Published in final edited form as:

Expert Rev Proteomics. 2016 May ; 13(5): 513–522. doi:10.1586/14789450.2016.1172965.

Intact glycopeptide characterization using mass spectrometry

Li Cao¹, Yi Qu², Zhaorui Zhang³, Zhe Wang⁴, Iya Prykova⁴, and Si Wu⁴

¹Pharma Research and Development, R&D Platform Technology & Science, GSK, King of Prussia, PA 19406

²ChemEco Division, Evans Analytical Group, Hercules, CA 94547

³Process Research & Development, AbbVie, North Chicago, IL 60064

⁴. Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK73019

Abstract

Glycosylation is one of the most prominent and extensively studied protein post-translational modifications. However, traditional proteomic studies at the peptide level (bottom-up) rarely characterize intact glycopeptides (glycosylated peptides without removing glycans), so no glycoprotein heterogeneity information is retained. Intact glycopeptide characterization, on the other hand, provides opportunities to simultaneously elucidate the glycan structure and the glycosylation site needed to reveal the actual biological function of protein glycosylation. Recently, significant improvements have been made in the characterization of intact glycopeptides, ranging from enrichment and separation, mass spectroscopy (MS) detection, to bioinformatics analysis. In this review, we recapitulated currently available intact glycopeptide characterization methods with respect to their advantages and limitations as well as their potential applications.

Keywords

glycosylation; glycopeptide; post-translational modification; LC-MS/MS; proteomics; bioinformatics

1. Introduction

Glycosylation, an enzymatic process that attaches glycans to proteins, is one of the most prominent protein posttranslational modifications that play various structural and functional roles in numerous cellular activities[1–5]. The biological function of glycoprotein is closely related to its glycosylation pattern, as demonstrated in the glycosylation of the IgG-Fc moiety[6]. Furthermore, the glycosylation pattern can be altered by various diseases, such as cancer. For example, the glycan pattern of prostate cancer biomarker, prostate specific

[†]To whom correspondence should be addressed: Si Wu, Ph.D., Department of Chemistry and Biochemistry, 101 Stephenson Parkway, Room 2210, Norman, Oklahoma 73019-5251 si.wu@ou.edu.

Declaration of Interest

The authors were supported by OU startup funds (Wu) and NIH 5U01AI101990-04 BRI subaward FY15109843 (Wu). Portions of this work were supported by funds from EMSL intramural research projects and EMSL capability development projects. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

antigen (PSA), changed significantly between the normal and prostate cancer cells[7]. Enormous efforts have been devoted to identifying glycoprotein disease biomarkers and to determining the glycosylation difference between normal and disease conditions. Up to date, the FDA has approved more than ten different glycoproteins as biomarkers for different types of cancer[8]. In addition to serving as biomarkers for disease prognosis and diagnosis, glycoproteins are also extremely important drug and vaccine candidates. For example, glycosylated monoclonal antibodies are promising drug candidates for diseases such as autoimmunity, inflammation, cancer, and HIV[9–11], and nanoparticles assembled by glycoproteins are developed as vaccines for infectious diseases[12, 13].

Generally, in protein glycosylation, glycans can be conjugated to at least nine different amino acid side chains[14]. The two most common and extensively studied glycosylated proteins are *N*-linked glycoproteins, in which glycans are covalently attached to the amide group of asparagine (Asn), and *O*-linked glycoproteins, in which glycans are linked to the hydroxyl group of threonine (Thr) or serine (Ser). The *N*-linked glycoprotein has a consensus sequence: $NX(S/T)$ where X is any amino acid except proline (Pro)[15]. Structurally, the *N*-linked glycoprotein has a universal core glycan structure: $\text{GlcNAc}_2\text{Man}_3$ (where $\text{GlcNAc}=\text{N-acetylglucosamine}$, and $\text{Man}=\text{Mannose}$). By contrast, the *O*-linked glycoprotein has neither consensus sequences nor universal core structures, although GalNAc is the most frequently observed anchor of *O*-linked glycan. In addition to above-mentioned three types of monosaccharides, more than a dozen different sugars were observed, including galactose, xylose, fucose, 5-*N*-acetylneuraminic acid, and other kinds of sialic acids.

Similar to other types of proteomics studies, mass spectrometry (MS) based approaches operating at peptide level have been predominantly applied for glycoprotein characterization, including peptide sequence, glycan composition, and attachment site. However, inherent complexities of intact glycopeptides due to structure heterogeneities present great analytical challenges. First, glycopeptides are often low in abundance, the ion signals of glycopeptides can be suppressed by the majority of co-existing unmodified peptides. Therefore, an efficient enrichment method is necessary. Second, the commonly used collisional based MS fragmentation method like collision-induced dissociation (CID) preferentially dissociates glycan while produces little peptide backbone fragments, which makes identification of peptide sequences difficult. In addition, the low m/z glycan oxonium ions, which are signature ions of glycopeptides[16, 17], may not be detected in the CID MS/MS spectra in ion-trap type MS. Therefore, an alternate fragmentation method is crucial for producing peptide backbone fragments. Third, the lack of sophisticated bioinformatics tools for intact glycopeptide characterization severely hampers glycoprotein research. Even for glycan-only analysis, monosaccharides attached to amino acids can form a tree structure instead of the linear peptide chain; therefore, new algorithms are needed to accommodate the challenge.

Due to the presented obstacles, the conventional practices for *N*-linked and *O*-linked glycopeptide characterization analyze peptide and glycan separately through deglycosylation. Glycopeptides are firstly enriched from complex biological samples, followed by glycan removal through either enzymatic treatment (e. g., PNGase F for *N*-

linked glycoprotein[18, 19]) or chemical reactions[20, 21]. The deglycosylated peptides and/or detached glycans are sequentially collected and analyzed through LC-MS to determine peptide sequences, glycosylation sites, and/or glycan compositions. This approach has successfully demonstrated identification of hundreds of glycosylation sites (*N*-glycosylation and/or *O*-glycosylation) in a single analysis [22–24]. Additionally, detailed glycan composition and structure information can be obtained separately [25, 26]; however, this information cannot be directly connected to specific glycosylation sites. An intact glycopeptide strategy (Figure 1) is emerging with encouraging results as an alternative to deglycosylation[17, 27, 28]. This new strategy keeps the glycan attached to its amino acids, which enables to elucidate the glycopeptide sequence, the glycosylation site, as well as the glycan composition simultaneously. In this review paper, we will review current options for each analytical challenge, their pros and cons, as well as applications[17, 27, 29].

