Review

N-glycoproteomics: mass spectrometry-based glycosylation site annotation

Yvonne Pasing¹, Albert Sickmann¹² and Urs Lewandrowski¹*¹

¹Department of Bioanalytics, Leibniz-Institut für Analytische Wissenschaften, ISAS eV, D-44227 Dortmund, Germany
²Medizinisches Proteom Center, Ruhr-Universität, D-44801 Bochum, Germany
*Corresponding author
e-mail: Urs.Lewandrowski@isas.de

Abstract

Glycosylations are ubiquitous and, in many cases, essential protein modifications. Yet comprehensive and detailed analysis of glycosylations on a proteome-wide scale is a daunting and still unsolved challenge. However, a common workflow has emerged over the last decade for large-scale N-glycosylation site annotation by application of proteomic methodology. Thereby, the qualitative and quantitative assessment of hundreds or thousands of modification sites is enabled. This review presents a short overview about common enrichment techniques and glycosylation site detection for N-glycopeptides, including benefits and challenges of analysis.

Keywords: electrostatic repulsion hydrophilic liquid interaction chromatography; glycopeptide enrichment; glycoproteome; glycosylation; lectin; titanium dioxide.

Introduction

Glycosylations are among the most abundant posttranslational modifications of proteins. Being essentially involved in numerous processes such as cell-cell recognition, subcellular trafficking (Sleat et al., 2006), and even signal transduction (Hart et al., 2007); glycosylations are also related to diseases such as cancer (Adamczyk et al., 2011) or congenital disorders of glycosylation (Schachter, 2001). The term glycosylation comprises a heterogeneous group of carbohydrate containing modification of proteins. The subgroup of N-linked glycans represents one of the most widely distributed forms, with glycans being attached to the amide group of asparagine by their reducing end N-acetylglucosamine residue (Spiro, 2002). In contrast to O-linked glycans, N-glycan attachment requires the presence of a consensus sequence motif NXS/NXT or, to a lesser extent, NXC (X can be any amino acid except proline) (Bause and Legler, 1981; Apweiler et al., 1999). As approximately 50% of all proteins are estimated to be glycosylated (Apweiler et al., 1999) and with thousands of sugar structures having already been deposited in databases (Ranzinger et al., 2008), this leads to potentially millions of glycoprotein isoforms in each cell. Furthermore, microheterogeneities can be frequently observed, as numerous different carbohydrate structures may occupy the same glycosylation site within a (glyco-)protein population (Rush et al., 1995; Harmon et al., 1996). Detailed analysis of this complex riddle is still an unresolved issue. Basically, three types of information have to be assured: the identification of glycan attachment sites within the protein sequence, the structure of respective glycans of a glycoprotein, and specifically which glycan structure is associated to which site (Figure 1A). On a single protein level, these questions can readily be answered by techniques such as mass spectrometry (MS) or nuclear magnetic resonance (NMR). Although NMR is, so far, unsurpassed in providing detailed structural information of glycans such as linkages, branching, or anomery (Agrawal, 1992), the technique requires relatively high amounts of purified proteins. On the contrary, MS analysis achieves absolute sensitivities in the low femtomole range (Wuhrer et al., 2009) even when dealing with complex mixtures of analytes. However, many structural features of glycans and, moreover, glycopeptides cannot be elucidated yet in large-scale applications due to limitations of MS sequencing as well as data interpretation.

So far, most trends in glycoproteomics mainly focus on the elucidation of (primary N-) glycosylation sites derived from complex biological samples in either qualitative or quantitative manner. Although the reduction to N-glycosylations is certainly limiting the amount of potential information being inherent in a complex cell system, it is the first necessary step for a future comprehensive description. In itself, the confirmation of so far unknown glycosylation sites already represents a range of valuable information. First of all, the site information provides a basis for future detailed characterization regarding the in vivo relevance of the respective site, e.g., by mutational analysis (Stroud et al., 2011). Furthermore, glycosylation sites may indicate the orientation of transmembrane proteins (Zielinska et al., 2010), as glycosylations are commonly encountered at the extracellular domains of proteins. In addition, site information supplements already obtained data on protein structure such as X-ray crystallography, where flexible glycans are usually removed before analysis (Grueninger-Leitch et al., 1996). Lastly, by focusing
thematically and technically on glycoprotein or, more precisely, glycosylation site analysis, a range of issues such as the huge dynamic range of protein abundance can be diminished (Stahl-Zeng et al., 2007). This includes the detection of proteins that are extensively glycosylated and not usually accessible to proteomic analysis.

