

## Opinion

## Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis

Yang Yang,<sup>1,2,†</sup> Vojtech Franc,<sup>1,2,†</sup> and Albert J.R. Heck<sup>1,2,\*</sup>

**Glycoproteomics is an important subdiscipline of proteomics. Mass spectrometry (MS)-based glycoproteomics has relied so far on two levels of analysis, targeting either the released glycans or digested glycopeptides. However, limitations of these approaches, caused by the vast structural microheterogeneity that glycoproteins can exhibit, have become evident. Therefore, concomitant methods enabling deep characterization of all discrete glycoforms are essential to extend our functional understanding of the glycoproteome. Here, we discuss recent progress, particularly in protein-centric approaches. High-resolution native MS appears to be especially promising in revealing the glycoform profile of glycoproteins. We propose that systematically integrating MS data acquired at the glycan, glycopeptide, and glycoprotein levels best enhances our understanding of the glycoproteome.**

**Mass Spectrometry for the Analysis of Protein Glycosylation**

To completely understand the proteome, the role of protein post-translational modifications (PTMs), which extend and diversify protein function beyond gene transcription, needs to be elucidated. Most PTMs introduce a well-defined functional group, such as a phosphate or acetate moiety, to specific amino acids on the protein. Distinct from most PTMs, glycosylation involves the attachment of a range of different carbohydrate molecules. There are multiple types of protein glycosylation. The two main types of glycosylation are *N*- and *O*-linked glycosylation, with *C*-mannosylation and *S*-glycosylation occurring less frequently. Approximately half of the proteins in the human genome are glycoproteins [1,2]. Glycoproteins are important because they regulate key biological processes, including cell signaling, embryogenesis, neuronal development, fertilization, hormone activity, immune regulation, and the proliferation of cells and their organization into specific tissues. Over the past decade, substantial progress has been made to obtain detailed information on protein glycosylation, with the aim of assigning functional properties of glycoproteins with respect to their defined structural features. Traditionally, research on protein glycosylation has been perceived from two perspectives: the 'glycomics' field, which studies the glycan structures that have been released from proteins and all other sugar-containing moieties in the cell, and the 'glycoproteomic' field, which largely focuses on the localization and structural elucidation of glycans on proteins. MS is one of the most powerful analytical tools used by both fields for the mass analysis and identification of glycoproteins and for the evaluation of glycosites and determination of oligosaccharide structures [3]. Despite its widespread utilization in glycoprotein analysis, peptide-centric glycoproteomics suffers from several drawbacks, such as difficulties in even detecting glycopeptides among the other highly abundant non-modified peptides, a lack of suitable MS fragmentation techniques, and/or complicated and time-consuming data interpretation. Nevertheless, new ways of overcoming such drawbacks are becoming available following recent technological

## Trends

Glycoproteomics is emerging as an important subdiscipline of proteomics, focusing on the role of protein glycosylation in various biological processes.

MS has taken a central role as an analytical platform in glycoproteomics, focusing mostly on two levels of analysis, targeting either the released glycans or the digested, enriched, glycopeptides. High-throughput profiling of *N*- or *O*-glycans from cell lysates or body fluids has now become achievable, through which, for instance, disease samples can be distinguished from healthy controls.

However, limitations of these high-throughput approaches have become apparent, caused by the structural microheterogeneity exhibited by many glycoproteins. Therefore, complementary methods enabling deep characterization of all discrete glycoprotein proteoforms are indispensable to broaden our understanding of the glycoproteome.

<sup>1</sup>Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

<sup>2</sup>Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

<sup>†</sup>These authors contributed equally.

\*Correspondence: [a.j.r.heck@uu.nl](mailto:a.j.r.heck@uu.nl) (Albert J.R. Heck).

advances in MS instrumentation. Here, we describe these emerging methodologies and their robust implementation towards the goal of understanding the function of protein glycosylation. [Figure 1](#) provides an overview of the current MS-based strategies for glycoprotein analysis that are presented and discussed here.

### Glycomics and Bottom-Up Glycoproteomics for Glycoprotein Analysis

The glycomics field has witnessed extensive development over the past decade. In the most common approaches, protein-linked *N*-glycans and *O*-glycans are first released by enzymatic and/or chemical methods. Matrix-assisted laser desorption/ionization (MALDI), with its ease of analysis of glycans directly from MALDI targets, and electrospray ionization (ESI) coupled online with liquid chromatography (LC), provide versatile, complementary ionization methods that can be applied to glycan analysis [4]. Contemporary MS can rapidly determine the masses of glycans released from a biological source (e.g., cell lysates or body fluids), with low sample consumption and high throughput, facilitating clinical studies, such as cancer glycomic biomarker discovery, and studies of age-related diseases [5,6]. Moreover, due to significant advances in MS-based techniques, it is now possible to obtain detailed structural information, such as the stereochemistry of glycosidic linkages and branching patterns, from complex mixtures of glycoconjugates [7,8]. Unfortunately, sugar molecules are challenging for MS analysis because of their inherent structural complexity, forming myriad linear and branched oligosaccharide structures as well as exhibiting chirality. Diastereomeric glycoconjugates have typically been rather indistinguishable by MS; however, in combination with ion mobility spectrometry, such structural analysis is now possible, at least to some extent [9].

The other approach, still somewhat less developed for investigating the functional role of glycoproteins, is defined as glycoproteomics [10–12]. Currently, MS-based glycoproteomics relies largely on peptide-centric analysis, wherein proteins are first digested by specific proteases into (glyco)peptides carrying individual modified glycosylation sites. After proteolysis, the complex and diverse peptide mixtures are selectively enriched for glycopeptides or directly subjected to LC-ESI-MS/MS analysis. MS detects glycosylation efficiently by scanning for the diagnostic fragment ions created by the facile glycosidic bond cleavage, or by looking for the neutral losses in the fragmentation spectra corresponding to the mass of a certain saccharide moiety. Once a glycopeptide is detected, a variety of tandem MS (MS/MS) techniques can be applied to elucidate the glycan composition, glycosylation site, and peptide sequence. This workflow enables the simultaneous identification of glycoproteins as well as their glycans and can be used to monitor localization, occupancy, and the microheterogeneity present within a glycoprotein sample. It can handle complex peptide–glycopeptide mixtures extracted from serum, cells, or tissues, enabling the discovery of, among others, biologically implicated glyco-epitopes. Current state-of-the-art technology allows site-specific identification of several hundred or even thousands of nonredundant intact *N*-glycopeptides [13–17]. Benefiting from these advantages, peptide-centric glycoproteomic approaches have been widely adopted and have been extensively reviewed elsewhere [18–23].

