Review Article

The Hitchhiker’s guide to glycoproteomics

Tiago Oliveira1, Morten Thaysen-Andersen2, Nicolle H. Packer1,2,3 and Daniel Kolarich1,3

1Institute for Glycomics, Griffith University, Gold Coast Campus, Gold Coast, Queensland, Australia; 2Department of Molecular Sciences, Macquarie University, Sydney, New South Wales, Australia; 3ARC Centre of Excellence for Nanoscale BioPhotonics, Griffith University, QLD and Macquarie University, NSW, Australia

Correspondence: Daniel Kolarich (d.kolarich@griffith.edu.au)

Protein glycosylation is one of the most common post-translational modifications that are essential for cell function across all domains of life. Changes in glycosylation are considered a hallmark of many diseases, thus making glycoproteins important diagnostic and prognostic biomarker candidates and therapeutic targets. Glycoproteomics, the study of glycans and their carrier proteins in a system-wide context, is becoming a powerful tool in glycobiology that enables the functional analysis of protein glycosylation. This ‘Hitchhiker’s guide to glycoproteomics’ is intended as a starting point for anyone who wants to explore the emerging world of glycoproteomics. The review moves from the techniques that have been developed for the characterisation of single glycoproteins to technologies that may be used for a successful complex glycoproteome characterisation. Examples of the variety of approaches, methodologies, and technologies currently used in the field are given. This review introduces the common strategies to capture glycoprotein-specific and system-wide glycoproteome data from tissues, body fluids, or cells, and a perspective on how integration into a multi-omics workflow enables a deep identification and characterisation of glycoproteins — a class of biomolecules essential in regulating cell function.

Protein glycosylation — the cells’ Swiss Army knife

Protein post-translational modifications (PTMs) enable the cell to produce profound structural and functional diversity from a limited number of protein-encoding genes [1]. Glycosylation plays an essential role across all domains of life [2]. Glycoproteins, together with other glycoconjugates, form the glycocalyx surrounding every living cell [3]. In this highly complex microenvironment, cell-surface receptors, signalling and cell adhesion molecules mediate and regulate cellular communication processes [4]. Intracellularly, O-GlcNAc glycosylation acts within the cytosol in a dynamic interplay with phosphorylation and is biosynthetically independent from the membrane and soluble glycoproteins trafficked to the extracellular environment after their formation [5,6].

In Eukaryotes, glycosylation is crucial for cell functions such as protein folding, regulating signalling or protein activity [2,7,8]. Congenital disorders of glycosylation (CDGs) are often embryonically lethal or phenotypically severe for affected individuals, emphasizing the essential role of glycosylation to life [2,9]. There are also examples of glycosylation ‘defects’ that do not impact normal development (e.g. human ABO blood groups [10]), which, however, can influence the susceptibility to infectious diseases and create crucial population diversity [11]. Changes in cell glycosylation have been associated with systemic pathologies such as (but not limited to) inflammation [12,13], cancer [14–19] or Alzheimer’s disease [20,21]. Disease-associated changes in protein glycosylation are now considered a hallmark in many diseases, making glycans and glycoproteins promising molecular features with enormous diagnostic and prognostic value and potential therapeutic targets for precision medicine [22].

This review aims to provide an ‘easy-to-digest’ introduction to the analytical approaches relevant for studying protein glycosylation. For a comprehensive introduction to the diverse biological functions of protein glycosylation, readers are referred to the freely available Essentials of Glycobiology textbook [23]. Understanding the molecular basis of how glycans are involved in health and disease requires technologies to precisely determine both the glycan structures (glycomics), and characterise their location and structure at discrete sites on glycoproteins (glycoproteomics) expressed by a cell or...
in an entire organ, body fluid, tissue or organism of interest. While the literature harbours many excellent technical reviews covering specific aspects of glycomics (such as [24–34]) and glycoproteomics (examples include [35–44]) technologies and methodologies, there is a gap in the literature surveying the methods and practical issues of modern glycoproteomics relevant to beginners in the field. This mini-review intends to provide a concise introduction to the current strategies available to generate glycoproteomics data and to provide some guidance for designing tailored glycoproteomics experiments.

**Strategies to identify glycoproteins and their glycosylation features from complex samples**

There are five common stages in glycoproteomics analyses (Figure 1). Within each of these stages, a variety of techniques and tools are available that can be combined in different ways. The selection of specific tools will inevitably impact the generated data, as each technique comes with specific advantages and limitations that can impact the success of an experiment.

**Stage 1: glycoprotein extraction**

The first step of any glycoproteomics experiment is to access the glycoproteins of interest from the complex biological matrix. Methods for tissue or cell lysis and protein extraction are diverse, and often depend on the type and available amount of the biological sample. Following tissue homogenisation or sonication and cell-disruption, the extraction step facilitates access to the (glyco)proteins of interest, in particular as many membrane glycoproteins require the presence of detergents to ensure sufficient solubility [45]. Established protocols using ultracentrifugation are available to enrich membrane glycoproteins [46] but sample amount is a limiting factor.

---

**Figure 1. Schematic representation of the five key stages of a glycoproteomics experiment.**

*Stage I:* Extraction of glycoproteins from biological samples. *Stage II:* Proteolysis of glycoproteins, optional glycopeptide enrichment and labelling and offline fractionation to prepare the samples for MS analysis. *Stage III:* Online separation and fragmentation-based identification of glycopeptides. *Stage IV:* Bioinformatic (operator supervised) analyses of the data generated and integration of orthogonal data (e.g. glycomics data) to perform qualitative and quantitative glycoproteome profiling. *Stage V:* Data sharing and accurate reporting of experimental parameters provide a solid basis for integration with other -omics research and reuse in the glycoscience community.