Techniques in MS-based glycosylation site assignment

The central technique for glycosylation site identification is liquid chromatography (LC) coupled online to MS detection. The automated fragmentation of peptide precursor ions results in characteristic fragment ion series that are searched against protein sequence databases (Perkins et al., 1999). In essence, a single identified peptide may be sufficient for the identification of a protein in this manner. Consequently, glycosylation site analysis also requires a single identified (glyco-)peptide for correct site assignment. Regarding the MS fragmentation mode, mostly collision-induced dissociation (CID) is used to generate fragment spectra during proteomic and glycoproteomic workflows. In theory, MS fragmentation of glycopeptides combines all sorts of information – site assignment, glycan structure or composition, and lastly, the allocation of the respective glycan to the glycosylation site. Unfortunately, CID of glycopeptides mainly causes fragmentation of the glycan moiety and only few relevant peptide backbone fragmentations (Hogan et al., 2005; Scott et al., 2010). Hence, insufficient information about the glycopeptide sequence and the attachment site is available, although the fragmentation yields information on the glycan composition. Complementary to CID, extensive peptide chain fragmentation can also be achieved by electron transfer dissociation (ETD; Syka et al., 2004) and electron capture dissociation (ECD; Mirgorodskaya et al., 1999). These techniques use either radical anions (e.g., from fluoranthenes) or direct emission of electrons via an emitter cathode to transfer electrons onto the peptide chain and initiate its fragmentation. The glycan moiety remains intact (Hakansson et al., 2001), and thus glycosylation site identification is enabled (Hogan et al., 2005; Scott et al., 2010). ETD/ECD-based fragmentation suffers from several drawbacks limiting its use for N-glycopeptide analysis. ETD requires at least twofold-charged precursor ions for later fragment ion detection and higher peptide charge states result in a general improvement of fragmentation behavior (Wiesner et al., 2008). Therefore, low-resolution ion traps may suffer from false assignment of multiply charged fragment ions. Furthermore, >2+ peptide charge states during MS ionization frequently require a higher number of basic amino acids within the respective peptide. Therefore, alternate proteases such as Lys-C, Lys-N, or Arg-C have to be used (Taouatas et al., 2008), which nevertheless suffer from reduced specificity and sequence coverage. Even disregarding further drawbacks such as lower fragmentation efficacy in comparison to CID and prolonged fragmentation
times, ETD-based approaches are still not readily compatible with current database search algorithms (Perkins et al., 1999) because, with respect to glycan heterogeneity, virtually thousands of variable sugar modifications would have to be taken into account. The increased number of possible glycan-peptide combinations would dramatically increase the search space and thus reduce the significance of peptide identifications in favor of random hits.

For these reasons, MS-based glycosylation site identification is mainly performed following either complete removal of the glycan moiety (Zhang et al., 2003) or at least trimming of the carbohydrate to a defined minimal structure while still retaining the information of the former attachment site within the peptide sequence in both cases (Hagglund et al., 2004). Most frequently, peptide:N-glycosidase F (PNGaseF) is used for N-glycans, resulting in complete removal of the glycan moiety and a parallel deamidation of asparagine to aspartic acid at the former site of glycosylation (Tarentino et al., 1985). This defined mass increment can be conveniently included in algorithm-based database searches and therefore enables an automated glycosylation site assignment (Kaji et al., 2003; Zhang et al., 2003; Lewandowski et al., 2006).

**N-Glycoproteomics**

Over the last decade, most studies focusing on large-scale N-glycosylation site assignments followed a basic workflow (Figure 1B), which can be divided into four major steps: (1) adequate sample preparation including protein subfractionation and, ultimately, proteolytic digest, which results in complex peptide mixtures still containing only minor amounts of glycopeptides. (2) Therefore, specific glycopeptide enrichment ensures accessibility of low abundant glycopeptides for MS detection by removal of nonglycosylated peptide species. Depending on the enrichment method, even some structural information of the respective glycan moieties can be deduced. (3) Trimming of the glycopeptide removes larger part of the glycan moiety, while retaining the information of the glycosylation site for subsequent (4) LC-MS/MS detection and data analysis. For data interpretation, various protein databases as well as probability-based search and scoring algorithms (Perkins et al., 1999; Lundgren et al., 2005) are available. In the following, a brief overview of the presented workflow along with particular challenges and opportunities will be provided.