Many achievements have been made to address particular methodological challenges in glycoproteomics [18–20,24,25]. Known as a long-accepted and understood principle, glycosylation is a nontemplated process, thus, there are no available biological amplification methods, such as the well-known polymerase chain reaction for DNA. Therefore, glycoproteins have to be analyzed at their physiological concentrations, which requires sensitive analytical tools. **Shotgun proteomic** (see [Glossary](#)), the analysis of peptides generated from the digestion of proteins using appropriate proteases, enables the identification and quantification of thousands of proteins within a single experiment [26–29]. As the name implies, shotgun analysis generically targets a few proteotypic peptides from each protein that is abundant in the mixture. Consequently, glycosylated peptides are often neglected in such proteomics

### Glossary

#### Hydrazide chemistry methods:

based on the conjugation of glycoproteins to a solid support using hydrazide chemistry, stable isotope labeling of glycopeptides, and the specific release of *N*-linked glycosylated peptides using the glycosidase PNGase F. The recovered glyco-site-containing peptides are then identified and quantified by LC MS/MS. Since the glycans remain attached to the solid support, this approach only provides information about the glycosylation sites, and not about the nature of the attached glycans.

#### *N*-glycan antennae:

*N*-glycans exhibit a common core pentasaccharide, Man<sub>3</sub>GlcNAc<sub>2</sub>. Further processing in the Golgi results in three main classes of *N*-linked glycan extensions: (i) high mannose; (ii) hybrid; and (iii) complex. While high-mannose glycans contain unsubstituted terminal mannose sugars, hybrid and complex glycans are characterized by substituted mannose residues with an *N*-acetylglucosamine linkage. These GlcNAc sequences added to the *N*-linked glycan core are called ‘antennae’. Based on the number of these branches, the resulting glycan structures are then named biantennary, triantennary, tetraantennary, and so on.

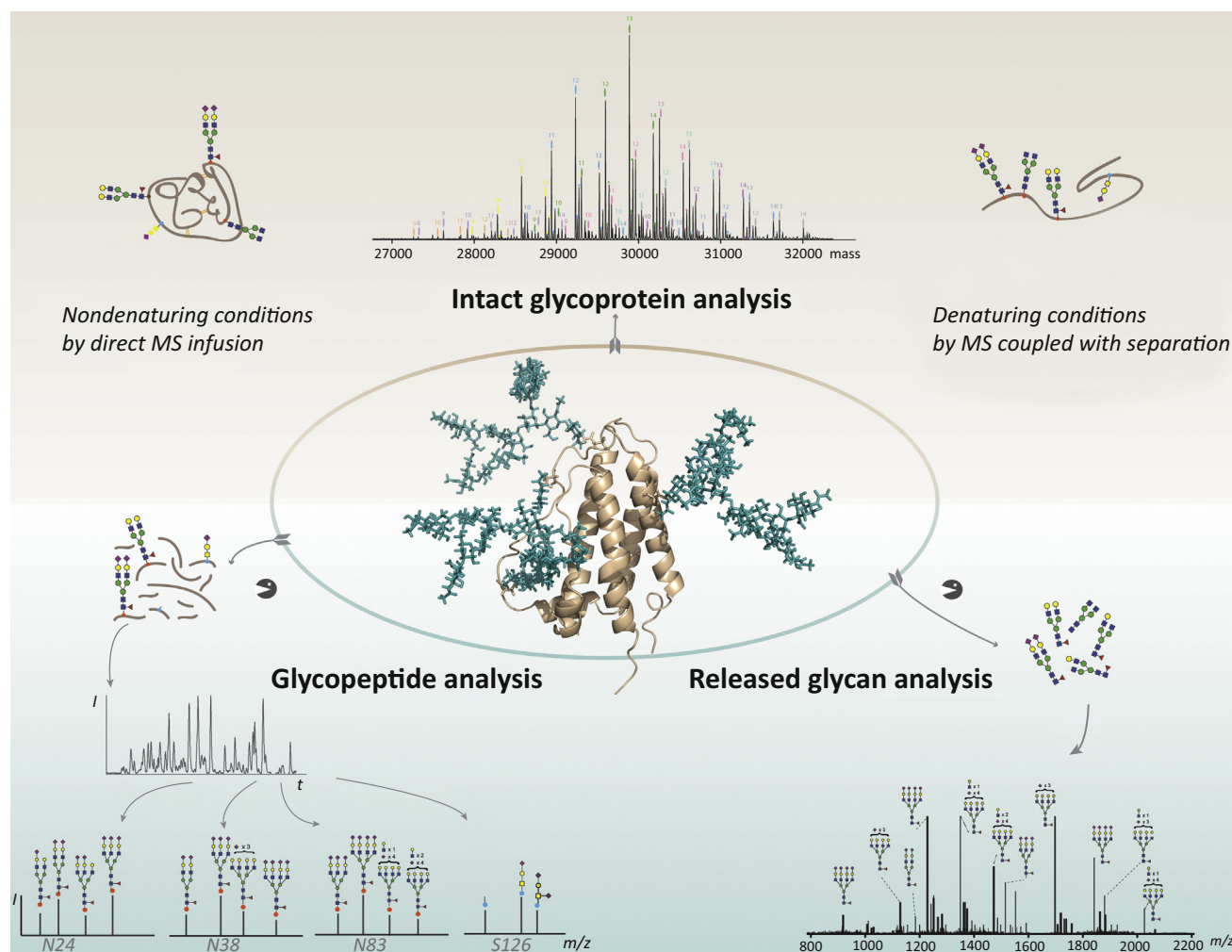
#### Nonconventional tryptic (glyco) peptides:

regular tryptic peptides contain two basic groups that are located at the N terminus and at the C-terminal lysine or arginine. Such peptides become typically double charged upon electrospray ionization. For such peptides, HCD-MS usually provides the best sequencing results. However, during sample preparation before MS analysis, the presence of a glycan on a protein can result in incomplete digestion, for instance due to charge effects or steric hindrance. Such nonconventional tryptic (glyco)peptides often feature a larger size, can take up more charges during the ionization process, and may fragment by nonstandard mechanisms. All these features complicate their analysis by standard HCD-MS techniques.