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).
factor. In the case of frozen or formalin-fixed tissues, more intensive physical disruption techniques such as pressure-assisted extraction can provide better yields [38,47]. Protein extraction from body fluids is comparably straightforward, as these specimens already contain soluble glycoproteins. Please note that changing buffer and/or salt concentrations or depletion of highly abundant proteins, often used in proteomics experiments, can result in the unintended loss of glycoproteins at the sample preparation step [48].

If other biomolecules such as DNA, RNA, metabolites, proteoglycans, glycosaminoglycans, lipids, glycolipids or glycans released from proteins are also to be analysed as part of a multi-omics study, the extraction conditions need to be adapted accordingly. In the case of glycolipids, for example, the frequently used chloroform–methanol precipitation method can be used to separate glycolipids from glycoproteins and other lipids [49,50] (Figure 2). Importantly, the composition and pH of the cell lysis buffer will also affect the solubility and integrity of the extracted glycoproteins, as will the presence of certain detergents, salts, denaturing agents and protease inhibitors. Finally, technologies such as filter-aided sample preparation (FASP), facilitate the use of complex MS-incompatible buffers for cell lysis, thus enabling the downstream processing of the extracted glycoproteins in Stage 2 (Figure 1) [51].

**Stage 2: fractionation and sample preparation of glycopeptides**

In Stage 2, the extracted glycoproteins need to be prepared for downstream MS or LC–MS-based characterisation (see Stage 3). While there are some examples where top-down glycoproteomics of semi-purified, intact proteins is successfully applied to diagnose CDGs (reviewed in [52]) and native MS has been used to gain a more holistic view of multi-glycosylated proteins (e.g. myeloperoxidase [53] and neutrophil elastase [54]), bottom-up strategies using proteases are the most widely used approaches for clinical glycoproteomics. The two key steps in Stage 2 are (i) proteolytic digestion of the extracted glycoproteins and (ii) enrichment of glycopeptides/glycoproteins from the complex biological mixtures, though not necessarily in this order.

Intact glycoproteins or glycoprotein complexes can be fractionated by targeted affinity strategies such as immunoprecipitation, 2D gel electrophoresis, SDS–PAGE or lectin chromatography before proteolysis. Alternatively, glycopeptide enrichment is performed after the proteolytic digestion, or the glycopeptides may be pre-fractionated using conventional separation techniques to increase the analytical coverage of the sample of interest [42,55].

In systems-wide glycoproteomics, glycopeptide enrichment remains an essential step since glycopeptides exhibit reduced ionisation efficiency in mass spectrometry compared with unglycosylated peptides, which is further aggravated by the intrinsic macro- and micro-heterogeneity of their glycan moieties [38,56]. Together, in un-enriched samples, these factors result in lower glycopeptide signal intensities relative to their non-glycosylated counterparts and a risk that these molecules are not being selected for fragmentation analyses.

**Proteolytic digestion of complex glycoprotein mixtures**

Proteases are the ‘scissors’ used in bottom-up glycoproteomics experiments, producing a mixture of glycopeptides and peptides [57,58]. Trypsin is the most widely used protease due to its high specificity, availability, and efficiency over a range of conditions (e.g. pH, salts, detergents) [59,60]. As an additional benefit, the resulting C-terminal arginine/lysine residues carry a positive charge, enhancing the ionization and fragmentation of (glyco)peptides [61]. Other proteases such as chymotrypsin, endoproteinase Glu-C, Asp-N and Lys-C, are equally useful for comprehensive glycoproteomics, due to their complementary cleavage specificities [62]. Using dual-protease approaches often increases glycoprotein identification and sequence coverage for the in-depth characterisation of the glycoproteome [63–65].

However, these proteases often inefficiently digest mucin or mucin-like glycoproteins [41,66,67]. Their dense glycosylation makes their already few conventional protease cleavage sites in mucin-domains less accessible, posing unique challenges for successful MS analysis [68]. Excitingly, a suite of novel proteases, the so-called mucinases, have recently become available that facilitate the glycoproteomics analysis of mucins [44]. For example, the OpeRATOR® O-protease requires the presence of an O-glycan on a serine/threonine residue to cleave N-terminally before this site of glycosylation. Thus, this cleavage preference generates peptides that feature N-terminal O-glycosylation, and its application has resulted in the successful mapping of approximately 3000 O-glycosites [69,70]. Nevertheless, OpeRATOR® does not always follow this cleavage pattern and it is strongly advised to perform additional glycopeptide sequencing to confirm glycosylation site localisation. The
Figure 2. Example of an approach to integrate a representative glycoproteomics workflow into a multi-omics study.

Stage I: After tissue lysis, material for genomic, transcriptomics or metabolomic analyses can be retrieved before the separation of lipids and glycolipids from glycoproteins for example by chloroform:methanol:water extraction.

Stage II: 
- Protease digestion
- Peptide labelling
- Sample mixing

Stage III: RP-nanoLC-ESI-MS/MS

Stage IV: Data analyses
- Manual validation

Stage V: Experimental reporting

Glyco-integrative multiomics analyses:
- Data helps to generate: Glycomics “Glycan search space” Proteomics “Protein search space”

Part 1 of 2
secreted protease of C1 esterase inhibitor (StcE), a zinc metalloprotease [71], has also recently been used to
generate O-glycopeptides from mucins, which significantly improved O-glycosite mapping [68].