**Sample preparation**

The initial step of sample preparation for glycoproteomic analysis has to meet two major prerequisites. First of all, it needs to ensure reproducible and targeted isolation of proteins in the respective sample (cells, tissues, etc.). This includes removal of interfering cellular components such as glycolipids or phospholipids as well as metabolites or DNA, as they might be copurified by glycopeptide-targeted enrichment techniques (Engholm-Keller and Larsen, 2011). Furthermore, a partial removal of high abundant nonglycosylated proteins may be of benefit to reduce sample complexity before glycopeptide enrichment. As many glycoproteins are localized to plasma membranes or other cellular organelles, fractionation by density gradient centrifugation (Khanna et al., 2010) or aqueous two-phase partitioning in combination with carbonate extraction and protein precipitation methods have been used to gain organelle or membrane fractions already enriched in glycoproteins (Lewandowski et al., 2007). Alternatively, immunodepletion-based procedures can be used to remove high abundant proteins such as serum albumin from plasma (Ueda et al., 2010; Zeng et al., 2010). In case initial chemical cell lysis was performed with no further protein fractionation intended, procedures such as filter-assisted sample preparation have been successfully used in glycoproteomic workflows (Zielinska et al., 2010).

However, the common limitations of proteomic workflows have to be considered for all sample preparation techniques to ensure reproducibility and avoid unspecific changes in the sample leading, e.g., to false-positive regulation during relative quantitation of glycosylation sites. Therefore, necessary sample preparation steps should be kept at a minimum.

The second major prerequisite for sample preparation is the cleavage of proteins to smaller peptides. Although some strategies use both – enrichment on the glycoprotein and glycopeptide level (Kaji et al., 2003) – most recent studies directly proceed to the peptide level for enrichment. Thereby, issues concerning protein precipitation effects or unspecific adhesion, e.g., of hydrophobic membrane proteins to stationary supports can be avoided. For cleavage, most studies use tryptic cleaving with high-specificity C-terminal of arginine and lysine residues. Alternatively, Lys-C (Nielsen et al., 2005), Arg-C (Muller et al., 1997), Glu-C, Asp-N, chymotrypsin, or unspecific enzymes such as pronase (An et al., 2003; Temporini et al., 2007) or proteinase K (Zauner et al., 2010) may be used, as complementary digestion methods can increase coverage of glycosylation sites otherwise not accessible to tryptic cleavage (Parker et al., 2011). However, missing enzyme specificity and missed cleavage sites commonly lead to reduced reliability of peptide identifications due to the necessary increase in search space during data evaluation.

**Glycopeptide enrichment**

Online coupled MS detection is a powerful tool enabling the sequencing of thousands of peptides per hour. Yet only a limited number of peptide precursors may be sequenced within a given MS cycle time and typically low abundant species are underrepresented compared with high abundant ones (Wang et al., 2010). Therefore, a glycopeptide-enrichment step before MS analysis has several immediate advantages for glycoproteomic applications. The removal of nonglycopeptides essentially reduces analysis time, thus enabling more site identifications of low abundant glycopeptides. High abundant structural proteins, such as intracellular actin, or housekeeping proteins, such as serum albumin, are not glycosylated and can be efficiently removed (Stahl-Zeng et al., 2007). Even in case high abundant proteins are indeed glycosylated, the number of peptides requiring analysis time and sequencing is drastically reduced. Furthermore, the detection of species featuring few accessible tryptic peptides or extensively glycosylated
extracellular domains is enabled, e.g., for glycoproteins with multiple transmembrane domains such as G-protein-coupled receptors (Lewandrowski et al., 2007). In the following, some of the most common enrichment techniques are presented.