**Shotgun proteomics:** refers to the combined use of high-performance liquid chromatography and MS to identify thousands of peptides from digested proteins in complex

experiments, since they represent only a minor fraction of the total number of peptides and are inherently more difficult to detect. In addition, most glycoproteins represent a structurally diverse heterogeneous population of proteins [30,31]. Glycoproteins often exhibit a variable number of site occupancies (macroheterogeneity) and/or variable glycan structures at each of their specific glycosylation sites (microheterogeneity). Interestingly, a single glycoprotein can have tens to hundreds of different glycan attachments [30,32]. This enormous diversity of glycan structures segregates MS signals of glycopeptides into a broad spectrum of ion species of various closely related masses, making their detection in shotgun proteomics especially difficult. To overcome these barriers, specifically for glycoproteins, many innovative methods and techniques have been introduced that have substantially improved the analysis of the glycoproteome. These include various enrichment methods using the specific physicochemical

mixtures. The name refers to the somewhat analog method used to sequence DNA, in that the identification of the genes and/or proteins is based on gene and/or peptide fragments, respectively.



Trends in Biotechnology

**Figure 1.** Overview of the Mass Spectrometry (MS)-Based Toolbox for the Analysis of Protein Glycosylation. Shown in the center of the figure is a representative structure of a complex glycoprotein, erythropoietin (EPO), which contains multiple *N*- and *O*-glycosylation sites. In the most widely used approach, the released *N*-glycan pool of EPO is analyzed by matrix-assisted laser desorption/ionization (MALDI)/MS or electrospray ionization (ESI) coupled online with liquid chromatography (LC)-MS/MS (glycan analysis). By digesting EPO into glycopeptides, it is possible to study site-specific glycosylation patterns using a glycopeptide-centric approach. The third-level analyses directly analyze the intact EPO, under either denaturing or nondenaturing conditions. The former is usually conducted using a prefractionation by either capillary electrophoresis (CE) or LC, while the latter is still mostly achieved by directly analyzing the sample by native MS using direct infusion.

properties of proteins and/or peptides, and glycans. The most common glycoprotein and/or glycopeptide enrichment strategies include lectin affinity chromatography [33,34], hydrophilic interaction chromatography (HILIC) [19,35], titanium dioxide chromatography [36,37], and **hydrazide chemistry methods** [38–40].

All of these methods have their pros and cons for enrichment as a consequence of their inherent differences in affinity separation. For example, lectin chromatography can be applied for the functional profiling of a wide spectrum of glycoproteins, using various types of specific lectin to capture different glycan moieties. Titanium dioxide chromatography excels in its ability to selectively enrich sialic acid-containing peptides, due to the specific affinity of titanium dioxide towards negatively charged sialic acid residues. HILIC-based methods selectively enrich glycopeptides by utilizing the hydrophilicity of the glycans. The efficiency of the HILIC material for glycopeptide enrichment has been significantly improved by introducing a customized hydrophilic matrix, named 'click maltose', which enhances the hydrogen-bonding interactions between the glycans of the glycopeptides and the silica-based matrix [41]. A popular variant is zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC), which exhibits an even better hydrophilic interaction towards glycopeptides [42]. In this method, glycopeptides are selectively captured from complex peptide mixtures through hydrophilic and electrostatic interactions on a surface carrying both positive and negative charges. Some of the highest specificity can be obtained by using hydrazide chemistry-based methods. Despite the inherent covalent enrichment principle, which does not allow facile elution of the intact glycopeptides from the beads, recent modifications to these methods have circumvented this obstacle [43,44]. Finally, although these specific approaches can be used for higher throughput analysis, none capture an unbiased view of the whole (glyco)proteome. On the positive side, most of these methods are highly complementary, with their combined use bringing about significant enhancements in glycoproteome depth and coverage [45].

Sample preparation for peptide-centric approaches represents another considerable challenge in the analysis of glycoproteins. For instance, one of the functions of protein glycans is to protect the amino acid backbone of proteins against proteolytic degradation. Therefore, tryptic digestion of glycoproteins often generates longer **nonconventional tryptic (glyco)peptides**, due to the high frequency of miscleavages. Such nonstandard tryptic cleavage products not only complicate LC separation and MS measurements, but may also hamper subsequent quantification and data interpretation. Therefore, the most popular strategy to determine *N*-linked glycosylation sites still entails the addition of a deglycosylation step before their identification by MS [46,47]. Glycosidase PNGase F is typically used for the release of *N*-glycans, leading to a peptide in which the asparagine residue is converted to aspartic acid, due to deamidation. However, the deglycosylation protocol using PNGase F has several drawbacks, such as long incubation times, incomplete deglycosylation, and spontaneous nonenzymatic deamidation of asparagine residues, caused by the applied high temperature and pH, which significantly affect the accuracy of the *N*-linked glycosylation site determination [48]. The implementation of an online deglycosylation protocol has substantially improved this workflow [49]. Nevertheless, new comprehensive strategies providing site-specific glycan occupancies are needed that do not ignore the appended glycans. One promising recent chemoenzymatic method uses solid-phase extraction of *N*-linked glycans and glycosite-containing peptides (NGAG). In principle, this method can simultaneously determine the glycan heterogeneity, the total *N*-glycan pool, and the sequence of the glycosite-containing peptides [50].

In parallel, much effort has been invested into developing better sequencing methods for glycopeptides, including new fragmentation schemes and targeted algorithms for glycopeptides. Labile PTMs, including *O*-linked glycosylations, are lost easily during CID-based analysis [51]. Newer approaches in MS fragmentation techniques can partially overcome these issues

[52–55]. In addition to glycopeptide-tailored instrumentation, several valuable software tools are now available to analyze site-specific glycosylation [56]. A nice example of one of the recent developments for the identification of intact glycopeptides uses complementary MS techniques and MS modes, such as HCD-MS/MS, CID-MS/MS, and MS3. This novel pipeline, called pGlyco, integrates information obtained by this strategy to provide an improved estimation of the false discovery rate (FDR) of the glycopeptides [57]. However, unambiguous peptide backbone identification and glycan (FDR) estimation remain challenges, necessitating manual data inspection.

In summary, over the past decade, glycomics and peptide-centric glycoproteomics methods for the analysis of complex glycoproteins have matured substantially. Although these two approaches are still mostly performed independently, an increasing number of researchers are realizing the importance of integrating the data from these levels of analysis to improve the functional annotation of protein glycosylation.