Generally, glycoproteomics workflows rely on proteases with defined specificity such as the ones mentioned
above. However, multiprotease mixtures such as Pronase, or proteases with very broad substrate specificity such as
Proteinase K, can be of value when characterising single glycoproteins or simple mixtures [72,73]. These pro-
teases exhibit a broad substrate specificity to produce glycopeptides with variable lengths (down to a single amino
acid residue), which can be useful for the in-depth analysis of purified glycoproteins and to cover otherwise
difficult-to-access regions within a protein. Such broadly specific proteases generate extremely heterogeneous
glycopeptide mixtures, making them unsuitable for the analysis of complex samples.

Enrichment using glyco-epitope binding agents

Enrichment strategies that take advantage of the presence of the glycan moiety are fundamental to improve gly-
coproteome coverage [42,74]. Antibodies, lectins or comparable binding agents (e.g. aptamers) have found
widespread application to enrich or fractionate complex mixtures of glycoproteins or glycopeptides [45,75–77].
Plant-derived lectins are the most widely used agents for this purpose, but these often show reactivity to several
different glyco-epitopes, particularly in the presence of a large dynamic range of glyco-structures [78]. As a
consequence of the broad binding affinity patterns of most lectins, any conclusions about the nature of the
enriched glyco-structures need to be carefully considered and ideally backed with additional experiments
(e.g. glycomics) that provide a higher level of compositional and structural information [79].

Lectin affinity chromatography (LAC) has been used to successfully enrich protease-produced glycopeptides
with short O-GalNAc structures such as the ones derived from glycoengineered SimpleCell lines [80,81] or on
cytosolic O-GlcNAc glycoproteins [6,82]. Chemical strategies based on releasing the glycans with simultaneous
labelling of O-GlcNAc glycosylation sites, followed by thiol-Sepharose affinity-enrichment of these modified
peptides have also been successfully employed for O-GlcNAc glycoproteomics [83]. Multi-LAC, the combina-
tion of two or more lectins within one column, has also been successfully employed to enrich glycopeptides
for glycoproteomics experiments [42].

Probes based on bacterial and human lectins or specific anti-glycan antibodies generally appear to exhibit
affinity to more specific glyco-epitopes than plant-derived lectins, but their commercial availability can be
limited [84]. Unfortunately, the quality and purity of these agents varies drastically between vendors, and many
show considerable levels of impurities that can jeopardise the interpretation of glycoproteomics experiments
(Kolarich D, personal observations).

Physicochemical agents for the enrichment of glycopeptides

A variety of enrichment strategies are based on non-biological reagents that target the physicochemical properties
of the glyco-moieties of glycopeptides such as hydrophilicity, size, negative charge or the chemical properties of
specific monosaccharides. These include approaches such as acetone precipitation [85,86], titanium dioxide
(TiO₂) for the enrichment of sialylated glycopeptides [87–89], boronic acid functionalised beads [90,91],
electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) [92,93] or the widely used hydrophilic
interaction liquid chromatography (HILIC) [55,94–96]. These technologies enable the enrichment of intact glyco-
peptides, and in some cases different subsets of intact glycopeptides. The hydrazine coupling approach widely
used for N-glycopeptide enrichment [97–99] is not suitable for intact glycopeptide analysis, as the glycan
components remain covalently attached to the hydrazine beads. Peptides are enzymatically released using peptide \(N\)-glycosidase F (PNGase F), inevitably leading to a loss of structural information of the formerly attached glycan moiety. For more details on the many methods available for glycopeptide enrichment, the readers are referred to several excellent reviews on this topic [35,42,100–104].

No enrichment strategy (or combinations thereof) can quantitatively capture all glycopeptides in complex mixtures, and enrichment also results in the loss of quantitative information on site occupancy levels. Thus, compromises between selectivity and enrichment efficiency will have to be made based on the specific experimental aims of a project.

**How glycosidases can support a glycoproteomics experiment**

\(N\)-glycans can be enzymatically removed using the hydrolytic enzyme PNGase F that efficiently cleaves between the innermost GlcNAc residue of all types of \(N\)-glycans [105], unless these contain an \(\alpha1\)–3 linked core fucose residue as frequently found in plants and invertebrates [106], or are present in truncated forms (e.g. GlcNAc, GlcNAc-Fuc) [107,108]. PNGase F has frequently been used after enrichment as a strategy to reduce sample complexity and facilitate downstream proteomics analyses [109]. The enzymatic release of \(N\)-glycans by PNGase F converts asparagine to aspartate, and care should be taken to avoid misinterpretation of this conversion with the same mass increment (+0.98402 Da) induced by spontaneous deamidation that frequently occurs on asparagine residues in an asparagine–glycine (…NG…) sequon [76,110,111]. The use of heavy water during the PNGase F de-glycosylation reaction can introduce an \(^{18}\text{O}\) into the newly generated aspartate residue, and be used to discriminate spontaneous deamidation from de-glycosylation of asparagine residues [112,113].

Applying de-glycosylation enzymes with different specificities can avoid such false positive identifications of glycosylation sites while indirectly providing some limited, but still useful structural information. Combining endo-\(\beta\)-N-acetylgalactosaminidase (Endo) H (which only cleaves oligomannosidic type \(N\)-glycans between the two GlcNAc residues of the chitobiose core, leaving a single GlcNAc attached to the glycopeptide) and PNGase F, using \(^{18}\text{O}\)-labeling, Cao and co-workers screened and successfully quantified site occupancy levels on HIV gp120 [64]. This approach enables relative quantitation of the macroheterogeneity since the resulting peptides, de-glycosylated peptides and single GlcNAc carrying glycopeptides exhibit similar ionisation efficiencies [56].