**Lectin affinity chromatography** The probably most renowned method for glycoprotein or glycopeptide enrichment is based on carbohydrate-lectin interactions. Lectins play various roles in cellular recognition via their specific binding to defined carbohydrate epitopes and take part in many biochemical processes such as protein biosynthesis, cell division, and cell agglutination. The advantage of lectins is *per se* their broad spectrum of available specificities and thus the choice for the epitope to screen for. Prominent examples are wheat germ agglutinin, which prefers dimers and trimers of N-acetylgalactosamine and N-acetyleneuraminic acid-bearing glycans (Debray et al., 1981; Kronis and Carver, 1982), concanavalin A from jack bean, which is preferentially used to enrich for N-glycopeptides, as it binds to core α-mannose residues (Baenziger and Fiete, 1979), or the 2–6 linked sialic acid-specific *Sambucus nigra* lectin. Moreover, the combination of two or more lectins with differing, redundant specificity either subsequently during multidimensional lectin affinity chromatography or in parallel by mixed lectin populations in batch mode (Zielinska et al., 2010) can be advantageous to obtain higher coverage of global glycosylation sites. Lectin-based enrichment has been performed with a variety of stationary phases for immobilization of lectins such as agarose, sepharose, or even monolithic supports. Apart from column chromatography (Ueda et al., 2010) or spin columns, these supports are also widely used in batch processing (Kaji et al., 2003). Alternatively, lectins were applied in their free form using filter-assisted sample preparation (Zielinska et al., 2010). Despite the advantage of their specificity, the use of lectins has some major drawbacks, which may influence glycoproteomic data quality. As the interaction of carbohydrates with lectins is transient and lectin binding depends on the respective buffer compositions, harsh washing conditions usually cannot be applied. Therefore, unspecific enrichment, e.g., by interaction of peptides to the stationary support are difficult to diminish. Furthermore, many lectins exhibit secondary specificities (Debray et al., 1981), which may limit their use for enrichment of structurally distinct glycan features. Lastly, recent studies have reported problems regarding specificity and binding efficiency during lectin-based enrichment in vastly complex protein mixtures (Lee et al., 2010). Therefore, the use of lectins especially for quantitative glycoproteomic measurements will have to be critically evaluated in the future. In general, lectins enrich for larger subpopulations of a given glycoproteome, thereby, restricting or influencing global views on the N-glycoproteome.

**Oxidative hydrazide coupling** A broad range of glycopeptides can be enriched by covalent coupling upon oxidation of vicinal cis-diols within the carbohydrate part to aldehyde groups by periodic acid treatment (Zhang et al., 2003; Zeng et al., 2010; Parker et al., 2011). The resulting aldehydes can then be used to form covalent hydrazone bonds with hydrazide functionalized solid supports or bifunctional linkers (Wollscheid et al., 2009). Aldehyde groups are commonly absent in peptides, thereby rendering the technique overall specific for glycopeptides. Moreover, it permits rigorous washing procedures due to covalent coupling, thereby minimizing copurification effects. The coupling reaction can be performed on the glycoprotein (Zeng et al., 2010) or glycopeptide level, whereas elution of glycopeptides is only achievable by bond cleavage, mainly performed enzymatically using endoglycosidases such as PNGaseF (Tarentino et al., 1985). In most cases, however, structural information about the glycan is obviously lost to further analysis. Moreover, this method depends upon a multistage sample preparation process (Lewandrowski and Sickmann, 2009) and requires careful optimization for each step. Lastly, periodate oxidation may also result in modification of N-terminal serine and threonine residues in peptides, which are therefore not accessible to further sequencing (Chelius and Shaler, 2003).

**Hydrophilic liquid interaction chromatography** As glycopeptides feature a rather large and – through their hydroxyl groups – very hydrophilic glycan moiety, hydrophilic liquid interaction chromatography (HILIC) can be used for separation from the bulk of nonmodified tryptic peptides (Hagglund et al., 2004). HILIC is based upon the retention of polar, hydrophilic molecules by the aqueous layer covering the stationary phase under mostly hydrophobic solvent conditions (e.g., 80% v/v acetonitrile; Myslins et al., 2010). For enrichment of glycopeptides and glycans, a wide range of solid supports ranging from pure silica or sepharose beads down to sulfobetain-modified supports (Parker et al., 2011) or simple cotton-based tips (Selman et al., 2011) have been used. Unfortunately, copurification effects can occur with other hydrophilic peptide species due to a general hydrophobicity overlap. The purification efficiency can, however, be enhanced by modifying the concentration of ion pairing reagents such as trifluoroacetic acid, thereby reducing the overlap between glycopeptides and nonmodified species (Mysling et al., 2010). However, secondary effects of the procedure have to be taken into account, as proteins (Kay et al., 2008) and peptides can be precipitated at increased ratios of organic solvent.

