tool for assigning mass spectrometry data of N-linked glycopeptides on the basis of their electron transfer dissociation spectra. *Anal. Chem.* **85**, 5023–5032
 48. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics*
 49. Kim, J. W., Hwang, H., Lim, J. S., Lee, H. J., Jeong, S. K., Yoo, J. S., and Paik, Y. K. (2016) gFinder: A web-based bioinformatics tool for the analysis of N-glycopeptides. *J. Proteome Res.* **15**, 4116–4125
 50. Park, G. W., Kim, J. Y., Hwang, H., Lee, J. Y., Ahn, Y. H., Lee, H. K., Ji, E. S., Kim, K. H., Jeong, H. K., Yun, K. N., Kim, Y. S., Ko, J. H., An, H. J., Kim, J. H., Paik, Y. K., et al. (2016) Integrated GlycoProteome analyzer (I-GPA) for automated identification and quantitation of site-specific N-glycosylation. *Sci. Rep.* **6**, 21175
 51. Wu, Y., Mechref, Y., Klouckova, I., Mayampurath, A., Novotny, M. V., and Tang, H. (2010) Mapping site-specific protein N-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **24**, 965–972
 52. Liu, G., Cheng, K., Lo, C. Y., Li, J., Qu, J., and Neelamegham, S. (2017) A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. *Mol. Cell. Proteomics* **16**, 2032–2047
 53. He, L., Xin, L., Shan, B., Lajoie, G. A., and Ma, B. (2014) GlycoMaster DB: Software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J. Proteome Res.* **13**, 3881–3895
 54. Wu, S. W., Liang, S. Y., Pu, T. H., Chang, F. Y., and Khoo, K. H. (2013) Sweet-Heart - an integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *J. Proteomics* **84**, 1–16
 55. Wu, S. W., Pu, T. H., Viner, R., and Khoo, K. H. (2014) Novel LC-MS(2) product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* **86**, 5478–5486
 56. Froehlich, J. W., Dodds, E. D., Wilhelm, M., Serang, O., Steen, J. A., and Lee, R. S. (2013) A classifier based on accurate mass measurements to aid large scale, unbiased glycoproteomics. *Mol. Cell. Proteomics* **12**, 1017–1025

57. Palaniappan, K. K., Pitcher, A. A., Smart, B. P., Spicciarich, D. R., Iavarone, A. T., and Bertozzi, C. R. (2011) Isotopic signature transfer and mass pattern prediction (IsoStamp): An enabling technique for chemically-directed proteomics. *ACS Chem. Biol.* **6**, 829–836
58. Woo, C. M., Iavarone, A. T., Spicciarich, 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
59. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989
60. Chi, H., Liu, C., Yang, H., Zeng, W. F., Wu, L., Zhou, W. J., Wang, R. M., Niu, X. N., Ding, Y. H., Zhang, Y., Wang, Z. W., Chen, Z. L., Sun, R. X., Liu, T., Tan, G. M., et al. (2018) Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **36**, 1059–1061
61. Zeng, W. F., Liu, M. Q., Zhang, Y., Wu, J. Q., Fang, P., Peng, C., Nie, A., Yan, G., Cao, W., Liu, C., Chi, H., Sun, R. X., Wong, C. C., He, S. M., and Yang, P. (2016) pGlyco: A pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Sci. Rep.* **6**, 25102
62. Stadlmann, J., Taubenschmid, J., Wenzel, D., Gattlinger, A., Durnberger, G., Dusberger, F., Elling, U., Mach, L., Mechtler, K., and Penninger, J. M. (2017) Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **549**, 538–542
63. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214
64. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2**, 667–675
65. Zhu, Z., Su, X., Go, E. P., and Desaire, H. (2014) New glycoproteomics software, GlycoPep evaluator, generates decoy glycopeptides de novo and enables accurate false discovery rate analysis for small data sets. *Anal. Chem.* **86**, 9212–9219
66. Shipman, J. T., Su, X., Hua, D., and Desaire, H. (2019) DecoyDeveloper: An on-demand, de novo decoy glycopeptide generator. *J. Proteome Res.* **18**, 2896–2902
67. Pioch, M., Hoffmann, M., Pralow, A., Reichl, U., and Rapp, E. (2018) glyXtool(MS): An open-source pipeline for semiautomated analysis of glycopeptide mass spectrometry data. *Anal. Chem.* **90**, 11908–11916
68. Xiao, K., and Tian, Z. (2019) GPSeeker enables quantitative structural N-glycoproteomics for site- and structure-specific characterization of differentially expressed N-glycosylation in hepatocellular carcinoma. *J. Proteome Res.* **18**, 2885–2895