2. Glycopeptide enrichment and separation

Enzymatic digestion is usually the first sample preparation step before glycopeptide enrichment and MS characterization. Trypsin is the most commonly used enzyme in glycoprotein digestion because it specifically cleaves after arginine and lysine, and produces more specific and relatively long glycopeptides for better glycosylation site localizations and more confident peptide identifications. However, larger glycopeptides are often more difficult to be ionized and detected in the routine analysis. Additionally, the cleavage site may be covered by the massive oligosaccharides, which results in miss-cleavage and long and even larger size glycopeptides. On the other hand, non-specific enzymes, such as Pronase E or Proteinase K, have been applied to generate smaller size glycopeptides with four to six amino acids[30, 31] for better detection in MS. However, many of these small nontryptic glycopeptides are not unique peptides and additional information is needed to confidently assign the glycosylated sites. In addition, the fragmentation patterns of nontryptic glycopeptides need to be better understood.

Because of the often-low abundance of glycopeptides in the digestion mixture, the enrichment step is the prerequisite of glycopeptide characterization. Many enrichment and separation methods have been developed for general or specific types of glycopeptides. Here, we will review various enrichment and separation strategies including lectin affinity, hydrazide chemistry, affinity enrichment, hydrophilic interaction liquid chromatography (HILIC), porous graphitized carbon (PGC) separation, and ion mobility spectrometer (IMS) separation. General peptide separation approaches such as reverse phase liquid chromatography (RPLC) and gel separation are not included in this review. However, to improve the enrichment and separation efficiency, orthogonal two-dimensional enrichment/separation composed of glycopeptide specific approaches only or with general peptide separation approaches are often applied[17, 27].

2.1 Lectin affinity

Lectins are a group of non-enzymatic proteins derived from plants, animals, or microbes, which recognize and bind carbohydrates[32, 33]. More than 60 lectins are commercially available[34]. The specificities of the commonly used lectins were described in detail in

other review papers[32, 34]. Some lectins, including wheat germ agglutinin (WGA), concanavalin A (Con A), and jacalin (JAC), recognize a wide range of glycan structures, so they can be applied for the enrichment of larger portion of glycoproteome. Others, including *Sambucus nigra* agglutinin (SNA) and *Maackia amurensis* leucoagglutinin (MAL), have narrower specificity, so they can be used for the enrichment of a small subset of the glycoproteome[32, 34]. Due to the binding specificity of lectins, single[35], serial[36], or multi-lectin[37] affinity chromatography can be used for glycoprotein purification. Another approach so called SimpleCell utilize a universal and stable generic engineering strategy to block the elongation of *O*-glycans. Combining lectin enrichment and ETD/HCD based mass spectrometry analysis, Steentoft identified 600 *O*-glycoproteins with about 3000 *O*-glycosylation sites on these engineered human cells[38].

However, none of the available lectins, or even the combination of different lectins, can capture the entire glycoproteome, so some glycans might be lost during the lectin enrichment[34, 39]. The efficiency of lectin enrichment is also contingent on lectin's accessibility to the target glycan within glycoproteins[32]. In addition, since the concentrated saccharide elution media is incompatible with MS, lectin affinity chromatography cannot be coupled with MS directly.

2.2 Other affinity enrichment approaches

Analysis of sialylated glycoproteins is of particular interest, partly due to the association of this type of glycoproteins with some diseases[40–42]. Titanium dioxide (TiO₂) enrichment is a method highly efficient for glycopeptides containing sialic acids as terminal sugars[43]. Under highly acidic conditions, both carboxylic acid and hydroxyl group of sialic acids interact with TiO₂ through multipoint binding, while non-modified peptides or neutral glycopeptides exhibit no affinity to TiO₂[43]. This method has been successfully applied to explore human plasma and saliva sialiomes: nearly 200 glycopeptides and 100 glycosylation sites were identified[43]. A detailed protocol is also available[44]. The drawback of this strategy is that elution of oligo and polysialylated species is difficult [45].

Azido sugar metabolic labeling is another type of affinity enrichment, which is particularly efficient for enriching low abundance glycoproteins in secretome with the presence of serum[46]. Azido sugars are incorporated into glycan structures by using existing biosynthetic pathways of mammalian cells, and the azide moiety can be utilized to enrich glycoprotein or glycopeptides through click-chemistry and affinity purification. The drawbacks of this method are the time-consuming sugar incorporation process, and many types of samples cannot be obtained through culture processes.

2.3 Hydrazone chemistry

The principle of hydrazone chemistry enrichment is that the *cis*-diol groups of glycan residues are oxidized by sodium periodate to form aldehydes, the oxidized glycans are captured using hydrazone beads, and then the attached *N*-linked glycoproteins/glycopeptides are released by enzymatic deglycosylation using PNGase F[47]. The application of hydrazone chemistry is primarily limited in *N*-linked glycoprotein analysis, due to the lack of an efficient approach to cleave *O*-linked glycoproteins and glycopeptides[47, 48].

Furthermore, this enrichment approach only delivers deglycosylated glycoproteins and glycopeptides, but not intact glycoproteins and glycopeptides. Recently, this approach was extended to characterize intact sialylated glycopeptides. The selectively oxidized sialylated glycoproteins were covalently bound to hydrazide beads and were released by acid hydrolysis with the selective removal of the sialic acids at the terminal positions of glycan chains of the enriched glycopeptides[49, 50].

2.4 Hydrophilic interaction liquid chromatography (HILIC)

HILIC has been employed extensively to enrich and separate carbohydrates[51]. In general, the samples are loaded in the non-polar mobile phase, usually acetonitrile, and analytes are retained to the polar stationary phase through hydrogen bonding, ionic interactions, and dipole-dipole interactions, followed by elution by increasing the water content in the mobile phase[51]. Among the numerous types of HILIC stationary phases, amide and zwitterionic (ZIC-HILIC) are particularly useful to enrich and separate glycopeptides[17, 27, 39, 52, 53]. We have recently demonstrated that the glycan enrichment using ZIC-HILIC followed by dialysis to remove non-specifically bound hydrophilic small molecules (e.g., Lys, and Arg) is a simple, fast, and efficient approach for the downstream NMR analysis of *N*- and *O*-linked glycans[54].