**Charged-based enrichment modes** A subgroup of glycopeptides features sialic acid residues as nonreducing end-capping structures. The degree of sialylation has been implicated in disease states such as cancer (Adamczyk et al., 2011) but also in plasma protein turnover and platelet survival (Sorensen et al., 2009). The enrichment of sialylated species can, e.g., be achieved by charge-based separation modes. At low pH, these peptides feature a lower overall net charge due to their partially negatively charged sialic acid residues in comparison to common tryptic peptides. This property is exploited for purification by strong cation exchange chromatography or electrostatic repulsion hydrophilic interaction chromatography (ERLIC). During strong cation exchange chromatography, sialylated
glycopeptides are retrieved in bulk within the flow-through or early eluting fractions due to repulsion effects between negatively charged sialic acid residues and the likewise negatively charged stationary phase (Lewandrowski et al., 2007). However, other charge-reduced peptide species may be present, which bear either additional negative charges, such as phosphopeptides (phosphate group), or a reduced number of positive charges, such as N-terminally acetylated (blocked primary amine) or semitryptic peptides (lysine or arginine missing at C-terminus). These peptides coelute with sialylated species during strong cation exchange chromatography and may hamper further analysis. In contrast, ERLIC enables the retention and separation of sialylated glycopeptides on a weak anion exchange resin overlaid with hydrophilic interaction chromatography at low pH (Lewandrowski et al., 2008; Hao et al., 2011a,b). Meanwhile, most nonmodified peptides are found and discarded within the flow-through. Copurified species such as phosphopeptides can be removed by previous phosphatase treatment. Although initial buffer systems still required intermediate purification of collected fractions due to buffer conditions, this drawback has been overcome by the use of volatile buffer systems (Hao et al., 2011a).

**Metal oxide affinity chromatography** Sialylated glycopeptides can be purified by metal oxide affinity chromatography. Mostly used in combination with titanium dioxide-based solid supports, metal oxide affinity chromatography has been successfully applied for the enrichment of – mostly phosphorylated – biopolymers, as reviewed recently by Engholm-Kellner and Larsen (2011). Regarding the analysis of posttranslational protein modifications, this method is widely used for phosphopeptide enrichment and recently also for enrichment of sialylated glycopeptides (Pal misano et al., 2010; Parker et al., 2011). Under acidic binding conditions, negatively charged sialic acid residues of glycopeptides presumably interact by a multidentate binding with the TiO₂ surface in a manner not yet clearly defined. Copurification of phosphopeptides can be conveniently eliminated by phosphatase treatment, whereas enrichment of otherwise negatively charged peptides can be reduced by competitors such as glycolic acid (Parker et al., 2011).

**Glycopeptide trimming and LC-MS detection** As mentioned previously, a large-scale MS-based sequencing of glycopeptides is not yet feasible due to the heterogeneity of the modification form and the vastly differing fragmentation properties of the carbohydrate and peptide moiety. After enrichment of glycopeptides, most workflows comprise the trimming or complete removal of the glycan moiety before final MS detection to enable reliable peptide sequencing by automated database searches (Figure 2).

Most workflows use PNGaseF for glycopeptide derivatization. It cleaves all N-glycans between the reducing end N-acetylglucosamine and the asparagine residue (Tarentino et al., 1985), except those carrying an α-1-3-linked fucose attached to the innermost N-acetylglucosamine residue, as observed in plants and some insects (Tretter et al., 1991). Furthermore, PNGaseF results in an artificial deamidation of asparagine to aspartic acid at the former site of N-glycosylation and thereby to a mass shift of 0.9840 Da, which can be readily detected by MS. Hence, the subsequent sequencing and mapping of the deamidation site is judged as an adequate indirect assignment of former N-glycosylation sites (Zhang et al., 2003; Beck et al., 2011). However, from the viewpoint of accuracy, this statement needs to be clarified. First of all, deamidations are not only induced by PNGaseF but may also be caused by natural processes within the cell (Robinson and Robinson, 2004) and can subsequently result in false-positive glycosylation site annotation (Parker et al., 2011). This effect can be reduced by two factors. On the one hand, high-efficiency enrichment procedures for glycopeptides will remove a major portion of peptides, which are per chance deamidated within the consensus sequence but not truly glycosylated. For many procedures, enrichment efficiencies of over 90% may be achieved (Woll scheid et al., 2009; Ueda et al., 2010; Beck et al., 2011). On the other hand, use of ¹⁸O-water during PNGaseF digestion will result in a 2.9883-Da mass shift due to the introduction of ¹⁸O to the amide group of asparagine residues (Kaji et al., 2003; Ueda et al., 2010; Beck et al., 2011). The isotopic labeling is chemically identical to the ¹⁶O isotope but is readily differentiable from the 0.9840-Da mass shift caused by natural deamidation. However, although natural deamidation may be ruled out, ¹⁶O-digests do not exclude the possibility for unspecific and spontaneous chemical deamidation during sample processing, e.g., proteolytic or endoglycosidic cleavage reactions. The deamidation rate of asparagine residues is indeed more pronounced at alkaline pH and dependent on buffer and temperature. In addition, chemical deamidation rates are sequence dependent; glycine residues in direct proximity of asparagine have been determined to be prone to spontaneous deamidation (Tyler-Cross and Schirch, 1991). Therefore, spontaneous deamidation cannot be fully excluded at any point of the sample preparation workflow because the accessible pH range is dependent on the used enzymes. The frequency of chemical asparagine deamidation may, however, be reduced by adjusting buffer conditions during proteolytic and endoglycosidic digests. In this context, Hao et al. (2011b) published optimized sample preparation conditions reducing spontaneous deamidations using decreased pH values during tryptic digestion (pH 6) and deglycosylation (pH 5).

