Biochimica et Biophysica Acta xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the *N*- and *O*-glycoproteomes

Morten Thaysen-Andersen *, Nicolle H. Packer

Biomolecular Frontiers Research Centre, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia

ARTICLE INFO

Article history:
Received 2 March 2014
Received in revised form 23 April 2014
Accepted 5 May 2014
Available online xxxx

Keywords:
Glycoproteomics
Glycopeptide
Glycoprotein
Mass spectrometry
Liquid chromatography
Protein glycosylation
Glycoproteome
Site-specific
N-glycan
O-glycan

ABSTRACT

Site-specific structural characterization of glycoproteins is important for understanding the exact functional relevance of protein glycosylation. Resulting partly from the multiple layers of structural complexity of the attached glycans, the system-wide site-specific characterization of protein glycosylation, defined as glycoproteomics, is still far from trivial leaving the *N*- and *O*-linked glycoproteomes significantly under-defined. However, recent years have seen significant advances in glycoproteomics driven, in part, by the developments of dedicated workflows and efficient sample preparation, including glycopeptide enrichment and prefractionation. In addition, glycoproteomics has benefitted from the continuous performance enhancement and more intelligent use of liquid chromatography and tandem mass spectrometry (LC–MS/MS) instrumentation and a wider selection of specialized software tackling the unique challenges of glycoproteomics data. Together these advances promise more streamlined *N*- and *O*-linked glycoproteome analyses. Tangible examples include system-wide glycoproteomics studies detecting thousands of intact glycopeptides from hundreds of glycoproteins from diverse biological samples. With a strict focus on the system-wide site-specific analysis of protein *N*- and *O*-linked glycosylation, we review the recent advances in LC–MS/MS based glycoproteomics. The review opens with a more general discussion of experimental designs in glycoproteomics and sample preparation prior to LC–MS/MS based data acquisition. Although many challenges still remain, it becomes clear that glycoproteomics, one of the last frontiers in proteomics, is gradually maturing enabling a wider spectrum of researchers to access this new emerging research discipline. The next milestone in analytical glycobiology is being reached allowing the glycoscientist to address the functional importance of protein glycosylation in a system-wide yet protein-specific manner.

1. Introduction to system-wide site-specific analysis of protein glycosylation

1.1. Protein N- and O-glycosylation structures and functions

Protein glycosylation is the covalent attachment of complex carbohydrates or oligosaccharides (here collectively called glycans) to specific amino acid residues of the polypeptide backbone of proteins. The biosynthetic machinery of mammals primarily allows the attachment of glycans to asparagine (*N*-glycosylation) and serine/threonine (*O*-glycosylation) residues thereby forming two major classes of protein glycosylation. These types i.e. *N*-GlcNAc and *O*-GalNAc (mucin-type) glycosylation, which are both the focus of this review, are synthesized in a non-template driven manner by a spectrum of glycosylation enzymes through different routes in the secretory pathway. The *N*-linked glycans, which in mammals are usually restricted to occupy asparagine residues in NXS/T, X ≠ P consensus sequences (sequons), consist of a common chitobiose core ($Man_3GlcNAc_2$) from which a variety of monosaccharides and other glycan modifications may be added to the non-reducing termini in specific linkage configurations [1]. The status of the cellular glycosylation machinery and the nature of the proteins undergoing glycosylation together determine the repertoire of glycans being presented on the protein carriers thus creating the important features of cell- and protein-specific glycosylation [2]. *N*-linked glycans are usually larger by mass and volume than their *O*-linked counterparts, which, in contrast, are more