HILIC coupled with MS is a versatile technique for glycomics analysis[51]. HILIC performs excellently in enriching glycopeptides from non-glycosylated peptides[17, 27, 39]. Furthermore, HILIC separates *N*-linked and *O*-linked glycopeptides as well as isomeric glycopeptides[51–53]. By using MS compatible elution buffers, HILIC can be on-line coupled to nano-electrospray MS, which allows direct analysis of the enriched glycopeptides[52, 53]. The disadvantage of HILIC enrichment is that the solubility of analytes is lower in organic mobile phase [45].

2.5 Porous graphitized carbon (PGC)

The PGC stationary phase is composed of pure graphitized carbon, which is highly resistant to extreme LC conditions (e.g. pH, temperature) and thus is advantageous over the conventional silica-based phase[55–57]. PGC has long been adopted to enrich and separate glycan and glycopeptides[17, 58–61]. The retention mechanism of polar analytes by PGC, although not well understood, is drastically different from that of conventional nonpolar phases. Two major explanations exist for the polar retention effect on graphite: dipole-dipole interaction induced by charged analytes, or electronic repartition on the graphite surface[62].

An and colleagues[61] pioneered using PGC solid phase extraction (SPE) to enrich non-specifically digested glycopeptides. In addition to carbohydrate enrichment, PGC has unique strength in separating glycan isomers[58–61], since the retention of analytes on graphitized carbon does not only depend on its hydrophobicity, but also on the interaction of polarized groups with the graphite. The interaction strength is determined by the type of interaction and interaction area. Stronger interaction or larger interaction area results in better retention. Because of PGC's consistent performance and efficient separation of both neutral and anionic oligosaccharides, PGC is also widely used for oligosaccharide separation [62].

2.6 Ion mobility spectrometer (IMS) separation

Isomeric monosaccharides, such as GlcNAc and GalNAc, are a challenge for MS-based glycan structure reconstruction because these isomers have the same atomic compositions and hence have identical masses. The employment of additional gas-phase separation dimension using IMS, provides an avenue to overcome the limitation, because ionized molecules with identical mass but different structures can be separated by IMS according to their collisional cross section size[63]. It has been demonstrated that the GlcNAc and GalNAc oxonium ions were partially separated by IMS, and the same glycopeptides with either GlcNAc or GalNAc attachments were detected with different collisional cross section size[64]. Furthermore, glycan isomers with different linkage or branching styles showed different drift time patterns, which provided insightful information to resolve glycan structures[65]. With the development of IMS technique and availability of commercialized instruments, IMS application for carbohydrate isomers separation and glycan structure characterization is increasing significantly [64–71]. However, the software development for IMS data analysis is currently behind the instrument development, which hinders the application and utilization of IMS.

Overall, affinity based enrichment approaches such as lectin or TiO₂ enrichment are highly specific, and can be applied to enrich low concentration glycoproteins or peptides in a complex sample such as serum or secretome. However, many of these affinity interactions are specific to certain carbohydrate structures, making them impractical for global glycopeptide analysis. On the other hand, HILIC or PGC enrichments are more generic because they utilize polar retention of glycopeptides on the column; however, these methods have limited distinguishing power between hydrophilic peptides and glycopeptides. Additional separation dimensions such as IMS can be directly coupled on-line with MS for glycopeptide isomer separations and glycan structure characterization.

3. MS Fragmentation

A thorough elucidation of intact glycopeptides requires detailed information from both the peptide backbone and its associated monosaccharide. In this section, we will discuss multiple tandem MS (MS/MS) strategies that have been developed and applied to characterize intact glycopeptides, including collision induced dissociation (CID), electron transfer dissociation (ETD), and higher energy collisional dissociation (HCD) (Figure 2 as an example of the comparison of different fragmentation approaches on a glycopeptide[27]). Various fragmentation approaches can be combined either sequentially (i.e., MS³) or in parallel (i.e., alternating) to provide more comprehensive information.

3.1 CID

CID is the earliest and most prevalently utilized tandem MS approach to generate diagnostic fragment ions. CID breaks the CO-NH bonds that produce *b* (N-terminal fragment) and *y* (C-terminal fragment) ions of non-modified peptides. However, glycosidic linkages of glycopeptide are preferentially broken to produce carbohydrate Y series ions because they are more labile than the amide bonds (nomenclature for describing the fragmentation of carbohydrates as in Figure 3[72]), which reveal predominantly the glycan moiety structure.

Due to this limitation, CID is seldom applied alone to study intact glycopeptides. Nonetheless, by combining CID produced Y series ions and high mass measurement accuracy on short glycopeptide, Nwosu et al., (2011) identified more than 200 *N*- and *O*-linked glycopeptides corresponding to 18 glycosites from a mixture of glycoprotein standards.

An alternative approach is adding one more fragmentation stage (MS^3) [49, 50, 53, 73]. In this approach, the selected Y ion generated by MS/MS was subjected to a second ion fragmentation cycle to generate *b* and *y* ions from a peptide backbone [49, 50, 73]. Halim et al., (2012) applied this approach in combination with electron capture dissociation (ECD) fragmentation to study urine glycoproteins, and identified more than 100 *N*- and *O*-linked glycopeptides. However, this approach is limited by the requirement of sufficient parent ion intensity to generate useful information from the MS^3 experiment. Moreover, currently there is no available software to automatically interpret and combine information from different fragmentation stages, which further limited the applicability of the MS^3 approach, particularly for complex biological samples.

3.2 Electron transfer dissociation (ETD)

ETD is an alternative MS fragmentation method and is particularly effective for glycopeptide. During ETD fragmentation, a multiply protonated peptide receives an electron from a radical anion and cleaves at the N-C $_{\alpha}$ bond to produce *c* (N-terminal fragment) and *z* (C-terminal fragment) ions [74]. In contrast to CID, labile post-translational modifications, such as glycosylation, are generally retained on the peptide backbone and are largely unaffected by the ETD fragmentation process.

ETD can be used alone for glycopeptide characterization [28, 75, 76]. Trinidad et al., (2013) conducted a study of *N*- and *O*-linked glycoproteins in mouse synaptosome. By applying ETD to lectin enriched glycopeptides, over 2500 unique *N*- and *O*-linked glycopeptides on 453 proteins in murine synaptosome were identified. However, since ETD does not cleave glycosidic linkages effectively, only monoisotopic masses of oligosaccharides were obtained, and no information of the monosaccharide units was provided.