As PNGaseF treatment provides only an indirect evidence of the former glycosylation site, endo-β-N-acetylglucosaminidase H can be used to hydrolyze the chitobiase unit in high mannose and some hybrid oligosaccharides of N-linked glycoproteins (Tarentino et al., 1974). The remaining innermost N-acetylgalactosamine residue unambiguously marks the glycosylation site during subsequent analysis (Hagglund et al., 2004), but this defined modification is compatible with large-scale database searches. However, MS detection of this species remains challenging, as fragment spectra are often dominated by neutral loss of the labile N-acetylexosamine residue and return few peptide sequence-related information (Wiesner et al., 2008). Thus, interpretation of derived spectra is still compromised.
Although this drawback can be avoided using ETD or ECD fragmentation, the smaller substrate range of endo-β-N-acetylglucosaminidase H generally favors the initially described PNGaseF treatment.

Deglycosylation may be achieved by MS fragmentation. Using appropriate high voltages within the MS ion source region, in-source fragmentation takes place (Ivancic et al., 2010), largely removing the glycan moiety except for the innermost glycan. A similar effect can be achieved by a further stage of fragmentation (MS<sup>2</sup>). During MS<sup>2</sup>, precursors that form after initial glycan fragmentation of glycopeptides on the MS<sup>2</sup> level can be selected for an additional fragmentation cycle to gain further information of the peptide chain (Halim et al., 2011). However, both modes suffer from partial fragmentation of labile peptide bonds as well as the incomplete fragmentation of the glycan moiety, thereby still limiting the automated form of data evaluation.

Besides the mode of deglycosylation, the false-positive or false-negative assignment of glycosylation sites may also be due to errors in the sequencing process. Algorithms such as Mascot (Perkins et al., 1999), Sequest (Lundgren et al., 2005), or Xtandem are routinely used to match experimental spectra to sequences derived from databases such as Swissprot, NCBInr, or IPI. Nevertheless, the spectrum to sequence match is probability based and includes a certain level of error probability for incorrect assignments. For large-scale MS data sets, the use of a so-called false discovery rate has been extensively used to describe the percentage of estimated false positive identifications (usually below 1% on the peptide level) (Nesvizhskii, 2010). Yet this rate does not ensure correct site assignment. Peptides containing more than single or even consecutive asparagine residues may exhibit, e.g., an incorrect assignment of deamidations as the N-glycosylation site marker. Usually, this incorrect assignment is manually revised by searching for incomplete ion series and ascertaining the presence of ions that are in favor of a reliable sequence and site assignment. However, due to the increasingly fast acquisition cycles of modern mass spectrometers, automated procedures also need to be implemented. As shown recently for phosphopeptides, modification site assignment can be improved by comparison of Mascot scores for first- and second-ranking peptide assignments for the same experimental spectrum (Beck et al., 2011). Near-indistinguishable search engine scores argue for less reliable site assignment, and such results should be generally removed from the data set because the assignment of the distinct site of modification within a peptide, e.g., with two neighboring asparagines cannot be assured. This procedure will hopefully help reduce the number of glycosylation sites that are either not reliably assigned or reported outside of the common consensus sequence motif without further proof, e.g., on frequently chemically deamidated NG sites (Tyler-Cross and Schirch, 1991). Yet recent detailed studies indicate that nonconsensus sequence-related N-glycosylation may occur in mammalian systems (Valliere-Douglass et al., 2009, 2010).