### A Snapshot from the Top: Intact Glycoprotein Analysis

We described above how glycoproteins and the glycoproteome can be investigated by analyzing released glycans or digested glycopeptides. However, (glyco)proteins can also be directly analyzed by MS, circumventing digestion. This third level had not been explored extensively until recently, partly because of the enormous structural microheterogeneity that a glycoprotein may exhibit, which complicates the mass analysis, and because no adequate (glyco)protein separation techniques had been available. However, protein-centric approaches have intrinsic advantages because they can capture and identify proteoforms of glycoproteins, allowing for the direct assessment of PTMs and combinatorial PTMs, and providing unique information that cannot be extracted easily at the peptide-centric level. For example, molecular-weight profiling of intact glycoproteins gives an overview of the type, abundance, and extent of glycosylation. This approach has proven to be a powerful tool in bacterial glycoproteomics, resulting in the discovery of novel glycans [58,59]. Protein-centric approaches have also gained attention in the field of biopharmaceutical analyses, largely due to the important role of glycosylation in monoclonal antibodies (mAbs). Most protein therapeutics (e.g., antibodies) are glycosylated and glycosylation is known to affect their potency, with aberrant glycosylation having potential risks to patients. Therefore, tools for the detailed characterization of heterogeneous glycoprotein therapeutics are necessary during the design, development, and manufacturing of such therapeutics.

Some separation techniques have been explored to reduce the complexity of heterogeneous glycoproteins. Currently, glycoproteins are often separated using capillary zone electrophoresis (CZE), isoelectric focusing (IEF), or size exclusion chromatography (SEC) [60]. In this way, sample heterogeneity may be reduced, making subsequent MS analysis feasible. Some of these separation methods have been successfully applied in the analysis of proteins with various molecular weights, from human interferon- $\beta$  (~20 kDa) [61] and erythropoietin (~35 kDa) [62] to intact IgG monoclonal antibodies (~150 kDa) [63]. These approaches provide a direct quantitative analysis of co-occurring glycoproteoforms, which represent a valuable indicator of therapeutic product quality control.

Top-down MS [64], whereby intact proteins are introduced into the mass spectrometer for direct sequencing, has been used to some extent for the analysis of site-specific PTMs on proteins, for instance in the analysis of protein acetylation, methylation, and phosphorylation [65,66]. Top-down MS offers the ability to sequence intact proteins and their PTMs and, uniquely, the combination of PTMs at the individual proteoform level. Top-down characterization of glycoproteins is not yet well developed, largely due to the complexity of the analyte [67]. The massive, spacious glycan chains can hamper informative MS/MS fragmentation, reducing



the obtained sequence coverage of the protein polypeptide backbone. One solution is to cut the intact protein into large polypeptides before MS/MS analysis. This so-called ‘middle-down’ approach enables more efficient sequencing that can often differentiate closely related protein isoforms. For instance, the immunoglobulin-G (IgG)-degrading enzyme from *Streptococcus pyogenes* (Ides) is an ideal tool for the analysis of IgG-based antibodies, since it specifically cleaves the IgG at the hinge region and releases two smaller Fab fragments and an Fc [68,69]. Despite their successful application for the analysis of IgG, generic enzymes providing restricted proteolysis with high efficiency and specificity for glycoproteins are not available. As an example of what is now possible, Bush *et al.* reported a study in which CZE was coupled via a sheath-less interface to an Orbitrap mass analyzer for the intact analysis of recombinant human interferon- $\beta$ 1, enabling the authors to separate 138 proteoforms, of which 55 could be quantified. Charged species caused by deamidation and sialylation could be efficiently separated, as well as some triantennary isomers [i.e., **N-glycan antennae** on  $\alpha(1-3)$  or  $\alpha(1-6)$  arms] [61]. This work demonstrates the potential of high-resolution CZE-MS for the top-down analysis of glycoproteins.

### Measuring under Native, Nondenaturing Conditions: Are There Extra Gains?

Most of the protein-centric MS methods described above use LC/MS- or capillary electrophoresis (CE)/MS-compatible solvents, which typically include a combination of water, organic solvent, and sometimes even detergent, at low pH to enhance the protein solubility and ion desolvation during the ESI process. These denaturing conditions inevitably disrupt protein tertiary and quaternary structures, and induce complex dissociation, loss of oligomeric states, and so on. To avoid the loss of specific information during protein denaturation, it is also possible to analyze glycoproteins under nondenaturing conditions, using a technique called native MS [70,71]. Additionally, native MS also reveals information simultaneously about the function of glycoproteins, including the stoichiometry, protein-ligand binding, and protein-protein interactions. Named in analogy to native gel electrophoresis, this field was initially largely focused on the structural and functional analysis of protein-ligand and protein complexes. In 2012, Rosati and colleagues performed high-resolution native MS to qualitatively and semi-quantitatively characterize antibody mixtures, the result of which proved the ability of native MS to tackle heterogeneity in protein mixtures [72]. Later, with the development of the high mass Orbitrap platform, the subtle mass differences in glycoprotein isoforms created by PTMs could be mass separated and even base line resolved directly from the intact proteins under native conditions [73]. This technical breakthrough largely promoted a comprehensive PTM analysis of protein phosphorylation (+80 Da), acetylation (+42 Da), or oxidation (+16 Da) by native MS at the intact protein level [74,75]. An advantage of native MS over MS under denaturing conditions is the resulting inherently less-congested mass spectra, because, under native conditions, the ion signals are distributed over substantially fewer charges and over a wider  $m/z$  window. Following these developments, native MS was also applied to the field of glycoproteomics, analyzing in-depth, for instance, the glycoproteins chicken ovalbumin, several glycoprotein therapeutics (IgGs) and even human plasma-derived glycoproteins [32,76,77]. This relatively new high-resolution MS platform provides a nice addition to the toolbox of protein glycosylation analysis, as described in more detail below. This system facilitates the detailed analysis of low-abundant size variants in recombinant antibodies and could also provide essential validation before admission to market.