More frequently, ETD is utilized in combination with complementary fragmentation approaches, like CID, to collect fragments from both the peptide backbone and attached glycan [27, 77]. Alley et al., (2009) demonstrated that the CID/ETD alternative fragmentation is capable of obtaining comprehensive information of the entire glycopeptide from model glycoproteins. The major drawback of ETD is its relatively narrow useful *m/z* range [27, 75, 77]. The fragmentation efficiency of ETD is relatively low (~20%) [77], especially for precursors with *m/z* higher than 850 [75]. This disadvantage is critical for trace amount of analytes in biological samples. In addition, ETD fragmentation works most efficiently on glycopeptides with single HexNAc [78] or GalNAc [38, 79].

3.3 Higher energy collisional dissociation (HCD)

HCD, a newly developed dissociation approach which is unique to Orbitrap MS by Thermo Scientific, also produces *b* and *y* type fragment ions. Compared with CID, HCD has several advantages. First, HCD is performed in a dedicated octopole collision cell at the end of the

“C-trap”, which enhances trap efficiency in the low m/z range, while CID fragments are often limited by the “one-third” m/z cutoff rule[80]. Unlike CID, small diagnostic oxonium ions of monosaccharaides are detectable in HCD[16, 17, 27]. Second, it efficiently dissociates the peptide backbone while preserving the attached labile modifications[27, 81, 82].

Many studies have demonstrated that the HCD did provide beneficial information for glycopeptide characterization[17, 27, 29, 82, 83]. Segu and Mechref (2010) [82] first noticed that Y1 ions (peptide + GlcNAc) of *N*-linked glycopeptides were the most abundant ions in the HCD MS/MS spectra. The Y1 ion was selected and subjected to a second HCD fragmentation to collect *b* and *y* ions from the peptide backbone. Cao et al., (2014) [17] further examined different fragmentation pattern of *N*- and *O*-linked glycopeptide according to different normalized collisional energy (NCE) value. Their result suggested that HCD with lower NCE values preferentially fragmented the glycan, while HCD with higher NCE values preferentially fragmented the peptide backbone. By applying HCD with alternating low and high NCE values, comprehensive structure information of both the sugar chain and the amino acid chain were collected.

In many applications, HCD is also combined with CID or ETD for complimentary fragmentation[27, 29, 84]. Scott et al., (2011) [27] applied both CID/HCD and CID/ETD to analyze *N*-linked glycoproteome of bacteria *C. jejuni*. Compared with CID/ETD, CID/HCD has a higher duty cycle, since ions are split between the ion trap and C-trap for CID/HCD, while delay time is needed for the switch between capturing ions for CID and ETD[27]. HCD/ETD is another combination which showed promising results[29, 84]. Yin et al., (2013) [29] applied and compared alternating HCD-ETD and HCD-product-dependent ETD methods to study endothelial cell secretome. The two methods selected different sets of glycopeptides for fragmentation; therefore, the identified glycopeptides had little overlap. Overall, complimentary fragmentation in general provides greater confidence in the identification of glycopeptides than studies relying on a single approach.

HCD and higher stage CID fragmentation were also applied to provide glycan structure information[85]. Halim et al., (2014) demonstrated that oxonium ion spectrum generated using HCD or CID MS3/MS4 were distinctly different between GlcNAc and GalNAc or among glycans with the same monosaccharide composition but different linkages[85].

4. Bioinformatics

Due to the complexity of glycoprotein, many different combinations of glycan and peptide composition can be isobaric. Based only on intact mass, early software like GlycoMod[86] yields multiple compositional possibilities. To determine the correct composition, tandem MS (MS/MS) is often performed in addition to MS. However, fragmentation increases the complexity of the data and makes the analytical task more challenging. Manual analysis of each tandem spectrum is laborious, time-consuming and requires significant expertise. Recently, a number of algorithms have been developed with automated approaches for glycopeptide spectra interpretation, including SimGlycan[87], GlypID 2.0 / GlycoFragWork[16, 88], ByOnic[89], GlycoPep Grader[90], SweetSEr[91], GP

Finder[92], Sweet-Heart[93], GPQuest[94], GlycoFinder[17], GlycopeptideSearch (GPS) [95], and GlycoMasterDB[96]. The attributes of each software are summarized in Table 1.

With large amounts of nonglycosylated peptides co-produced during protease digestion, glycopeptides are often suppressed or not selected for fragmentation. Using nonspecific proteases[92] or performing glycopeptide enrichment prior to testing are two approaches to circumvent this issue[97]. Recently, a software enrichment strategy on accurate mass measurement and mass defects was developed to quickly discriminate between peptide and N-glycopeptide signals in mass spectrometry. The method based on intrinsic mass defects of elements can be applied to diverse glycoproteomic problems without the need for prior knowledge of protein sequence and glycans[97].

5. Expert commentary and Five-year view

The field of intact glycopeptide study is growing substantially, due to advancements in sample preparation methodologies, instrumentation, and specialized software. Recent studies reported confident identification of hundreds to thousands intact glycopeptides from complex biological samples[27, 28, 94, 99]. However, the identified glycopeptides still only comprise a small percentage of the whole glycoproteome. However, there are remaining challenges for intact glycoprotein analysis. *(1) There is the lack of high throughput, universal intact glycopeptide enrichment techniques.* As we discussed earlier, affinity based enrichment approaches are only specific to a subclass of glycan structures, while hydrophilic-interaction based separation approaches cannot really distinguish glycopeptides with hydrophilic nonglycopeptides. To address this issue, a multiplexed affinity purification approach such as combined schemes of affinity purifications can be incorporated to improve the throughput. An attractive direction is to develop a chemical reaction based Azido sugar labeling approach to combine multiple labeling reagents for various glycan structure groups. After labeling, a single step click-chemistry as well as the following affinity purification can be performed for the following MS detection. In addition, despite of the recent advancements and improvements of the intact glycopeptide purification and separation, the intact glycopeptide sample preparation workflow is time consuming, and there is a critical need for developing automation approaches in both sample preparation and data analysis step. *(2) There are no efficient glycan structure elucidation techniques.* Glycan structure annotations using current fragmentation approaches in MS often rely on oxonium ions and neutral mass loss. It is very difficult to distinguish structure isomers, or define monosaccharides with the same mass (i.e., mannose and glucose). Recently, several studies based on the oxonium ion spectrum produced by different collision methods and IMS suggested improved isomeric glycan separation and provided more information about glycan structure elucidation[64, 65]. This review restricted discussion in qualitative characterization of glycoprotein, however, quantitative description is another crucial challenge of characterization. With rapid technology improvement, these issues are expected to be resolved in the near future.