Figure 2 Benefits and drawbacks of MS-based N-glycosylation site workflows with respect to the trimming and detection step. Depending on the use of intact, trimmed, or deglycosylated forms, different levels of information content are found, which are counterbalanced by the need for automated large-scale annotation of N-glycosylation sites. Although analysis of complete glycopeptides may result in the most complete information content (glycan and peptide part), it is currently not feasible for high-throughput applications. Vice versa, deglycosylated peptides may be conveniently sequenced but retain no information on the former glycan moiety.

Although this drawback can be avoided using ETD or ECD fragmentation, the smaller substrate range of endo-β-N-acetylglucosaminidase H generally favors the initially described PNGaseF treatment.
Applications and outlook

In summary, method development has led to a versatile toolbox for glycosylation site analysis including new enrichment modes as well as reliable site detection by MS. Certainly, continuous method development will be required to extend the accuracy and also the information content of current analysis platforms. However, a rapidly increasing number of studies have already successfully used large-scale glycosylation site analysis and expanded its use beyond the immediate identification of sites toward comparative and quantitative assessment of protein and glycosylation changes. The use of methods from proteomic workflows will for sure support and enhance the development of applications in the glycoproteomic field as demonstrated by the following few examples.

As many cell surface proteins are glycosylated, Wollscheid et al. (2009) proposed label-free relative quantitation of surface protein-derived glycopeptides as an alternative to flow cytometry measurements for monitoring, e.g., the differentiation of embryonic stem cells. They took advantage of an oxidative hydrazide labeling procedure with a soluble bifunctional linker (biocytin hydrazide) and analyzed former glycopeptides after PNGaseF treatment. Thereby, a broad spectrum of the modulated membrane proteome from embryonic stem cells over embryonic bodies toward neural progenitor cells could be characterized. MS detection of glycosylation sites thereby successfully substituted for the lack of few available antibodies for flow cytometry.

Further efforts have been made in the field of tumor marker analysis, e.g., derived from serum samples of lung cancer patients. Subsequent to removal of high abundant plasma proteins by immunoaffinity depletion, Zeng et al. (2010) used a common oxidative hydrazide enrichment followed by PNGaseF release. MS analysis of former glycopeptides was performed for adenocarcinoma and squamous cell carcinoma patients. Subsequent to removal of high abundant plasma proteins. By assessing not only the changes in glycosylation motifs toward comparative and quantitative assessment of protein and glycosylation changes. The use of methods from proteomic workflows will for sure support and enhance the development of applications in the glycoproteomic field as demonstrated by the following few examples.

As many cell surface proteins are glycosylated, Wollscheid et al. (2009) proposed label-free relative quantitation of surface protein-derived glycopeptides as an alternative to flow cytometry measurements for monitoring, e.g., the differentiation of embryonic stem cells. They took advantage of an oxidative hydrazide labeling procedure with a soluble bifunctional linker (biocytin hydrazide) and analyzed former glycopeptides after PNGaseF treatment. Thereby, a broad spectrum of the modulated membrane proteome from embryonic stem cells over embryonic bodies toward neural progenitor cells could be characterized. MS detection of glycosylation sites thereby successfully substituted for the lack of few available antibodies for flow cytometry.

Further efforts have been made in the field of tumor marker analysis, e.g., derived from serum samples of lung cancer patients. Subsequent to removal of high abundant plasma proteins by immunoaffinity depletion, Zeng et al. (2010) used a common oxidative hydrazide enrichment followed by PNGaseF release. MS analysis of former glycopeptides was performed for adenocarcinoma and squamous cell carcinoma pools. The authors screened for changes in glycopeptide levels and could discriminate between control and cancer samples on the basis of extracted mass chromatograms. However, it was not yet possible to reliably assess differences between both carcinoma pools. In contrast, Ueda et al. (2010) used a modified lectin-based approach in combination with isobaric tagging for parallel quantitation of glycopeptides from eight sera of patients with obstructive pulmonary disease, lung cancer, and healthy controls in a single analysis. Again, immunoaffinity depletion was used to initially remove abundant plasma proteins. By assessing not only the changes in glycopeptides but in parallel also of nonenriched unmodified peptides, differences originating either from the protein profile or from the glycosylation profile could be separately evaluated. This enables screening for truly glycosylation-dependent biomarkers during cancer diagnosis.