In contrast with these endoglycosidases, exoglycosidases digest specific terminal monosaccharide residues from glycans, providing an opportunity to gain insights into key structural features including biologically relevant glyco-epitopes. Exoglycosidase-assisted glycopeptide analysis can be used to determine the level of e.g. \(\alpha2\)–3 linked NeuAc residues in a protein- and site-specific manner [114], or unambiguously determine the presence of sialyl Lewis X epitopes on specific glycans attached to specific sites when performed on glycopeptides from isolated glycoproteins [115,116]. However, such strategies are still to be applied at the glycoproteome-wide scale.

**Glycopeptide labelling strategies can facilitate quantitation and glycopeptide enrichment**

Peptide labelling approaches such as tandem mass tags (TMT) [117–119] or isobaric tags for relative and absolute quantification (iTRAQ) [120,121] have been successfully employed in quantitative glycoproteomics workflows [122–126]. TMT provides accurate relative MS2 (or MS3) -based quantitation and an opportunity for multiplexing to reduce instrument time. Furthermore, it also improves the ET(hc)D fragmentation of glycopeptides by increasing the charge density of labelled glycopeptides [127]. TMT labelling strategies are easily implementable into any clinical glycoproteomics workflow but add sample handling steps and increase costs.

Metabolic derivatization methods are available based on the incorporation of isotopically labelled amino acids (e.g. SILAC) [128–130] (polypeptide-centric labelling) or the incorporation of non-natural monosaccharides [131,132] (glycan-centric labelling) into glycoproteins produced by cultured cells. Monosaccharide-specific click chemistry [133] has been leading the field, where monosaccharides modified with otherwise inert azido-groups replace the natural monosaccharides occurring within the cell. These are eventually incorporated into glycoconjugates and facilitate their enrichment and in situ visualisation. If used for enrichment, the glycan moiety is frequently removed from the peptide for subsequent de-glycosylated peptide analysis [134]. A limitation of these strategies is the variable incorporation efficiency, often generating unlabelled glycoconjugates, as well as the impact of the labelling conditions on cell growth and physiology. Importantly, these reagents have been optimised with respect to applicability in vivo by significantly reducing their cellular toxicity [135]. Even
though click chemistry metabolic labelling is a helpful tool for in vitro and animal model focussed glycoproteomic studies [136], its implementation into clinical glycoproteomics in the foreseeable future is unlikely.

A variety of different labelling strategies have been developed for the characterisation of glycoprotein mixtures of low complexity. Generally, these aim to either increase ionisation efficiency and/or to stabilise specific glycosylation features. Permethylation, a chemical derivatisation method, has been applied to glycopeptides from isolated glycoproteins to obtain more detailed glycan structural information [137]. However, this modification significantly increases the overall hydrophobicity of glycopeptides, which complicates their separation by reversed phase (RP) chromatography, making them unsuitable for glycoproteome-wide analysis. Other forms of glycan derivatisation such as methylamidation have been successfully employed to derivatise and stabilise labile sialic acid residues for MALDI-TOF based glycopeptide profiling [138]. A sialic acid labelling approach using different stabilisation reagents can also be used to distinguish α2–3 from α2–6 linked sialic acid linkages at the glycopeptide level [139].

While these derivatisation strategies have unique benefits, chemical modifications can lead to unexpected reactions that unintentionally increase overall sample heterogeneity, subsequently affecting data analyses. Hence, it is important to consider the benefits and limitations before including glycopeptide labelling in the experimental design.

**Stage 3: separation and identification of glycopeptides**

Most glycoproteomics workflows use advanced online nano-scale separations such as nano-flow liquid chromatography (nano-LC) or electrokinetic separation (e.g. capillary electrophoresis, CE) coupled with ESI–MS to detect and characterise intact glycopeptides. The past decade has seen tremendous advancements in both off- and online separation and detection technologies that have increased the sensitivity, accuracy and throughput of glycoproteomics workflows [140,141]. It would go beyond this mini-review to provide a detailed account on the many different aspects of MS-based glycopeptide separation and detection, hence here we are focussing on a high-level overview of the most important advantages and challenges of widely used techniques. For more details, we refer readers to recent literature on this topic [38–40,44,142].

**MS-coupled separation techniques for glycoproteomics**

In principle, any ultra- or high-performance liquid chromatography (UPLC/HPLC) based separation method that can be performed using MS-compatible solvents can be used to separate complex glycopeptide mixtures prior to MS analysis. RP-LC is without doubt the most widely used separation technique for this purpose due to its unmatched peak capacity, versatility, simplicity and robustness [64,128,143]. A wide selection of different RP materials is available, with a variety of additional separation functionalities, such as improved retention of more hydrophilic compounds (such as glycopeptides) or better separation capacity resulting in reduced LC peak width and thus improved MS signal intensities. In principle, the same LC conditions used for peptide separation can also be employed for the separation of glycopeptides.

Glycopeptides are usually less hydrophobic than their non-glycosylated counterparts, and very hydrophilic glycopeptides might not exhibit sufficient interaction with the RP-matrix when loading the sample in low concentrations of organic solvent as it is commonly done in standard peptide RP-LC. Using pre- and analytical columns optimised to work under completely aqueous conditions can help to capture such very hydrophilic glycopeptides [57], particularly in the case of mucin-type glycopeptides where many sites of glycosylation can be occupied within a single glycopeptide [144]. The loss of hydrophilic glycopeptides can also be minimised by combining different stationary phases such as C18-RP and porous graphitized carbon (PGC), where the latter captures hydrophilic glycopeptides from the RP flowthrough before glycopeptides from both columns are consecutively eluted for MS analyses [73].