Acknowledgments

We thank Dr Ljiljana Pasa-Tolic at PNNL for the discussions related to potential applications on intact glycopeptide analysis.

References

*-Papers of interest

** -Papers of particular interest

1. Helenius A, Aebi M. Intracellular functions of N-linked glycans. *Science*. 2001; 291:2364–9. [PubMed: 11269317]
2. Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem*. 2004; 73:1019–49. [PubMed: 15189166]
3. Arnold JN, Wormald MR, Sim RB, et al. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol*. 2007; 25:21–50. [PubMed: 17029568]
4. Rudd PM, Elliott T, Cresswell P, et al. Glycosylation and the immune system. *Science*. 2001; 291:2370–6. [PubMed: 11269318]
- **5. Khidekel N, Ficarro SB, Peters EC, et al. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci U S A*. 2004; 101:13132–7. (first high-throughput O-GlcNAc proteome profiling through chemoenzymatic approach). [PubMed: 15340146]
6. Jefferis R. Glycosylation of recombinant antibody therapeutics. *Biotechnol Prog*. 2005; 21:11–6. [PubMed: 15903235]
7. Peracaula R, Tabares G, Royle L, et al. Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology*. 2003; 13:457–70. [PubMed: 12626390]
8. Kim EH, Misek DE. Glycoproteomics-Based Identification of Cancer Biomarkers. *International Journal of Proteomics*. 2011; 2011:10.
9. Chan AC, Carter PJ. Therapeutic antibodies for autoimmunity and inflammation. *Nat Rev Immunol*. 2010; 10:301–16. [PubMed: 20414204]
10. Weiner LM, Surana R, Wang S. Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nat Rev Immunol*. 2010; 10:317–27. [PubMed: 20414205]
11. Wu X, Yang ZY, Li Y, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*. 2010; 329:856–61. [PubMed: 20616233]
12. Akahata W, Yang ZY, Andersen H, et al. A virus-like particle vaccine for epidemic Chikungunya virus protects nonhuman primates against infection. *Nat Med*. 2010; 16:334–8. [PubMed: 20111039]
13. Kanekiyo M, Wei CJ, Yassine HM, et al. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature*. 2013; 499:102–6. [PubMed: 23698367]
14. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002; 12:43R–56R.
15. Pless DD, Lennarz WJ. Enzymatic conversion of proteins to glycoproteins. *Proc Natl Acad Sci U S A*. 1977; 74:134–8. [PubMed: 264667]
16. Mayampurath AM, Wu Y, Segu ZM, et al. Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity. *Rapid Commun Mass Spectrom*. 2011; 25:2007–19. [PubMed: 21698683]
17. Cao L, Tolic N, Qu Y, et al. Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal Biochem*. 2014; 452:96–102. [PubMed: 24440233]
18. Tarentino AL, Plummer TH Jr. Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from *Flavobacterium meningosepticum*. *Methods Enzymol*. 1994; 230:44–57. [PubMed: 8139511]
19. Plummer TH, Elder JH, Alexander S, et al. Demonstration of Peptide-N-Glycosidase-F Activity in Endo-Beta-N-Acetylglucosaminidase F Preparations. *Journal of Biological Chemistry*. 1984; 259:700–4.
20. Kakehi K, Susami A, Taga A, et al. High-performance capillary electrophoresis of O-glycosidically linked sialic acid-containing oligosaccharides in glycoproteins as their alditol derivatives with low-wavelength UV monitoring. *J Chromatogr A*. 1994; 680:209–15. [PubMed: 7952002]

21. Royle L, Mattu TS, Hart E, et al. An analytical and structural database provides a strategy for sequencing O-glycans from microgram quantities of glycoproteins. *Anal Biochem.* 2002; 304:70–90. [PubMed: 11969191]
22. Parker BL, Palmisano G, Edwards AV, et al. Quantitative N-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment. *Mol Cell Proteomics.* 2011; 10:M110 006833. [PubMed: 21441315]
23. Pasing Y, Sickmann A, Lewandrowski U. N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biol Chem.* 2012; 393:249–58. [PubMed: 23029648]
24. Chen W, Smeekens JM, Wu R. A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol Cell Proteomics.* 2014; 13:1563–72. [PubMed: 24692641]
25. Yabu M, Korekane H, Miyamoto Y. Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology.* 2014; 24:542–53. [PubMed: 24663386]
26. Houel S, Hilliard M, Yu YQ, et al. N- and O-glycosylation analysis of etanercept using liquid chromatography and quadrupole time-of-flight mass spectrometry equipped with electron-transfer dissociation functionality. *Anal Chem.* 2014; 86:576–84. [PubMed: 24308717]
- *27. Scott NE, Parker BL, Connolly AM, et al. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol Cell Proteomics.* 2011; 10:M000031–MCP201. (first intact glycopeptide profiling using combined fragmentation including HCD). [PubMed: 20360033]
- **28. Trinidad JC, Schoepfer R, Burlingame AL, et al. N- and O-glycosylation in the murine synaptosome. *Mol Cell Proteomics.* 2013; 12:3474–88. (>2500 unique N- and O-linked glycopeptides). [PubMed: 23816992]
29. Yin X, Bern M, Xing Q, et al. Glycoproteomic analysis of the secretome of human endothelial cells. *Mol Cell Proteomics.* 2013; 12:956–78. [PubMed: 23345538]
30. Yu YQ, Fournier J, Gilar M, et al. Identification of N-linked glycosylation sites using glycoprotein digestion with pronase prior to MALDI tandem time-of-flight mass spectrometry. *Anal Chem.* 2007; 79:1731–8. [PubMed: 17243769]
31. Froehlich JW, Barboza M, Chu C, et al. Nano-LC-MS/MS of glycopeptides produced by nonspecific proteolysis enables rapid and extensive site-specific glycosylation determination. *Anal Chem.* 2011; 83:5541–7. [PubMed: 21661761]
32. Alley WR Jr, Mann BF, Novotny MV. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem Rev.* 2013; 113:2668–732. [PubMed: 23531120]
33. Sharon N. Lectins: Carbohydrate-specific reagents and biological recognition molecules. *Journal of Biological Chemistry.* 2007; 282:2753–64. [PubMed: 17145746]
34. Fanayan S, Hincapie M, Hancock WS. Using lectins to harvest the plasma/serum glycoproteome. *Electrophoresis.* 2012; 33:1746–54. [PubMed: 22740463]
35. Chowdhury S, Ray S, Chatterjee B. Single step purification of polysaccharides using immobilized jackfruit lectin as affinity adsorbent. *Glycoconjugate Journal.* 1988; 5:27–34.
36. Sumi S, Arai K, Kitahara S, et al. Serial lectin affinity chromatography demonstrates altered asparagine-linked sugar-chain structures of prostate-specific antigen in human prostate carcinoma. *Journal of Chromatography B.* 1999; 727:9–14.
37. Zeng Z, Hincapie M, Pitteri SJ, et al. A Proteomics Platform Combining Depletion, Multi-lectin Affinity Chromatography (M-LAC), and Isoelectric Focusing to Study the Breast Cancer Proteome. *Analytical Chemistry.* 2011; 83:4845–54. [PubMed: 21513341]
38. Steentoft C, Bennett EP, Clausen H. Glycoengineering of human cell lines using zinc finger nuclease gene targeting: SimpleCells with homogeneous GalNAc O-glycosylation allow isolation of the O-glycoproteome by one-step lectin affinity chromatography. *Methods Mol Biol.* 2013; 1022:387–402. [PubMed: 23765677]
39. Wohlgenuth J, Karas M, Eichhorn T, et al. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Analytical Biochemistry.* 2009; 395:178–88. [PubMed: 19699707]