Abbreviations: CID, collision induced dissociation; Con A, concanavalin A; CSF, cerebrospinal fluid; DDA, data dependent acquisition; DIA, data independent acquisition; ECD, electron capture dissociation; EIC (or XIC), extracted ion chromatogram; ESI, electrospray ionization; ETD, electron transfer dissociation; FDR, false discovery rate; FT-ICR, Fourier transform ion cyclotron resonance; Fuc, fucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; HCD, higher-energy collisional dissociation; HILIC, hydrophilic interaction liquid chromatography; HPLC, high performance liquid chromatography; IRMPD, infrared multiphoton dissociation; LC–MS/MS, liquid chromatography tandem mass spectrometry; LTQ, linear trap quadrupole; MALDI, matrix assisted laser desorption ionization; Man, mannose; MRM, multiple reaction monitoring; MS^n, mass spectrometry to the nth power; nETD, negative electron transfer dissociation; NeuAc, *N*-acetyl-5-neuraminic acid; Q-TOF, quadrupole-time-of-flight; RP, reversed phase; SPE, solid phase extraction; TMT, tandem mass tag; UPLC, ultra-high pressure liquid chromatography

* Corresponding author at: Department of Chemistry and Biomolecular Sciences, Macquarie University, NSW 2109, Australia. Tel.: +61 2 9850 7487; fax: +61 2 9850 6192.
E-mail address: morten.andersen@mq.edu.au (M. Thaysen-Andersen).

http://dx.doi.org/10.1016/j.bbapap.2014.05.002
1570-9639/

frequently attached in the serine- and threonine-rich regions of mucin and mucin-like proteins. Several types of non-mucin *O*-glycosylation are also known to exist in mammals including *O*-GlcNAcylation [3], *O*-mannosylation [4] and *O*-fucosylation [5], but are not addressed in this review. The general structures, biosynthetic pathways and associated disorders of *N*- and *O*-linked glycosylation have been reviewed elsewhere in detail and readers are encouraged to use these resources for a more general introduction to protein glycosylation [1,6].

As expected from the conservation of mammalian protein glycosylation sites and glycan structures and the significant amount of cellular energy utilized to ensure proper regulation of the biosynthetic machinery, *N*- and *O*-linked glycans have repeatedly been shown to be important for generating and maintaining the structural and functional integrity of their carrier proteins [1]. Protein *N*- and *O*-glycosylation is also involved in development processes [7,8], in bacterial binding to the host [9] and to sustain the normal function of individual cells and indeed also of the whole tissue, organ and organism [10]. Finally, glycosylation is a highly regulated protein modification in cells and is known to be affected by various pathologies including, but not restricted to, cancer [11,12], inflammation [13], Alzheimer disease [14], multiple sclerosis [15], cystic fibrosis [9] relative to cells exposed to 'normal' physiology and homeostasis.

1.2. Multiple structural layers of protein glycosylation

Interpretation of the glycosylation code has been intriguing to past generations of glycobiologists. The ability to accurately characterize the structure of glycoproteins is of importance to unravel the diverse functions of glycans. Better understanding of protein glycosylation will facilitate the development of next generation of glycoprotein-based therapeutics [16]. However, relative to simple protein modifications, protein glycosylation is more challenging to characterize due to the sub-stoichiometry and heterogeneity formed by multiple layers of structural diversity at the primary structural level of the glycan. Variations in the monosaccharide compositions, overall topology/branching patterns and linkage types are common features creating a spectrum of closely related glycan species. The amount of information needed to unambiguously characterize a glycoprotein is consequently much larger than for unmodified proteins and proteins modified by other chemical groups e.g. phosphorylation and methylation. Even the most detailed characterization studies usually only capture part of the glycoprotein structure leaving some structural aspects unknown. Specific structural features may be assumed/predicted from the established 'synthetic restriction rules' derived from the well-studied biosynthetic machinery of mammalian protein glycosylation [17]. It is important to stress that glycobiological function often can be learnt from less-than-complete structural information of glycoproteins. In other cases, the functional understanding of glycobiology is hidden in the fine details of the glycoprotein structure supporting more thorough glycoprotein characterization requirements.