Another attractive separation technique, though less widely used, is based on CE separation that can provide excellent separation and sharp peaks with high signal-to-noise ratios [145,146]. CE offers the capacity to separate distinct glycoforms attached to peptides (e.g. linkage difference of sialic acids) based on their charge and physical characteristics [147]. However, the online coupling of CE separation to MS instruments can be challenging due the limited number of MS-compatible electrolytes necessary for electromigration [148], and CE is less frequently employed than LC. Recent advances have made CE-nanoESI systems commercially available. While this separation strategy remains to be applied for system-wide glycoproteomics, its capacity to separate glycans and glycopeptides of purified proteins has recently been shown [147,149,150].
Nano-LC can alternatively be coupled with MALDI-TOF-MS detection, which has the advantage of having the MS analyses decoupled from the LC step, facilitating the re-analyses of each spot at a later time if necessary. This has found some applications for glycoprotein-centric analyses such as IgM [151], or chemically glycosylated vaccine candidate glycoproteins such as cross-reactive material 197 (CRM197) [152]. While several publications describe this system for shotgun clinical proteomics (e.g. [153]), we are not aware of studies that have used LC-MALDI-TOF-MS for glycoproteomics.

Ion-mobility MS (IM-MS) is a recent technology that shows promise for improved analysis of glycopeptides [154,155]. IM-MS is a gas-phase separation method of ions based primarily on their mass and charge but also their size and shape [156]. Given that glycopeptides are usually considerably larger and thus occur in higher charge states than unglycosylated co-eluting peptides, IM-MS provides an opportunity to separate glycopeptides from unglycosylated peptides within an online experiment using field asymmetric ion mobility spectrometry (FAIMS) [157]. While these recent technology developments have not yet found widespread application in glycoproteomics [44], promising data have been published for the characterisation of isolated glycoproteins [158–160], where IM-MS has been reported to be able to distinguish sialic acid linkage isomers (α2–3 or α2–6) from otherwise isobaric glycopeptide precursors [161], and to enable characterisation of isomeric glycopeptides where different sites on the same peptide are glycosylated [162].

**Glycopeptide fragmentation — destructive approaches to decipher glycoproteomes**

Identification of glycopeptides from complex samples would be impossible without fragmentation techniques optimised to deliver information on both the peptide and glycan moieties [163–165]. In combination with MS analysers that acquire product ion spectra with high mass accuracy across a wide m/z range, the fragmentation of glycopeptides may in favourable cases generate sufficient product ion information to facilitate software-assisted identification [166,167]. Glycopeptide fragmentation is perhaps one of the most central aspects within a glycoproteomics experiment, as it generates the fragment ‘fingerprint’ of a specific glycopeptide that is then used to determine the composition of the glycan and the sequence of the peptide. Depending on the type of fragmentation, the site of glycosylation can also be determined from the same product ion spectrum [39,168,169] (Table 1). Hybrid-type MS instruments (e.g. Orbitrap Tribrid) can perform different fragmentation schemes in parallel [164] and excellent reviews have thoroughly discussed the pros and cons of current fragmentation technologies [39,163,165]. Hence, the selection of the fragmentation scheme most suitable for each experiment is crucial to generate informative product ions of both the peptide and glycan moieties. Here, we limit the discussion to the fragmentation technologies most commonly used in glycoproteomics (Table 1).

Collision-induced dissociation (CID) techniques are often used in glycoproteomics but fundamentally differ if performed on ion trap (resonance-type activation) or Q-ToF (beam-type activation) instruments. In principle, both CID-types result largely in fragmentation of the glycan backbone but leave the peptide backbone relatively intact when performed at lower excitation levels sufficient to fragment non-modified peptides (Table 1).

Increasing the energy to achieve higher-energy collisional dissociation (HCD) may result in the generation of sufficient peptide produced ions that facilitate peptide sequence assignment, next to glycan oxonium product ions. Stepped-HCD (sHCD), where the fragmentation energies are being modulated from low to high, delivers more balanced product ion spectra that usually contain more information on both the glycan and peptide moiety of glycopeptides [166].

These fragmentation techniques rarely deliver reliable information on the site of glycan attachment, as achieved by ion-induced dissociation techniques (e.g. electron-transfer dissociation (ETD), electron-capture dissociation (ECD)) [41]. While in most cases these are not necessary for site assignment of N-glycopeptides due to the well-known N-glycosylation sequon (N-X-S/T/C; X≠P) [111], these fragmentation methods become really important when the modification site cannot readily be predicted, such as in O-glycosylation or chemical glycosylation reactions [165]. Hybrid-type fragmentation techniques such as electron-transfer/higher-energy collision dissociation (EThC) can deliver informative product ions from both dissociation techniques [164]. However, EThC takes more time to perform, limiting the cycle time and the overall number of product ion spectra that can be generated within an LC–MS/MS experiment and might not always be necessary to address the research question.
Stage 4: data analyses and bioinformatics

Adequate software tools are the final key to successful glycoproteomics experiments. Different from many other types of PTMs, which can be considered as a single mass value that is either present or absent on the polypeptide chain, glycan modifications can range from a single monosaccharide residue to complex oligo- and polysaccharides, posing unique bioinformatics challenges. Like proteomics, the glycoproteome search space needs to be first appropriately defined. The current software for glycoproteomics differs in how the glycan search space is defined and incorporated into the process of glycopeptide identification [170]. Some tools allow not only the identification but also the relative or absolute quantitation of glycopeptides, and support annotation of