(i) Providing an unbiased overview of glycoproteins close to their native form

Several glycans attached to proteins, especially O-linked glycans, are known to be relatively easily dissociated from their glycosylation sites during the sample preparation and/or MS ionization process due to the chemically less stable O-glycosidic bond. Dissociation of the glycosidic bond can occur under acidic conditions, O-glycopeptides can undergo  $\beta$ -elimination upon treatment with strong bases or O-glycans can be lost during the

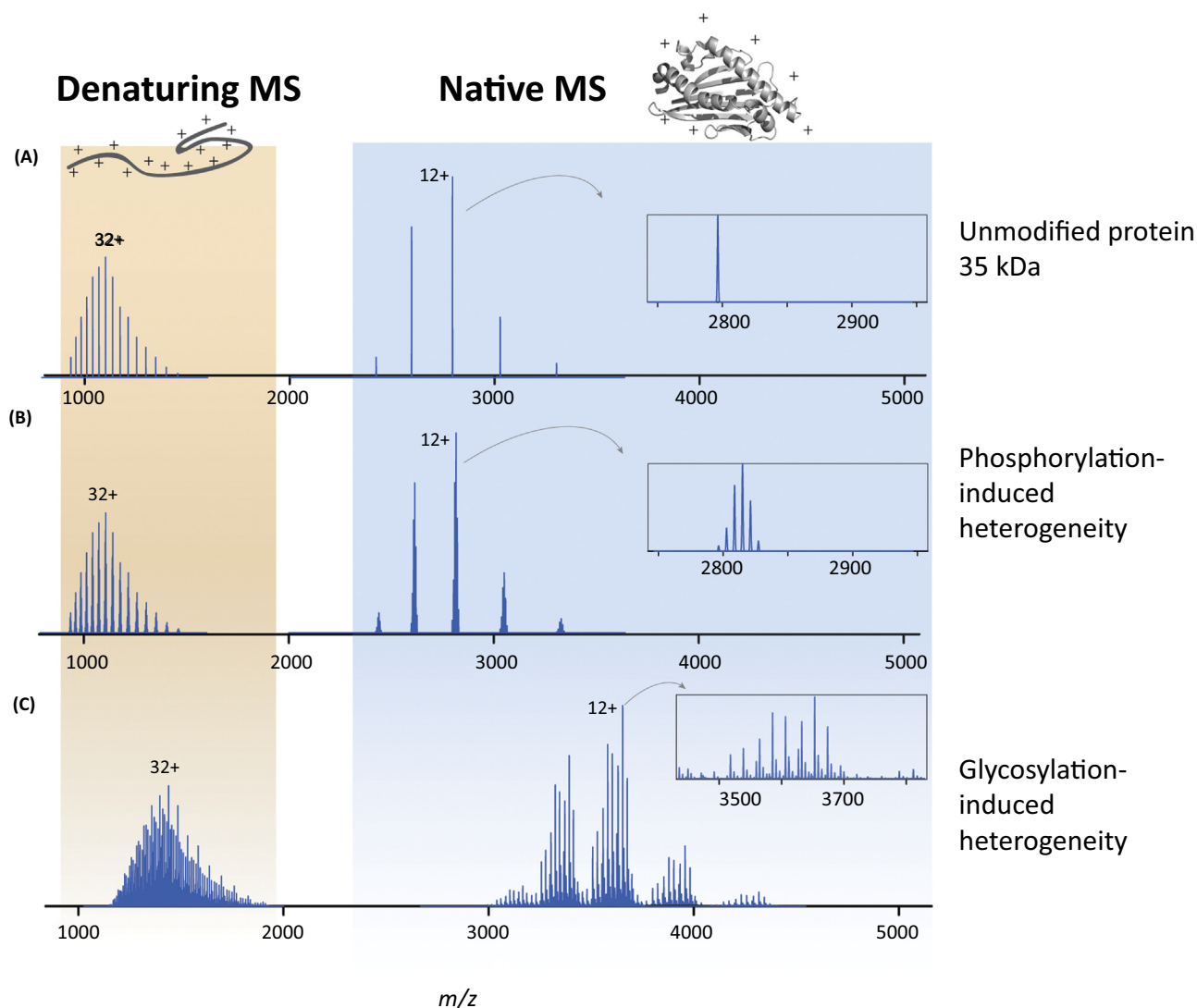
ionization and desolvation process. All these factors can introduce a bias into extracted-ion-chromatogram (XIC)-based quantification. To profile the proteoform distribution more accurately, it is preferable to use gentle conditions that preserve the protein in its most native form. Compared with denaturing MS, native MS does not require denaturation, reduction, alkylation, or digestion of the protein before analysis, simplifying sample handling, reducing sample loss, and enabling the accurate mass determination of all co-occurring proteoforms. Additionally, native MS is comparable to conventional glycosylation analysis techniques, such as 2-amino-benzamide (2-AB) labeling of the glycan pool followed by quantification by HPLC, in a simple, fast and robust fashion [78]. Therefore, native MS can be considered as an essential tool to provide a quantitative overview of the proteoform distribution of intact glycoproteins.

(ii) Decreasing spectral congestion, by providing extra  $m/z$  space to resolve glycoprotein heterogeneity

During native MS analysis, proteins are kept in their original folded, compact structure, which restricts the accommodation of a significant number of protons on their surface upon transition from solution to the gas phase during the ESI process. As a result, proteins take on a narrower charge state distribution and, on average, a lower charge state, compared with standard denaturing MS, resulting in the center of the charge envelop shifting towards a higher  $m/z$  region. This provides a larger space between adjacent charge states, which is valuable for separating protein ions coming from a heterogeneous mass distribution. Figure 2 shows how the increasing heterogeneity of a hypothetical protein with a backbone mass of 35 kDa influences the mass spectra obtained by denaturing versus native MS. The protein containing multiple *N*- and *O*-linked glycosylation sites exhibits a broad molecular-weight distribution caused by the decoration of glycosylation at different sites (Figure 2C). There is a mass difference of more than 3000 Da between the smallest and largest proteoform in this mass spectrum. Under native MS conditions, the ions appear around 3500  $m/z$ , taking on an average of 12 charges. The space between charge  $[M + 13H]^{13+}$  and  $[M + 12H]^{12+}$  is 224  $m/z$ , facilitating the separation of high heterogeneity originating from glycosylation. Empirically, when measuring under denaturing conditions, proteins of this size take on 30–50 charge states during analysis. Assuming the same glycoprotein is measured under denaturing conditions (taking on ~30 charges as an average), all the signal ions appear in a narrower  $m/z$  window, around  $m/z$  1500, creating undesired overlapping of charge states that is impossible to resolve. In real measurements, ESI under denaturing conditions usually results in a broader charge-state distribution, a higher background, and, thus, a lower signal:noise (S/N) ratio, hampering the efficient separation of low-abundance proteoforms.