Recently, Parker et al. (2011) used a comprehensive glycoproteomic workflow for the determination of changes in the cell surface glycoproteome after myocardial ischemia and reperfusion in rats. After membrane protein purification and labeling of samples by isobaric tagging, glycopeptides were enriched in parallel by HILIC, titanium dioxide affinity chromatography, and hydrazide capturing. In addition, quantitation of glycopeptides was validated using isotopic dimethyl labeling. A total of 80 glycopeptides with altered abundance were found in the isobaric tagging approach, and more than half of this number could also be confirmed by dimethyl labeling. However, the analysis of released N-glycan profiles suggested no significant alterations in the N-glycan structures. The study mainly retrieved proteins being implicated in cardiac remodeling and suggests an even earlier start of the remodeling process during reperfusion than anticipated.

In total, these studies already show promising aspects of glycoproteomic analysis beyond the mere identification of sites. However, the merge between glycoproteomic and glycomic profiling has to be conducted in the near future. One of the next essential steps is the combination of glycosylation site information and glycan composition at the respective site. Thereby, even subtle changes in the glycosylation profile could be detected, leading to yet another increase of information content in bioanalytical research. This detailed information is expected to assist in understanding, e.g., receptor-ligand binding during cellular contacts such as the rolling of platelets on injured vessel walls, as glycans and glycoproteins are one of the primary interaction interfaces. Refined knowledge of changes in glycosylation may also lead to new and specific biomarkers for cancer subtypes (Adamczyk et al., 2011), thereby enabling reliable screening methods. Lastly, in-depth knowledge of in vivo glycosylation patterns for a large range of glycoproteins may also assist in the design of adequate production cell lines required for the increasing demand of glycosylated biosimilars (Sola and Griebenow, 2009) in the pharmaceutical industry.

Acknowledgments

This work was supported by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen.

References


Received November 2, 2011; accepted January 26, 2012.
Albert Sickmann studied Biochemistry and completed his PhD thesis on the proteome analysis of human cerebrospinal fluid at the Ruhr University Bochum, Germany in 2001. He became assistant professor for Proteomics at the Medical Proteome Center, Bochum, Germany in 2002. Later, he moved to the Rudolf-Virchow-Center for Experimental Biomedicine, Würzburg, Germany where he was appointed professor for protein analytics in 2007. Subsequently, he changed to the Leibniz-Institute for Analytical Sciences, Dortmund, Germany in 2008 where he is now heading the newly established bioanalytic life science branch as scientific director. His main research topics are the analysis of protein networks in mitochondria and the signal transduction of thrombocytes with a major focus on method development for proteomic applications.

Urs Lewandrowski studied Biochemistry at the Ruhr University Bochum and graduated in 2003. For his PhD thesis on glycosylation analysis of human thrombocyte proteins, he moved to the Rudolf-Virchow-Center for Experimental Biomedicine, Würzburg, Germany – finishing in 2007. In 2009 he moved as a research assistant to the Leibniz-Institute for Analytical Sciences where he now heads a working group on glycoproteomic applications. His main interest is the technical development of bioanalytical methods for glycosylation analysis – in particular N-glycosylation site assignment.

Yvonne Pasing studied Chemical Biology at the Faculty of Chemistry, TU Dortmund, Germany and did her Master’s thesis on the biochemical characterization of the alkane monooxygenase enzyme system at the Laboratory of Chemical Biotechnology, Department of Biochemical and Chemical Engineering – graduating in August 2010. Subsequently, she joined the Institute for Analytical Sciences, Dortmund, Germany where she works on the automation of analysis workflows for glycoproteomic applications.

Yvonne Pasing studied Chemical Biology at the Faculty of Chemistry, TU Dortmund, Germany and did her Master’s thesis on the biochemical characterization of the alkane monooxygenase enzyme system at the Laboratory of Chemical Biotechnology, Department of Biochemical and Chemical Engineering – graduating in August 2010. Subsequently, she joined the Institute for Analytical Sciences, Dortmund, Germany where she works on the automation of analysis workflows for glycoproteomic applications.

Unauthenticated
Download Date | 6/24/18 5:18 PM