| Fragmentation technique | Schematic representation | Important considerations |
|-------------------------|--------------------------|--------------------------|
| Ion Trap-based CID      | ![Image](cid.png)        | Good fragmentation of the glycan moiety but typically scarce peptide fragmentation [39]. Typically used for glycan-composition determination. Commonly acquired using ion trap (resonance activation). |
| Beam-type CID and HCD   | ![Image](beam.png)       | More informative ion spectra generated than ion trap CID. At low energy levels, similar product ion as obtained in ion trap-based CID. Stepped collision energy HCD (shHCD) offers the advantage of using low and high collision energies on the same precursor for improved identification of the glycan and peptide [167,193–195]. At high energy levels, glycan information is largely limited to oxonium ions and B-ions. Orbitrap c-trap [196], and Q-TOF (beam-type activation) instruments [166]. |
| ETD                    | ![Image](etd.png)        | This fragmentation is very informative on the peptide sequence [197] while leaving the glycan-moiety intact [127,169]. Ideal for peptide and glycan site identification. Commonly acquired on ion trap, FT-MS and orbitrap instruments. |
| EThcD                  | ![Image](ethcd.png)      | Hybrid fragmentation method that provides a supplemental collision activation energy of the ETD reaction [164,198]. This is beneficial in order to identify both the peptide backbone, the glycan composition, and the glycan attachment site within a single spectrum [199–204]. Available on Orbitrap Tribrid instruments. |

The choice of fragmentation scheme depends on (i) instrument availability, (ii) specific aims of an experiment and (iii) available sample amount. Each technique has specific advantages and limitations that need to be balanced based on the individual project aims.
Table 2. Examples for commonly used and recently developed software tools for glycopeptide data analysis (in alphabetical order)

| Software and access | Availability and integration (current version*) | Glycopeptide search strategy and key features | Compatible file types |
|---------------------|------------------------------------------------|---------------------------------------------|----------------------|
| Byonic [205] https://proteinmetrics.com/byos/ | Commercial | *De novo* intact N- and O-glycopeptide identification based on MS/MS data | Mgf, Thermo raw, mzXML |
| GlycoBinder [206] https://github.com/IvanSilbern/GlycoBinder | Freeware | Integration in SugarQuant MS pipeline | Thermo raw |
| GlycoPAT [207] https://virtualglycome.org/glycopat | Freeware | Considers peptide and glycan fragmentation to calculate false discovery rate (FDR) scoring | mzML, dta |
| GlyXtoolMS [208] https://github.com/glyXera/glyXtoolMS | Freeware | Modular tool, allowing control over all phases of analysis | mzML |
| GPQuest [209] https://www.biomarkercenter.org/gpqquest | Freeware | N-glycopeptide analyses | mzML |
| iQ-GPA (GlycoProteome Analyzer) [210] https://www.iqpa.kr | Freeware | N-glycopeptide analyses | Thermo or Bruker raw |
| MetaMorphous [211] https://github.com/smith-chem-wisc/MetaMorphous | Freeware | O-Pair search methodology allows to improve site-specific identification, using paired HCD- and EThcD-MS/MS spectra from LC–MS/MS data | Mgf, Thermo raw, mzML |

Continued
The reliability of the software-based data analysis, however, is strongly influenced by the type and quality of the input data [166], the type of fragmentation method, and factors relating to the search engine and output filtering [171]. A suite of computational tools has been developed over the past decade, which have strongly contributed to the maturation and application of glycoproteomics (Table 2). While these informatics solutions have made impressive progress as summarised in recent reviews [39,103,170,172–174], some challenges remain. A recent inter-laboratory study conducted by the HUPO human glycoproteome initiative (HGI) to evaluate the performance of current software solutions and to identify high-performance search strategies for glycopeptide data analysis, identified several high-performance software solutions, and at the same time demonstrated the significant informatics challenges that remain for glycopeptide data analyses — an important step forward to improve glycoproteomics software performance [171]. We expect some exciting new developments in this space in the coming years, supported by this and other community efforts [171,175,176] and the active integration of multi-dimensional data from different -omics technologies [177] (Figure 2). In the authors’ experience, the ability to incorporate glycomics data into the glycoproteomics workflow, coined as “glycomics-assisted glycoproteomics”, is an example of a particular useful integration of multiple-omics data sets. These allow an informed definition of the glycan search space whilst providing detailed information on the attached glycan structures [79,126,129,141,178–180]. Notably, careful manual review of the data output is still needed to obtain reliable and reproducible results in large-scale glycopeptide data analysis.

### Stage 5: good reporting practice of glycoproteomics data

Data sharing and detailed reporting of MS-based glycoproteomics have become common practice [181,182], providing an opportunity for independent community review and data re-interrogation, but also a valuable resource for other researchers and software developers. However, with the increasing complexity and the enormous amount of data collected within a single experiment, the lack of a detailed and accurate reporting of experimental conditions limits the use by other scientists.
This was recognised several years ago for proteomics and led to the development of essential reporting guidelines (MIAPE) [183]. These guidelines have set an important standard that needs to be followed when submitting LC–MS/MS proteomics data to any of the data repositories under the ProteomeXchange consortium [182]. While many of the important experimental aspects for a glycoproteomics experiment are covered by the MIAPE guidelines, several key aspects of the glycan moiety of glycopeptides require particular attention in the reporting process. The Minimum Information Required for A Glycomics Experiment (MIRAGE) consortium has developed many guidelines (https://www.beilstein-institut.de/en/projects/mirage/) focused on requirements for glycomics experiments that also contain aspects relevant to MS-based glycoproteomics experiments [184–187]. These guidelines are continuously being updated based on community feedback to facilitate the comprehensive reporting of glycomics experimental conditions, with dedicated glycoproteomics guidelines currently being drafted.

To facilitate sharing of glycomics and glycoproteomics data, the GlycoPOST (https://glycopost.glycosmos.org/) platform has been established as a data repository that supports the storage of MS, LC and LC–MS glycomics raw and analysed data, in addition to glycoproteomics data [188].