40. Scanlin TF, Glick MC. Terminal glycosylation and disease: Influence on cancer and cystic fibrosis. *Glycoconjugate Journal*. 2000; 17:617–26. [PubMed: 11421353]
41. Arnold JN, Saldova R, Galligan MC, et al. Novel glycan biomarkers for the detection of lung cancer. *J Proteome Res*. 2011; 10:1755–64. [PubMed: 21214223]
42. Miyahara K, Nouse K, Saito S, et al. Serum glycan markers for evaluation of disease activity and prediction of clinical course in patients with ulcerative colitis. *PLoS One*. 2013; 8:e74861. [PubMed: 24116015]
43. Larsen MR, Jensen SS, Jakobsen LA, et al. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics*. 2007; 6:1778–87. [PubMed: 17623646]
44. Palmisano G, Lendal SE, Larsen MR. Titanium dioxide enrichment of sialic acid-containing glycopeptides. *Methods Mol Biol*. 2011; 753:309–22. [PubMed: 21604132]
45. Thaysen-Andersen M, Larsen MR, Packer NH, et al. Structural analysis of glycoprotein sialylation - Part I: pre-LC-MS analytical strategies. *RSC Advances*. 2013; 3:22683–705.
46. Roper SM, Zemskova M, Neely BA, et al. Targeted glycoprotein enrichment and identification in stromal cell secretomes using azido sugar metabolic labeling. *Proteomics Clin Appl*. 2013; 7:367–71. [PubMed: 23687070]
47. Zhang H, Li XJ, Martin DB, et al. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*. 2003; 21:660–6. [PubMed: 12754519]
48. Wang L, Aryal UK, Dai Z, et al. Mapping N-linked glycosylation sites in the secretome and whole cells of *Aspergillus niger* using hydrazide chemistry and mass spectrometry. *J Proteome Res*. 2012; 11:143–56. [PubMed: 22136231]
49. Nilsson J, Ruetschi U, Halim A, et al. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods*. 2009; 6:809–11. [PubMed: 19838169]
50. Halim A, Nilsson J, Ruetschi U, et al. Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol Cell Proteomics*. 2012; 11:M111 013649. [PubMed: 22171320]
51. Wuhrer M, de Boer AR, Deelder AM. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom Rev*. 2009; 28:192–206. [PubMed: 18979527]
52. Takegawa Y, Ito H, Keira T, et al. Profiling of N- and O-glycopeptides of erythropoietin by capillary zwitterionic type of hydrophilic interaction chromatography/electrospray ionization mass spectrometry. *J Sep Sci*. 2008; 31:1585–93. [PubMed: 18435510]
53. Zauner G, Koeleman CA, Deelder AM, et al. Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *J Sep Sci*. 2010; 33:903–10. [PubMed: 20222081]
54. Qu Y, Feng J, Deng S, et al. Structural Analysis of N- and O-glycans Using ZIC-HILIC/DIALYSIS Coupled to NMR Detection. *Fungal Genetics and Biology*. 2014 Accepted.
55. Teutenberg T, Tuerk J, Holzhauser M, et al. Evaluation of column bleed by using an ultraviolet and a charged aerosol detector coupled to a high-temperature liquid chromatographic system. *Journal of Chromatography A*. 2006; 1119:197–201. [PubMed: 16378619]
56. Barrett DA, Pawula M, Knaggs RD, et al. Retention behavior of morphine and its metabolites on a porous graphitic carbon column. *Chromatographia*. 1998; 47:667–72.
57. Monser L. Liquid chromatographic determination of four purine bases using porous graphitic carbon column. *Chromatographia*. 2004; 59:455–9.
58. Brokl M, Hernandez-Hernandez O, Soria AC, et al. Evaluation of different operation modes of high performance liquid chromatography for the analysis of complex mixtures of neutral oligosaccharides. *J Chromatogr A*. 2011; 1218:7697–703. [PubMed: 21628061]
59. Hua S, Hu CY, Kim BJ, et al. Glyco-analytical multispecific proteolysis (Glyco-AMP): a simple method for detailed and quantitative Glycoproteomic characterization. *J Proteome Res*. 2013; 12:4414–23. [PubMed: 24016182]