(iii) Providing opportunities for probing glycosylation regulated interactions

Native MS additionally provides the possibility to study the stoichiometry, protein-ligand binding, and protein-protein interactions of glycoproteins. The removal of glycosylation by different glycoside hydrolases has been applied at the protein level, as demonstrated using sialidase, galactosidase, and PNGase F [79]. This strategy, which stepwise applies glycan-specific glycosidases to sequentially trim the glycan tree, can effectively reduce the heterogeneity. It helps to reconstruct and verify the glycoproteoform assignment by native MS. In addition, by targeting certain linkages of the monosaccharide using a specific exoglycosidase, such as  $\alpha$ -2,3 sialidase or  $\alpha$ -2,6 sialidase, it is possible to achieve linkage-specific sequencing. The methodology is applicable to all types of glycosylation and for monitoring the kinetics of these glycosylation events. This method is particularly applicable in the analysis of proteins generated by glyco-engineering, and lesser-studied glycosylation types, such as *O*-GlcNAcylation. Native MS also has no limit to providing structural information on glycoprotein complexes and in assisting the identification of protein-protein interactions. The application was demonstrated using glycosylated antibody-antigen complexes, multimeric glycoproteins [80], glycosylated antibody-drug conjugates [81,82], and glycoprotein complexes involved in complement activation [83,84].



Trends in Biotechnology

**Figure 2.** Structural Heterogeneity in Proteins as Visualized by Mass Spectrometry. (A) Example mass spectra of a 35-kDa unmodified protein measured under denaturing and native electrospray ionization (ESI) conditions. When a similar size protein is modified at six phosphorylation sites, the resulting proteoforms can still be separated and resolved under both denaturing and native ESI conditions (B). (C) when a similar size protein becomes modified at six glycosylation sites (including *N*- and *O*-glycans), the mass heterogeneity is substantially enhanced, leading to a highly congested mass spectrum under denaturing conditions and, therefore, can best be tackled by native MS.

Despite of its potential, the downside of native MS is that it is still a low-throughput technique that requires specialized expertise, especially in the essential, extensive protein purification process before MS analysis to guarantee a good-quality mass spectrum. Data interpretation is also challenging, especially for multifaceted decorated glycoproteins, due to the escalating combinatorial arrangements of possible positional isomers. The concurrent development of native protein separation techniques will open new opportunities for analyzing highly heterogeneous protein species, particularly given that biotechnological and pharmaceutical companies are currently interested in reliable, high-throughput analyses of their products. Industrial enzymes and therapeutic proteins are key molecular entities with enormous potential in the food industry and human and/or animal healthcare. However, due to their large size, biology-driven production, and complicated structural features, these protein products are one of the



most challenging entities to be functionally and structurally characterized. A recently reported online coupling of SEC with native MS for the characterization and/or validation of bispecific antibody formats represents a nice example of such an analysis under native conditions [85].

Concluding Remarks: Combining the Strengths from all Levels of Analysis: Hybrid MS Approaches

Protein-specific glycosylations regulate biological events and diseases, such as viral infection and several immunological disorders and cancer [5,86]. This highlights an urgent need to establish a generic approach towards the deep characterization of individual glycoproteins, mapping all the sites and all (co-occurring) modifications. As mentioned above, a typical glycoprotein can be considered as a complex mixture of proteoforms, all having closely related masses, resulting from the microheterogeneity of their structures. All of the MS approaches discussed here (Figure 1 and Table 1) have inherent strengths and weaknesses in analyzing qualitatively and quantitatively the protein of interest.

Combined approaches for glycoprotein characterization, using the different levels of MS and chromatography touched upon here, represent an ideal way forward. Multilevel characterization will also facilitate a more comprehensive understanding of the structural and functional properties of these glycoproteins. Yet, to our knowledge, only a small number of such hybrid studies have been reported [87,88]. In our opinion, there are some technical improvements that could boost the more widespread use of hybrid MS approaches in glycoproteomics: (i) methods for measuring the intact glycoproteins need to be optimized, to achieve efficient, fast, and robust MS separation of the heterogeneity in complex glycoproteins; and (ii) additional, improved bioinformatics tools need to become available to compare the data acquired at these different levels. These combined will help us assess the completeness of glycoprotein characterization, resulting in complete in-depth investigations and, in the future, may find answers for some of the most challenging questions in the glycoproteomic field (see Outstanding Questions).

Recently, an integrative workflow was proposed to strengthen the analysis of complex glycoproteins [76]. The workflow combines high-resolution native MS for intact protein analysis and glycopeptide analysis for the detailed characterization of individual glycosites. Subsequently, an *in silico* construction of an intact glycoprotein profile was achieved via a novel integrative algorithm. In this way, data from both approaches could be directly compared using a simple correlation score. Such data integration bridges the gap between the two approaches,

Table 1. Comparison of Glycan, Glycopeptide, and Glycoprotein Centric Analyses

Feature	Native glycoproteins	Intact glycoproteins	Glycopeptides	Released glycans	Hybrid MS
Proteoform profiling; proteoform distribution; relative abundance; stoichiometry	✓✓✓✓	✓✓	✗	✗	✓✓✓✓
Site-specific analysis; localization; occupancy; composition; microheterogeneity; positional isomers	✓	✓	✓✓✓✓	✓	✓✓✓✓
Detailed glycan analysis; structure; glycan isomers	✓	✓	✓✓	✓✓✓✓	✓✓✓✓
Establishment	✓	✓	✓✓✓✓	✓✓✓✓	✓
High throughput	✓	✓	✓✓	✓✓✓✓	✓

<sup>a</sup> The level of development and/or investment in each technology is indicated by the number of check marks '✓'. Techniques that are not suited or have not yet been developed for addressing a certain issue are marked by a '✗'.

and allows the assessment of the integrity of the glycopeptide MS characterization, ultimately facilitating more complete characterization. This integrative workflow led to the discovery of three new C-mannosylation sites in human plasma-derived properdin, for which the stoichiometry could also be directly quantified. This approach is generic and applicable to all types of glycosylation, including those of the less-studied types, such as C-mannosylation, O-GlcNAcylation, or S-glycosylation. Another reported example was of the plasma protein complement protein C9, a 65-kDa glycoprotein that has an essential role in the membrane attack complex. Using the hybrid MS workflow, the structural microheterogeneity of C9 was described in unprecedented detail, including N-glycosylation, C-glycosylation, and O-glycosylation [77]. A similar concept was also recently reported integrating the intact protein analysis and glycan profiling to elucidate IgG glycosylation [89]. Clearly and excitingly, the long-expected integration of different MS platforms is starting to happen. The less-discussed third-level analysis should and will become increasingly valued in the blueprint of the in-depth characterization of complex glycoproteins. Demonstrated by an emerging number of examples, the future of MS-based glycoproteomics is indeed hybrid.