69. Lih, T. M., Choong, W. K., Chen, C. C., Cheng, C. W., Lin, H. N., Chen, C. T., Chang, H. Y., Hsu, W. L., and Sung, T. Y. (2016) MAGIC-web: A platform for untargeted and targeted N-linked glycoprotein identification. *Nucleic Acids Res.* **44**, W575–W580
70. Domagalski, M. J., Alocci, D., Almeida, A., Kolarich, D., and Lisacek, F. (2018) PepSweetener: A web-based tool to support manual annotation of intact glycopeptide MS spectra. *Proteomics Clin. Appl.* **12**, e1700069
71. Cheng, K., Pawlowski, G., Yu, X., Zhou, Y., and Neelamegham, S. (2019) DrawGlycan-SNFG & gPAnnotate: Rendering glycans and annotating glycopeptide mass spectra. *Bioinformatics* **36**, 1942–1943
72. 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
73. Mayampurath, A., Song, E., Mathur, A., Yu, C. Y., Hammoud, Z., Mechref, Y., and Tang, H. (2014) Label-free glycopeptide quantification for biomarker discovery in human sera. *J. Proteome Res.* **13**, 4821–4832
74. Wu, J., Qin, H., Li, T., Cheng, K., Dong, J., Tian, M., Chai, N., Guo, H., Li, J., You, X., Dong, M., Ye, M., Nie, Y., Zou, H., and Fan, D. (2016) Characterization of site-specific glycosylation of secreted proteins associated with multi-drug resistance of gastric cancer. *Oncotarget* **7**, 25315–25327
75. Yang, G., Hu, Y., Sun, S., Ouyang, C., Yang, W., Wang, Q., Betenbaugh, M., and Zhang, H. (2018) Comprehensive glycoproteomic analysis of Chinese hamster ovary cells. *Anal. Chem.* **90**, 14294–14302
76. Shu, Q., Li, M., Shu, L., An, Z., Wang, J., Lv, H., Yang, M., Cai, T., Hu, T., Fu, Y., and Yang, F. (2020) Large-scale identification of N-linked intact glycopeptides in human serum using HILIC enrichment and spectral library search. *Mol. Cell. Proteomics* **19**, 672–689
77. 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**, e1700436
78. Xu, F., Wang, Y., Xiao, K., Hu, Y., Tian, Z., and Chen, Y. (2020) Quantitative site- and structure-specific N-glycoproteomics characterization of differential N-glycosylation in MCF-7/ADR cancer stem cells. *Clin. Proteomics* **17**, 3
79. Choo, M. S., Wan, C., Rudd, P. M., and Nguyen-Khuong, T. (2019) GlycopeptideGraphMS: Improved glycopeptide detection and identification by exploiting graph theoretical patterns in mass and retention time. *Anal. Chem.* **91**, 7236–7244
80. An, Z., Shu, Q., Lv, H., Shu, L., Wang, J., Yang, F., and Fu, Y. (2018) N-linked glycopeptide identification based on open mass spectral library search. *Biomed. Res. Int.* **2018**, 1564136
81. Nasir, W., Toledo, A. G., Noborn, F., Nilsson, J., Wang, M., Bandeira, N., and Larson, G. (2016) SweetNET: A bioinformatics workflow for glycopeptide MS/MS spectral analysis. *J. Proteome Res.* **15**, 2826–2840
82. Mayampurath, A., Yu, C. Y., Song, E., Balan, J., Mechref, Y., and Tang, H. (2014) Computational framework for identification of intact glycopeptides in complex samples. *Anal. Chem.* **86**, 453–463
83. Sun, S., Shah, P., Eshghi, S. T., Yang, W., Trikanad, N., Yang, S., Chen, L., Aiyetan, P., Hoti, N., Zhang, Z., Chan, D. W., and Zhang, H. (2016) Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **34**, 84–88
84. Cao, L., Qu, Y., Zhang, Z., Wang, Z., Prytkova, I., and Wu, S. (2016) Intact glycopeptide characterization using mass spectrometry. *Expert Rev. Proteomics* **13**, 513–522
85. Chen, Z., Huang, J., and Li, L. (2019) Recent advances in mass spectrometry (MS)-based glycoproteomics in complex biological samples. *Trends Analyt. Chem.* **118**, 880–892
86. Qin, H., Chen, Y., Mao, J., Cheng, K., Sun, D., Dong, M., Wang, L., Wang, L., and Ye, M. (2019) Proteomics analysis of site-specific glycoforms by a virtual multistage mass spectrometry method. *Anal. Chim. Acta* **1070**, 60–68
87. Qin, H., Dong, X., Mao, J., Chen, Y., Dong, M., Wang, L., Guo, Z., Liang, X., and Ye, M. (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