60. Nwosu CC, Seipert RR, Strum JS, et al. Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *J Proteome Res.* 2011; 10:2612–24. [PubMed: 21469647]
61. An HJ, Peavy TR, Hedrick JL, et al. Determination of N-glycosylation sites and site heterogeneity in glycoproteins. *Anal Chem.* 2003; 75:5628–37. [PubMed: 14710847]
62. West C, Elfakir C, Lafosse M. Porous graphitic carbon: A versatile stationary phase for liquid chromatography. *Journal of Chromatography A.* 2010; 1217:3201–16. [PubMed: 19811787]
63. Pagel K, Harvey DJ. Ion mobility-mass spectrometry of complex carbohydrates: collision cross sections of sodiated N-linked glycans. *Anal Chem.* 2013; 85:5138–45. [PubMed: 23621517]
64. Both P, Green AP, Gray CJ, et al. Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat Chem.* 2014; 6:65–74. [PubMed: 24345949]
65. Zhu F, Trinidad JC, Clemmer DE. Glycopeptide Site Heterogeneity and Structural Diversity Determined by Combined Lectin Affinity Chromatography/IMS/CID/MS Techniques. *J Am Soc Mass Spectrom.* 2015; 26:1092–102. [PubMed: 25840811]
66. Clowers BH, Dwivedi P, Steiner WE, et al. Separation of sodiated isobaric disaccharides and trisaccharides using electrospray ionization-atmospheric pressure ion mobility-time of flight mass spectrometry. *Journal of the American Society for Mass Spectrometry.* 2005; 16:660–9. [PubMed: 15862767]
67. Shi T, Fillmore TL, Sun X, et al. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc Natl Acad Sci U S A.* 2012; 109:15395–400. [PubMed: 22949669]
68. Creese AJ, Cooper HJ. Separation and identification of isomeric glycopeptides by high field asymmetric waveform ion mobility spectrometry. *Anal Chem.* 2012; 84:2597–601. [PubMed: 22280549]
69. Li HL, Bendiak B, Siems WF, et al. Carbohydrate Structure Characterization by Tandem Ion Mobility Mass Spectrometry (IMMS)(2). *Analytical Chemistry.* 2013; 85:2760–9. [PubMed: 23330948]
70. Damen CW, Chen W, Chakraborty AB, et al. Electrospray ionization quadrupole ion-mobility time-of-flight mass spectrometry as a tool to distinguish the lot-to-lot heterogeneity in N-glycosylation profile of the therapeutic monoclonal antibody trastuzumab. *J Am Soc Mass Spectrom.* 2009; 20:2021–33. [PubMed: 19744865]
71. Isailovic D, Plasencia MD, Gaye MM, et al. Delineating diseases by IMS-MS profiling of serum N-linked glycans. *J Proteome Res.* 2012; 11:576–85. [PubMed: 22148953]
72. Morelle W, Michalski JC. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc.* 2007; 2:1585–602. [PubMed: 17585300]
73. Wührer M, Catalina MI, Deelder AM, et al. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2007; 849:115–28.
74. Syka JE, Coon JJ, Schroeder MJ, et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A.* 2004; 101:9528–33. [PubMed: 15210983]
75. Darula Z, Medzihradsky KF. Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol Cell Proteomics.* 2009; 8:2515–26. [PubMed: 19674964]
76. Thaysen-Andersen M, Wilkinson BL, Payne RJ, et al. Site-specific characterisation of densely O-glycosylated mucin-type peptides using electron transfer dissociation ESI-MS/MS. *Electrophoresis.* 2011; 32:3536–45. [PubMed: 22180206]
77. Alley WR Jr, Mechref Y, Novotny MV. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom.* 2009; 23:161–70. [PubMed: 19065542]
78. Alfaro JF, Gong CX, Monroe ME, et al. Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc Natl Acad Sci U S A.* 2012; 109:7280–5. [PubMed: 22517741]

79. Steentoft C, Vakhrushev SY, Vester-Christensen MB, et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods*. 2011; 8:977–82. [PubMed: 21983924]
80. Olsen JV, Macek B, Lange O, et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007; 4:709–12. [PubMed: 17721543]
81. Jedrychowski MP, Huttlin EL, Haas W, et al. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics*. 2011; 10:M111 009910. [PubMed: 21917720]
82. Segu ZM, Mechref Y. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom*. 2010; 24:1217–25. [PubMed: 20391591]
83. Yang W, Shah P, Toghi Eshghi S, et al. Glycoform analysis of recombinant and human immunodeficiency virus envelope protein gp120 via higher energy collisional dissociation and spectral-aligning strategy. *Anal Chem*. 2014; 86:6959–67. [PubMed: 24941220]
84. Singh C, Zampronio CG, Creese AJ, et al. Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J Proteome Res*. 2012; 11:4517–25. [PubMed: 22800195]
85. Halim A, Westerlind U, Pett C, et al. Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J Proteome Res*. 2014; 13:6024–32. [PubMed: 25358049]
86. Cooper CA, Gasteiger E, Packer NH. GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*. 2001; 1:340–9. [PubMed: 11680880]
87. Apte A, Meitei NS. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods Mol Biol*. 2010; 600:269–81. [PubMed: 19882135]
88. Mayampurath A, Yu CY, Song E, et al. Computational framework for identification of intact glycopeptides in complex samples. *Anal Chem*. 2014; 86:453–63. [PubMed: 24279413]
- *89. Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics*. 2012; Chapter 13(Unit13):20. (Good commercial software for intact glycopeptide characterization). [PubMed: 23255153]
90. Woodin CL, Hua D, Maxon M, et al. GlycoPep grader: a web-based utility for assigning the composition of N-linked glycopeptides. *Anal Chem*. 2012; 84:4821–9. [PubMed: 22540370]
91. Serang O, Froehlich JW, Muntel J, et al. SweetSEQer, simple de novo filtering and annotation of glycoconjugate mass spectra. *Mol Cell Proteomics*. 2013; 12:1735–40. [PubMed: 23443135]
92. Strum JS, Nwosu CC, Hua S, et al. Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. *Anal Chem*. 2013; 85:5666–75. [PubMed: 23662732]
93. Wu SW, Liang SY, Pu TH, et al. Sweet-Heart - an integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *J Proteomics*. 2013; 84:1–16. [PubMed: 23568021]
94. Toghi Eshghi S, Shah P, Yang W, et al. GPQuest: A Spectral Library Matching Algorithm for Site-Specific Assignment of Tandem Mass Spectra to Intact N-glycopeptides. *Anal Chem*. 2015; 87:5181–8. [PubMed: 25945896]
95. Chandler KB, Pompach P, Goldman R, et al. Exploring site-specific N-glycosylation microheterogeneity of haptoglobin using glycopeptide CID tandem mass spectra and glycan database search. *J Proteome Res*. 2013; 12:3652–66. [PubMed: 23829323]
96. He L, Xin L, Shan B, et al. GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J Proteome Res*. 2014; 13:3881–95. [PubMed: 25113421]
97. Froehlich JW, Dodds ED, Wilhelm M, et al. A classifier based on accurate mass measurements to aid large scale, unbiased glycoproteomics. *Mol Cell Proteomics*. 2013; 12:1017–25. [PubMed: 23438733]
98. Shen Y, Toli N, Hixson KK, et al. De Novo Sequencing of Unique Sequence Tags for Discovery of Post-Translational Modifications of Proteins. *Analytical Chemistry*. 2008; 80:7742–54. [PubMed: 18783246]