1.3. Bottom-up site-specific glycoprofiling of isolated glycoproteins

Ever since glycosylated proteins were discovered there has been a desire to accurately determine the glycoprotein structure. Historically, characterization of isolated glycoproteins of relatively high amounts (>pico-nanomolar) has been achieved using an array of analytical techniques to obtain in-depth information of the glycoprotein structure [18]. This has indeed advanced our understanding of the structure/function relationship of many glycoproteins [19–22]. The majority of these techniques are based on so-called bottom-up principles where glycoproteins are hydrolyzed and the fragments including monosaccharides, released glycans and proteolytically generated glycopeptides, are analyzed separately [23]. The structural information from the individual fragments is then pieced together to learn about the structure of the entire glycoprotein. Modern techniques are centered around high performance liquid chromatography (HPLC) [24,25] and liquid chromatography conjugated on-line with mass spectrometry (LC–MS) [26–33] of fluorescently labeled and native glycans or intact glycopeptides. Only few analytical techniques are presently capable of handling and generating information-rich data from intact glycoproteins, in particular when analyzed in complex mixtures e.g. from 2D gels [34], leaving the conventional bottom-up methods still the preferred analytical approach.

Analysis of glycans released from their carrier proteins represents a useful and relatively easy experimental approach with which to generate in-depth glycan structural information such as knowledge of the monosaccharide composition, branching topology and stereoisomeric linkage of the monosaccharide sequence. However, although profiling of released glycans can generate a holistic view of the global cell glycome or cellular sub-glycomes, glycomics alone intrinsically lacks the ability to provide information about the carrier protein and the specific attachment site of each glycan. In contrast, glycopeptides (and glycoproteins) provide an avenue for establishing direct evidence of the connectivity between the glycan and the polypeptide backbone. As discussed later, less information about the glycan structure can usually be obtained from glycopeptide-based approaches. Knowledge about the specific glycan attachment site becomes particularly needed for glycoproteins containing multiple glycosylation sites or when mixtures of glycoproteins are analyzed. Established protocols for site-specific glycosylation analysis of isolated proteins are available [33,35]; however, due to the extensive variety in glycoprotein architecture mostly arising from variations in the polypeptide sequences and positions of sequons relative to available proteolytical sites such methods must be customized to be suitable for the glycoprotein of interest. Often glycan- and glycopeptide-based analyses are carried out in parallel to obtain complementary and in-depth site-specific information of isolated glycoproteins [19,33,36–38]. Site-specific characterization and relative quantitation of approximately 170 mammalian glycoproteins has been recently curated by the authors based on published literature [2]. From this we estimate that glycan structures have been partly or completely characterized in a site-specific manner for approximately only 400–500 of the ~10,000 predicted human glycoproteins [39]. Together with the knowledge that protein glycosylation varies dramatically in a spatial and temporal manner, it is thus clear that the vast majority of the human (let alone the whole mammalian) glycoproteome remains uncharacterized. Hence, glycoscientists have long argued the need for the development of analytical techniques capable of performing site-specific analysis of protein glycosylation at the larger-scale to increase the coverage of the glycoproteome.

1.4. Defining glycomics and glycoproteomics

Glycomics is commonly defined as the conjugate-unspecific analysis of the glycome at a system-wide level. This includes, but is not limited to, the global analysis of glycans released from proteins derived from isolated cells, tissues or body fluids. The term glycomics often implicitly refers to the analysis of the *N*- and/or *O*-glycomes, but should more correctly be specified in case-by-case studies to avoid confusion with the global analysis of other glycoconjugates including glycolipids, GPI anchors or free glycans. Several semi-automated sensitive glycomics workflows are now available for low/medium through-put analysis of the *N*- and *O*-glycomes [23,35,40,41]. An excellent overview of recent developments in glycan analysis by mass spectrometry was recently published [42].

Glycoproteomics is the protein-specific counterpart to glycomics defined as the site-specific analysis of the glycoproteome at the system-wide level. Thus, glycoproteomics yields information about the protein carriers, the glycan attachment sites and the structure and occupancy of the glycan. Glycopeptides, and much less frequently intact glycoproteins, are the target analytes for glycoproteomics. As discussed in this work, techniques and workflows for glycoproteomics remain immature

Please cite this article as: M. Thaysen-Andersen, N.H. Packer, Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteomes, Biochim. Biophys. Acta (2013), http://dx.doi.org/10.1016/j.bbapap.2014.05.002