### Acknowledgments

We acknowledge support from The Netherlands Organization for Scientific Research (NWO) funding the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in The Netherlands Proteomics Centre. Y.Y. and A.J.R.H. are supported by the European Union (EU)-funded ITN project ManiFold, grant 317371. This project has received additional funding from the EU's Horizon 2020 Research and Innovation Programme for MSMed, under grant agreement No. 686547. A.J.R.H. acknowledges further support from the NWO TOP-Punt Grant 718.015.003.

### References

- Moremen, K.W. *et al.* (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* 13, 448–462
- Apweiler, R. *et al.* (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* 1473, 4–8
- Varki, A. *et al.* eds (2009) *Essentials of Glycobiology* (2nd edn), Cold Spring Harbor Laboratory Press
- Zaia, J. (2010) Mass spectrometry and glycomics. *Omic J. Integr. Biol.* 14, 401–418
- Ruhaak, L.R. *et al.* (2013) Developments in the identification of glycan biomarkers for the detection of cancer. *Mol. Cell. Proteomics* 12, 846–855
- Alley, W.R. *et al.* (2013) High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem. Rev.* 113, 2668–2732
- Kolarich, D. *et al.* (2012) Determination of site-specific glycan heterogeneity on glycoproteins. *Nat. Protoc.* 7, 1285–1298
- Devakumar, A. *et al.* (2008) Identification of isomeric N-glycan structures by mass spectrometry with 157 nm laser-induced photofragmentation. *J. Am. Soc. Mass Spectrom.* 19, 1027–1040
- Both, P. *et al.* (2014) Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat. Chem.* 6, 65–74
- Pan, S. *et al.* (2011) Mass spectrometry based glycoproteomics—from a proteomics perspective. *Mol. Cell. Proteomics* 10, R110.003251
- Wuhrer, M. *et al.* (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 849, 115–128
- Harvey, D.J. (2005) Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert Rev. Proteomics* 2, 87–101
- Trinidad, J.C. *et al.* (2013) N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* 12, 3474–3488
- Zhao, Y. *et al.* (2014) Online two-dimensional porous graphitic carbon/reversed phase liquid chromatography platform applied to shotgun proteomics and glycoproteomics. *Anal. Chem.* 86, 12172–12179
- Shah, P. *et al.* (2015) Integrated proteomic and glycoproteomic analyses of prostate cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. *Mol. Cell. Proteomics* 14, 2753–2763
- Parker, B.L. *et al.* (2016) Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF- $\alpha$ -induced insulin resistance in adipocytes. *Mol. Cell. Proteomics* 15, 141–153
- Medzhradszky, K.F. (2015) Tissue-specific glycosylation at the glycopeptide level. *Mol. Cell. Proteomics* 14, 2103–2110
- Leverly, S.B. *et al.* (2015) Advances in mass spectrometry driven O-glycoproteomics. *Biochim. Biophys. Acta BBA Gen. Subj.* 1850, 33–42
- Zhu, Z. and Desaire, H. (2015) Carbohydrates on proteins: site-specific glycosylation analysis by mass spectrometry. *Annu. Rev. Anal. Chem. Palo Alto Calif* 8, 463–483
- Gaunitz, S. *et al.* (2016) Recent advances in the analysis of complex glycoproteins. *Anal. Chem.* 89, 389–413
- Huffman, J.E. *et al.* (2014) Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol. Cell. Proteomics* 13, 1598–1610
- Reusch, D. *et al.* (2015) Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles—part 1: separation-based methods. *mAbs* 7, 167–179
- Reusch, D. *et al.* (2015) Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles—Part 2: mass spectrometric methods. *mAbs* 7, 732–742
- Mariño, K. *et al.* (2010) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* 6, 713–723
- Thaysen-Andersen, M. and Packer, N.H. (2014) Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide

### Outstanding Questions

How complex is the structural heterogeneity of individual glycoproteins?

How can all proteoforms of a given glycoprotein be best separated and identified?

Are there biases in the analytical workflows used in MS-based glycoproteomics?

How can the available analytical workflows in glycoproteomics best be integrated?