Several international efforts have been actively connecting glycomics and glycoproteomics data with other relevant glycobiological information to make current glycoproteomics knowledge more accessible. Initiatives such as Glyconnect (https://glyconnect.expasy.org/) [189], Glycomics@Expasy (www.expasy.org), GlyCosmos (https://glycosmos.org/) [190] and GlyGen (https://www.glygen.org/) [191] (see also http://www.glyspace.org/ [192]) have started to systematically curate glycoproteomics and glycomics data while linking information across each platform. These communal efforts will significantly facilitate the integration of glycoproteomics data into other -omics research resources.

Conclusion
This review provides a concise overview of the methods now available for glycoproteomics analyses, with the intention to inform researchers that are new to the field, as well as experienced proteomics scientists that are considering jumping into the exciting wild waters of glycoproteomics. Glycoproteomics technologies have experienced a tremendous evolution over the past two decades, starting from the profiling of single glycoproteins, to now allowing large-scale system-wide analyses of complex samples as an integral part of multi-omics studies. Transcriptomics of glycosylation pathway relevant enzymes informs on how these pathways could be affected under studied conditions, metabolomics delivers important information on glycosylation precursors such as nucleotide sugar substrates that, in concert with (glyco)lipidomics and proteoglycomics, proteomics, glycoproteomics and genomics, can deliver a detailed picture of the highly interconnected cellular glycosylation pathways and how these are affected in diseases (Figure 2). These exciting developments will undoubtedly lead to an increased understanding of the function of glycoproteins in health and disease. It is also clear that these technologies are opening a new era in glycoscience that will, in combination with the other -omics techniques, deliver previously overlooked functional insights into the ubiquitous modification of proteins by glycans. Glycoproteomics analysis is and will increasingly become an indispensable part of understanding the molecular basis of life.

Perspectives
- Glycoproteomics is becoming a powerful tool in glycobiology that enables the system-wide mapping of protein-specific glycosylation features.
- Understanding of protein-specific glycosylation and how it is impacted in diseases provides novel opportunities for precision diagnostics and therapies.
- Integration of glycoproteomics (and glycomics) into multi-omics studies is important to capture the glyco-language of cells and organisms.

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.
Funding
D.K. is the recipient of an Australian Research Council Future Fellowship (project number FT160100344) funded by the Australian Government. This work was made possible by a Griffith University Postgraduate Research Scholarship (GUPRS) awarded to TO. This work was supported by the Australian Research Council (CE140100003).

Open Access Statement
Open access for this article was enabled by the participation of Griffith University in an all-inclusive Read & Publish pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

Author Contributions
Manuscript conceptualization: T.O. and D.K.; Manuscript writing, comments and editing: T.O., M.T.-A., N.P., and D.K.

Abbreviations
CDGs, congenital disorders of glycosylation; CID, collision-induced dissociation; EThcD, electron-transfer/higher-energy collision dissociation; FASP, filter-aided sample preparation; HCD, higher-energy collisional dissociation; LAC, lectin affinity chromatography; PTMs, protein post-translational modifications; RP, reversed phase; TMT, tandem mass tags.

References
1 Moremen, K.W., Tiemeyer, M. and Nairn, A.V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. Nat. Rev. Mol. Cell Biol. 13, 448–462 https://doi.org/10.1038/nrm3383
2 Varki, A. (2017) Biological roles of glycans. Glycobiology 27, 3–49 https://doi.org/10.1093/glycob/cww086
3 Kuo, C.-H., Gandhi, J., Zia, J.G., and Paszek, R.N. and J, M. (2018) Physical biology of the cancer cell glycocalyx. Nat. Phys. 14, 658–669 https://doi.org/10.1038/s41567-018-0186-9
4 van Kooyk, Y. and Rabinovich, G.A. (2008) Protein-glycan interactions in the control of innate and adaptive immune responses. Nat. Immunol. 9, 593–601 https://doi.org/10.1038/nri2103
5 Yang, X. and Qian, K. (2017) Protein O-GlcNAcylation: emerging mechanisms and functions. Nat. Rev. Mol. Cell Biol. 18, 452–465 https://doi.org/10.1038/nrm.2017.22
6 Ma, J. and Hart, G.W. (2017) Analysis of protein O-GlcNAcylation by mass spectrometry. Curr. Protoc. Protein Sci. 87, 24.10.21–24.10.16
7 Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3, 97–130 https://doi.org/10.1093/glycob/3.2.97
8 Stanley, P., Taniguchi, N. and Aebi, M. (2015) N-Glycans. In Essentials of Glycobiology (Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L. and Seeberger, P.H., eds), pp. 99–111, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
9 Rely, C., Stewart, T.J., Rentfrow, M.B. and Novak, J. (2019) Glycosylation in health and disease. Nat. Rev. Nephrol. 15, 346–366 https://doi.org/10.1038/s41581-019-0129-4
10 Landsteiner, K. and van der Scheer, J. (1925) On the antigens of red blood corpuscles: the question of lipoid antigens. J. Exp. Med. 41, 427–437 https://doi.org/10.1084/jem.41.3.427
11 Ewald, D.R. and Sumner, S.C. (2016) Blood type biochemistry and human disease. Wiley Interdiscip. Rev. Syst. Biol. Med. 8, 517–535 https://doi.org/10.1002/wsbm.1355
12 Rabinovich, G.A. and Toscano, M.A. (2009) Turning ‘sweet’ on immunity: galectin-glycan interactions in immune tolerance and inflammation. Nat. Rev. Immunol. 9, 338–352 https://doi.org/10.1038/nri2536
13 Dias, A.M., Pereira, M.S., Padrao, N.A., Alves, I., Marcos-Pinto, R. and Sá, J.M. (2018) Glycans as critical regulators of gut immunity in homeostasis and disease. Cell Immunol. 333, 9–18 https://doi.org/10.1016/j.cellimm.2018.07.007
14 Christiansen, M.N., Chik, J., Lee, L., Anugraham, M., Abraham, J.L. and Packer, N.H. (2014) Cell surface protein glycosylation in cancer. Proteomics 14, 525–546 https://doi.org/10.1002/pmic.201300387
15 Stickwell, S.R., Ju, T. and Cummings, R.D. (2015) Protein glycosylation in cancer. Annu. Rev. Pathol. 10, 473–510 https://doi.org/10.1146/annurev-pathol-012414-040438
16 Pinno, S.S. and Reis, C.A. (2015) Glycosylation in cancer: mechanisms and clinical implications. Nat. Rev. Cancer 15, 540–555 https://doi.org/10.1038/nrc3982
17 Munkley, J. and Elliott, D.J. (2016) Hallmarks of glycosylation in cancer. Onco Targets Ther. 9, 35478–35489 https://doi.org/10.18632/oncotarget.8155
18 Mereiter, S., Balmana, M., Campos, D., Gomes, J. and Reis, C.A. (2019) Glycosylation in the era of cancer-targeted therapy: where are we heading? Cancer Cell 36, 6–16 https://doi.org/10.1016/j.ccell.2019.06.006
19 Chatterjee, S., Lee, L.Y., Kawahara, R., Abrahams, J.L., Adamczyk, B., Anugraham, M. et al. (2019) Protein paucimannosylation is an enriched N-glycosylation signature of human cancers. Proteomics 19, e1900010 https://doi.org/10.1002/pmic.201900010
20 Schedin-Weiss, S., Winblad, B. and Tjernberg, L.O. (2014) The role of protein glycosylation in Alzheimer disease. FEBS J. 281, 46–62 https://doi.org/10.1111/febs.12590
21 Fang, P., Xie, J., Sang, S., Zhang, L., Liu, M., Yang, L. et al. (2020) Multilayered N-glycoproteome profiling reveals highly heterogeneous and dysregulated protein N-glycosylation related to Alzheimer’s disease. Anal. Chem. 92, 867–874 https://doi.org/10.1021/acs.analchem.9b03555
1656