88. Gao, W., Li, H., Liu, L., Huang, P., Wang, Z., Chen, W., Ye, M., Yu, X., and Tian, R. (2019) An integrated strategy for high-sensitive and multi-level glycoproteome analysis from low micrograms of protein samples. *J. Chromatogr. A* **1600**, 46–54
89. Li, Q., Xie, Y., Wong, M., and Lebrilla, C. B. (2019) Characterization of cell glycolyx with mass spectrometry methods. *Cells* **8**, 882
90. Park, D. D., Xu, G., Wong, M., Phoomak, C., Liu, M., Haigh, N. E., Wongkham, S., Yang, P., Mavarakis, E., and Lebrilla, C. B. (2018) Membrane glycomics reveal heterogeneity and quantitative distribution of cell surface sialylation. *Chem. Sci.* **9**, 6271–6285
91. Suttapitugsakul, S., Sun, F., and Wu, R. (2020) Recent advances in glycoproteomic analysis by mass spectrometry. *Anal. Chem.* **92**, 267–291
92. Zhang, Y., Mao, Y., Zhao, W., Su, T., Zhong, Y., Fu, L., Zhu, J., Cheng, J., and Yang, H. (2020) Glyco-CPLL: An integrated method for in-depth and comprehensive N-glycoproteome profiling of human plasma. *J. Proteome Res.* **19**, 655–666
93. Zhu, J., Warner, E., Parikh, N. D., and Lubman, D. M. (2019) Glycoproteomic markers of hepatocellular carcinoma-mass spectrometry based approaches. *Mass Spectrom. Rev.* **38**, 265–290
94. Brown, C. J., Grassmyer, K. T., MacDonald, M. L., Clemmer, D. E., and Trinidad, J. C. (2019) Glycoproteome analysis of human serum and brain tissue. *bioRxiv*. <https://doi.org/10.1101/647081>
95. Zhang, S., Cao, X., Liu, C., Li, W., Zeng, W., Li, B., Chi, H., Liu, M., Qin, X., Tang, L., Yan, G., Ge, Z., Liu, Y., Gao, Q., and Lu, H. (2019) N-glycopeptide signatures of IgA2 in serum from patients with hepatitis B Virus-related Liver diseases. *Mol. Cell. Proteomics* **18**, 2262–2272
96. Lin, Y. H., Zhu, J., Meijer, S., Franc, V., and Heck, A. J. R. (2019) Glycoproteogenomics: A frequent gene polymorphism affects the

- glycosylation pattern of the human serum fetuin/alpha-2-HS-glycoprotein. *Mol. Cell. Proteomics* **18**, 1479–1490
97. Zhang, H., Lv, Y., Du, J., Shao, W., Jiao, F., Xia, C., Gao, F., Yu, Q., Liu, Y., Zhang, W., Zhang, Y., Qin, W., and Qian, X. (2020) A GSH functionalized magnetic ultra-thin 2D-MoS₂ nanocomposite for HILIC- based enrichment of N-glycopeptides from urine exosome and serum proteins. *Anal. Chim. Acta* **1098**, 181–189
 98. Zhao, X., Zheng, S., Li, Y., Huang, J., Zhang, W., Xie, Y., Qin, W., and Qian, X. (2020) An integrated mass spectroscopy data processing strategy for fast identification, in-depth, and reproducible quantification of protein O-glycosylation in a large cohort of human urine samples. *Anal. Chem.* **92**, 690–698
 99. Dang, L., Jia, L., Zhi, Y., Li, P., Zhao, T., Zhu, B., Lan, R., Hu, Y., Zhang, H., and Sun, S. (2019) Mapping human N-linked glycoproteins and glycosylation sites using mass spectrometry. *Trends Analyt. Chem.* **114**, 143–150
 100. Reiding, K. R., Franc, V., Huitema, M. G., Brouwer, E., Heeringa, P., and Heck, A. J. R. (2019) Neutrophil myeloperoxidase harbors distinct site-specific peculiarities in its glycosylation. *J. Biol. Chem.* **294**, 20233–20245
 101. Yang, L., Sun, Z., Zhang, L., Cai, Y., Peng, Y., Cao, T., Zhang, Y., and Lu, H. (2019) Chemical labeling for fine mapping of IgG N-glycosylation by ETD-MS. *Chem. Sci.* **10**, 9302–9307
 102. Di, Y., Zhang, L., Zhang, Y., Zhao, H., Yan, G., Yao, J., Zhang, S., and Lu, H. (2019) MdCDPM: A mass defect-based chemical-directed proteomics method for targeted analysis of intact sialylglycopeptides. *Anal. Chem.* **91**, 9986–9992
 103. Desaire, H. (2013) Glycopeptide analysis, recent developments and applications. *Mol. Cell. Proteomics* **12**, 893–901
 104. Mao, J., You, X., Qin, H., Wang, C., Wang, L., and Ye, M. (2019) A new searching strategy for the identification of O-linked glycopeptides. *Anal. Chem.* **91**, 3852–3859
 105. 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
 106. 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**, 902–910