- site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta* 1844, 1437–1452
26. Cox, J. and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 80, 273–299
  27. Bensimon, A. *et al.* (2012) Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* 81, 379–405
  28. Altelaar, A.F.M. *et al.* (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* 14, 35–48
  29. Richards, A.L. *et al.* (2015) Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* 24, 11–17
  30. Rudd, P.M. and Dwek, R.A. (1997) Glycosylation: heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* 32, 1–100
  31. Parker, B.L. *et al.* (2013) Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J. Proteome Res.* 12, 5791–5800
  32. Yang, Y. *et al.* (2013) Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Anal. Chem.* 85, 12037–12045
  33. Kaji, H. *et al.* (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* 21, 667–672
  34. Yang, Z. and Hancock, W.S. (2004) Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chromatogr. A* 1053, 79–88
  35. Hägglund, P. *et al.* (2004) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* 3, 556–566
  36. Larsen, M.R. *et al.* (2007) Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol. Cell. Proteomics* 6, 1778–1787
  37. Palmisano, G. *et al.* (2010) Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* 5, 1974–1982
  38. Zhang, H. *et al.* (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 21, 660–666
  39. Tian, Y. *et al.* (2007) Solid-phase extraction of N-linked glycopeptides. *Nat. Protoc.* 2, 334–339
  40. Chen, R. *et al.* (2009) Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J. Proteome Res.* 8, 651–661
  41. Yu, L. *et al.* (2009) Hydrophilic interaction chromatography based enrichment of glycopeptides by using click maltose: a matrix with high selectivity and glycosylation heterogeneity coverage. *Chem. Weinh. Bergstr. Ger.* 15, 12618–12626
  42. Mysling, S. *et al.* (2010) Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* 82, 5598–5609
  43. Nilsson, J. *et al.* (2009) Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* 6, 809–811
  44. Huang, J. *et al.* (2015) A peptide N-terminal protection strategy for comprehensive glycoproteome analysis using hydrazide chemistry based method. *Sci. Rep.* 5, 10164
  45. McDonald, C.A. *et al.* (2009) Combining results from lectin affinity chromatography and glyco-capture approaches substantially improves the coverage of the glycoproteome. *Mol. Cell. Proteomics* 8, 287–301
  46. Morelle, W. and Michalski, J.-C. (2007) Analysis of protein glycosylation by mass spectrometry. *Nat. Protoc.* 2, 1585–1602
  47. Zielinska, D.F. *et al.* (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897–907
  48. Rivers, J. *et al.* (2008) Asparagine deamidation and the role of higher order protein structure. *J. Proteome Res.* 7, 921–927
  49. Qu, Y. *et al.* (2011) Integrated sample pretreatment system for N-linked glycosylation site profiling with combination of hydrophilic interaction chromatography and PNGase F immobilized enzymatic reactor via a strong cation exchange precolumn. *Anal. Chem.* 83, 7457–7463
  50. Sun, S. *et al.* (2016) Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* 34, 84–88
  51. Budnik, B.A. *et al.* (2006) Global methods for protein glycosylation analysis by mass spectrometry. *Biochim. Biophys. Acta* 1764, 1870–1880
  52. Alley, W.R. *et al.* (2009) Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun. Mass Spectrom.* 23, 161–170
  53. Saba, J. *et al.* (2012) Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int. J. Proteomics* 2012, e560391
  54. Mechref, Y. (2012) Use of CID/ETD mass spectrometry to analyze glycopeptides. *Curr. Protoc. Protein Sci.* 12, 12.11.1–11
  55. Frese, C.K. *et al.* (2013) Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (ETDcD). *J. Proteome Res.* 12, 1520–1525
  56. Dallas, D.C. *et al.* (2013) Automated glycopeptide analysis—review of current state and future directions. *Brief. Bioinform.* 14, 361–374
  57. Zeng, W.-F. *et al.* (2016) pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Sci. Rep.* 6, 25102
  58. Nothaft, H. and Szymanski, C.M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* 8, 765–778
  59. Valguarnera, E. *et al.* (2016) Sugar and spice make bacteria not nice: protein glycosylation and its influence in pathogenesis. *J. Mol. Biol.* 428, 3206–3220
  60. Staub, A. *et al.* (2011) Intact protein analysis in the biopharmaceutical field. *J. Pharm. Biomed. Anal.* 55, 810–822
  61. Bush, D.R. *et al.* (2016) High resolution CZE-MS quantitative characterization of intact biopharmaceutical proteins: proteoforms of interferon- $\beta$ 1. *Anal. Chem.* 88, 1138–1146
  62. Haselberg, R. *et al.* (2013) Low-flow sheathless capillary electrophoresis-mass spectrometry for sensitive glycoform profiling of intact pharmaceutical proteins. *Anal. Chem.* 85, 2289–2296
  63. Sinha, S. *et al.* (2008) Comparison of LC and LC/MS methods for quantifying N-glycosylation in recombinant IgGs. *J. Am. Soc. Mass Spectrom.* 19, 1643–1654
  64. Kelleher, N.L. (2004) Top-down proteomics. *Anal. Chem.* 76, 197A–203A
  65. Siuti, N. and Kelleher, N.L. (2007) Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 4, 817–821
  66. Toby, T.K. *et al.* (2016) Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem. Palo Alto Calif* 9, 499–519
  67. Nicolardi, S. *et al.* (2015) Top-down MALDI-in-source decay-FTICR mass spectrometry of isotopically resolved proteins. *Anal. Chem.* 87, 3429–3437
  68. An, Y. *et al.* (2014) A new tool for monoclonal antibody analysis: application of IdeS proteolysis in IgG domain-specific characterization. *mAbs* 6, 879–893
  69. Fornelli, L. *et al.* (2014) Middle-down analysis of monoclonal antibodies with electron transfer dissociation Orbitrap Fourier transform mass spectrometry. *Anal. Chem.* 86, 3005–3012
  70. Heck, A.J.R. (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* 5, 927–933
  71. Leney, A.C. and Heck, A.J.R. (2017) Native mass spectrometry: what is in the name? *J. Am. Soc. Mass Spectrom.* 28, 5–13
  72. Rosati, S. *et al.* (2012) Qualitative and semiquantitative analysis of composite mixtures of antibodies by native mass spectrometry. *Anal. Chem.* 84, 7227–7232
  73. Rosati, S. *et al.* (2012) Exploring an Orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chem. Int. Ed.* 51, 12992–12996

74. van de Waterbeemd, M. (2014) Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. *Angew. Chem. Int. Ed. Engl.* 53, 9660–9664
75. Rosati, S. *et al.* (2013) In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *mAbs* 5, 917–924
76. Yang, Y. *et al.* (2016) Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* 7, 13397
77. Franc, V. *et al.* (2017) Proteoform profile mapping of the human serum Complement component C9 reveals unexpected new features of N-, O- and C-glycosylation. *Anal. Chem.* 89, 3483–3491
78. Rose, R.J. *et al.* (2013) Mutation of Y407 in the CH3 domain dramatically alters glycosylation and structure of human IgG. *mAbs* 5, 219–228
79. Rosati, S. *et al.* (2014) Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* 9, 967–976
80. Dyachenko, A. *et al.* (2015) Tandem native mass-spectrometry on antibody–drug conjugates and submillion Da antibody–antigen protein assemblies on an Orbitrap EMR equipped with a high-mass quadrupole mass selector. *Anal. Chem.* 87, 6095–6102
81. Rosati, S. *et al.* (2012) Exploring an Orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chem. Int. Ed. Engl.* 51, 12992–12996
82. Marcoux, J. *et al.* (2015) Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. *Protein Sci. Publ. Protein Soc.* 24, 1210–1223
83. Diebold, C.A. *et al.* (2014) Complement is activated by IgG hexamers assembled at the cell surface. *Science* 343, 1260–1263
84. Wang, G. *et al.* (2016) Molecular basis of assembly and activation of complement component C1 in complex with immunoglobulin G1 and antigen. *Mol. Cell* 63, 135–145
85. Habberger, M. *et al.* (2015) Rapid characterization of biotherapeutic proteins by size-exclusion chromatography coupled to native mass spectrometry. *mAbs* 8, 331–339
86. Theodoratou, E. *et al.* (2016) Glycosylation of plasma IgG in colorectal cancer prognosis. *Sci. Rep.* 6, 28098
87. Ayoub, D. *et al.* (2013) Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *mAbs* 5, 699–710
88. Leymarie, N. *et al.* (2013) Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: The ABRF Glycoprotein Research Multi-Institutional Study 2012. *Mol. Cell. Proteomics* 12, 2935–2951
89. Yang, Y. *et al.* (2017) Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *mAbs* Published online February 14, 2017. <http://dx.doi.org/10.1080/19420862.2017.1290033>