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Downloaded from http://portlandpress.com/biochemsoctrans/article-pdf/49/4/1643/919901/bst-2020-0879c.pdf by guest on 24 February 2022
203 Caval, T., Zhu, J., Tian, W., Remmelzwaal, S., Yang, Z., Clausen, H. et al. (2019) Targeted analysis of lysosomal directed proteins and their sites of mannose-6-phosphate modification. Mol. Cell Proteom. 18, 16–27. https://doi.org/10.1074/mcp.M118.000967

204 Caval, T., Zhu, J. and Heck, A.J.R. (2019) Simply extending the mass range in electron transfer higher energy collisional dissociation increases confidence in N-glycopeptide identification. Anal. Chem. 91, 10401–10406 https://doi.org/10.1021/acs.analchem.9b02125

205 Bern, M., Kil, Y.J. and Becker, C. (2012) Byonic: advanced peptide and protein identification software. Curr. Protoc. Bioinform. Chapter 13, Unit13 20 https://doi.org/10.1002/0471256953.bi1320s40

206 Fang, P., Ji, Y., Silbern, I., Doebele, C., Ninov, M., Lenz, C. et al. (2020) A streamlined pipeline for multiplexed quantitative site-specific N-glycoproteomics. Nat. Commun. 11, 5268 https://doi.org/10.1038/s41467-020-19052-w

207 Liu, G., Cheng, K., Lo, C.Y., Li, J., Qu, J. and Neelamegham, S. (2017) A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. Mol. Cell Proteom. 16, 2032–2047 https://doi.org/10.1074/mcp.M117.68239

208 Ploch, M., Hoffmann, M., Prašlova, A., Reichl, U. and Rapp, E. (2018) Glyxtool(MS): an open-source pipeline for semiautomated analysis of glycopeptide mass spectrometry data. Anal. Chem. 90, 11908–11916 https://doi.org/10.1021/acs.analchem.8b02087

209 Tgi Eshghi, S., Shah, P., Yang, W., Li, X. and Zhang, H. (2015) GPOuest: a spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact N-glycopeptides. Anal. Chem. 87, 5181–5188 https://doi.org/10.1021/acs.analchem.5b00024

210 Park, G.W., Kim, J.Y., Hwang, H., Lee, J.Y., Ahn, Y.H., Lee, H.K. et al. (2016) Integrated GlycoProteome analyzer (I-GPA) for automated identification and quantitation of site-specific N-glycosylation. Sci. Rep. 6, 21175 https://doi.org/10.1038/srep21175

211 Lu, L., Riley, N.M., Shortreed, M.R., Bertozzi, C.R. and Smith, L.M. (2020) O-pair search with metaMorpheus for O-glycopeptide characterization. Nat. Methods 17, 1133–1138 https://doi.org/10.1038/s41592-020-00985-5

212 Polasky, D.A., Yu, F., Teo, G.C. and Nesvizhskii, A.I. (2020) Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. Nat. Methods 17, 1125–1132 https://doi.org/10.1038/s41592-020-0967-9

213 Zeng, W.-F., Liu, M.-Q., Zhang, Y., Wu, J.-Q., Fang, P., Peng, C. 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 https://doi.org/10.1038/srep25102

214 Baker, P.R., Trinidad, J.C. and Chalkley, R.J. (2011) Modification site localization scoring integrated into a search engine. Mol. Cell Proteom. 10, M111.008078 https://doi.org/10.1074/mcp.M111.008078

215 Kossiner, D., Chambers, M., Burke, R., Agus, D. and Mallick, P. (2008) Proteowizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536 https://doi.org/10.1093/bioinformatics/btn323