99. Parker BL, Thaysen-Andersen M, Solis N, et al. Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J Proteome Res.* 2013; 12:5791–800. [PubMed: 24090084]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Key issues

- A universal approach to enrich both *O*-linked glycopeptides and *N*-linked glycopeptides needs to be established to improve the coverage of glycoproteome.
- Improved mass spectrometry fragmentation approaches should be developed to confidently assign both glycosylation site localization and glycopeptide sequencing.
- Improved bioinformatics tools are needed to confidently and automatically assign intact glycopeptides with confidence.
- Intact glycopeptide isomers are difficult to separate using current available separating techniques.
- Strategies for quantitative glycopeptide analysis are still to be developed.

Figure 1.
Overall workflow

Figure 2. Comparison of CID-, HCD-, and ETD-MS for fragmentation of a *C. jejuni* PEB3 glycopeptide.[27]

Figure 3.

Nomenclature for describing the MS fragmentation of carbohydrates. In this nomenclature, ions retaining the charge on the non-reducing terminus are named A, B and C, and the ions retaining charge on the reducing terminus are X, Y and Z. A and X correspond to cross-ring cleavages, whereas B, C, Y and Z correspond to glycosidic cleavages. Subscript numbers denote the cleavage position, starting at the reducing terminus for the X, Y and Z ions, and at the non-reducing terminus for the others. In the case of ring cleavages, superscript numbers are given to show the cleaved bonds.[72]

Table 1

Comparison of the currently available bioinformatics tools for glycopeptide analysis

Software	Description	Availability	Reference
GlycoMod	Uses intact MS without tandem MS data; searches against known peptide sequences for both <i>N</i> - and <i>O</i> -linked glycopeptides, and reports all results without confident scores	Non- commercial; available through downloading	Cooper, et al., 2001[86]
SimGlycan	Raw/standard MS data files input, MS/MS database searching for glycans, scoring and ranking glycan results	commercial	Apte&Meitei, 2010[87]
GlypID 2.0/ GlycoFragWork	MS and MS/MS input, raw/mzXML standard format. Use CID, ETD, and HCD spectra. Scoring available and false discovery rate (FDR) available in later version. Only applicable to <i>N</i> -linked glycopeptide.	non- commercial; available through downloading	Mayampurath et al., 2011[16]; Mayampurath et al., 2014[88];
ByOnic	Uses tandem mass spectra (MS/MS) spectra, both <i>N</i> - and <i>O</i> - linked glycopeptides, does not require known glycans, supports for all types of fragmentation	Commercial	Bern et al., 2012[89]
GlycoPep Grader	Uses intact MS and tandem MS, designed for <i>N</i> -linked glycopeptide, needs protein sequence, does not require known glycans, MS/MS analysis in charge state dependent fashion, works on low resolution CID data	Non- commercial; available through downloading	Woodin, et al., 2012[90]
SweetSEQer	de novo analysis of tandem MS for glycoconjugates, can be customized and added to other software	Open-source, non-commercial; available through downloading ^a	Serang et al., 2013[91]
GP Finder	Uses intact MS and tandem MS, needs protein sequence, needs known glycans, designed for both <i>N</i> - and <i>O</i> -linked glycopeptide, independent of the type of enzyme. An improved version of GlycoX	non- commercial; available through downloading ^b	Strum et al., 2013[92]
Sweet-Heart	Uses MS ² data to predict glycopeptides and design targeted MS ³ sequence. Information from MS ² and MS ³ experiments integrated to derive peptide backbone sequence and glycoform. Applicable to <i>N</i> -linked glycopeptides only.	Non- commercial	Wu et al., 2013[93]
GPQuest	MS and MS/MS input, HCD spectra. Uses glycan oxonium ions to select intact glycopeptide spectra, then compares the detected spectra to the theoretical glycopeptide spectra library to identify <i>N</i> -linked glycopeptides.	Non- commercial;	Eshghi et al., 2015[94]
GlycoFinder	Specifically designed for tandem MS (MS/MS) spectra with alternating HCD NCE parameters. Uses glycan oxonium ions to select candidate glycopeptide spectra. HCD spectra obtained with low NCE values are analyzed in a similar manner as de novo sequencing for possible monosaccharide composition; HCD spectra obtained with high NCE values are analyzed by an approach similar to USTags[98] and manually validated for peptide sequence.	Non- commercial ^c	Cao et al., 2014[17]
GlycoPeptideSearch (GPS)	Checks for oxonium ions in CID MS/MS spectra to select glycopeptide spectra. Each spectrum is checked for intact- peptide fragment ions and followed by glycan structure matches. Peptide-glycan match in each spectrum is evaluated and glycopeptide FDR is available. Designed for <i>N</i> -linked glycopeptides.	Non- commercial; available through downloading	Chandler et al., 2013[95]
GlycoMaster DB	Uses HCD-only or HCD/ETD MS/MS spectra. Checks for oxonium ions and ion ladders in HCD spectra to select glycopeptide spectra and generates glycan structure based on HCD fragments. Identifies	Non- commercial; available through downloading	He et al., 2014[96]

Software	Description	Availability	Reference
	peptide backbone of glycopeptide in ETD spectra. When ETD data is not available, peptide sequence is identified by matching the calculated peptide mass and considering the glycopeptide sequon. Designed for N-linked glycopeptides. Matching scores for glycan and peptide are calculated.		

Notes:

^a available at the following website: <http://software.steenlab.org/>

^b Link changed to <http://lebrilla.faculty.ucdavis.edu/links/glycopeptide-finder/>

^c available upon request