Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Microbial glycoproteomics

Adnan Halim[1,*] and Jan Haug Anonsen[2]

Mass spectrometry-based "–omics" technologies are important tools for global and detailed mapping of post-translational modifications. Protein glycosylation is an abundant and important post translational modification widespread throughout all domains of life. Characterization of glycoproteins, including identification of glycan structure and components, their attachment sites and protein carriers, remains challenging. However, recent advances in glycoproteomics, a subbranch that studies and categorizes protein glycosylations, have greatly expanded the known protein glycosylation space and research in this area is rapidly accelerating. Here, we review recent developments in glycoproteomic technologies with a special focus on microbial protein glycosylation.

Addresses
[1] Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark
[2] Center for Integrative Microbial Evolution, The Mass Spectrometry and Proteomics Unit, Department of Biosciences, University of Oslo, Blindernveien 31, 0371 Oslo, Norway

Corresponding authors: Halim, Adnan (halim@sund.ku.dk), Anonsen, Jan Haug (j.h.anonsen@ibv.uio.no)
[*] Present address: Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065, USA.

Current Opinion in Structural Biology 2017, **44**:143–150

This review comes from a themed issue on **Carbohydrates**

Edited by **Ute Krengel** and **Thilo Stehle**

http://dx.doi.org/10.1016/j.sbi.2017.03.005

Introduction

The elaborate process of protein glycosylation (pgl) is catalyzed by enzymatic machineries (transferases and hydrolases) capable of synthesizing hundreds of different glycan structures that collectively add an immense structural complexity and flexibility to the proteome. Evolutionary conservation and shared commonalities for protein glycosylations are present in all three domains of life, signifying the importance of widespread protein glycosylation [1,2]. Glycoproteins may display common and/or unique structures in individual organisms and/or cell types, inevitably presenting a major analytical challenge for investigators attempting to catalogue site-specific glycosylations on a global scale. However, recent technological advances in mass spectrometry coupled with methodological improvements for glycopeptide enrichment has led to a breakthrough in the field. Glycoproteomics, defined as "*the systems-level analysis of glycoproteins, including their protein identities, sites of glycosylation, and glycan structures*" [3] may involve a number of methods and technologies, but the sensitivity and speed of mass spectrometry (MS) has made it a popular choice for glycoproteomic investigations and the inclusion of MS is now almost universal. Together with complementary –omics technologies, the protein glycosylation landscape is now being resolved in much greater detail and depth. Below, we discuss the basic concepts of MS-based glycoproteomics and its applications for glycan structural elucidation and the identification of target proteins and sites in microbial (Bacteria, Archaea and yeast) glycoproteomes.

Microbial glycosylation

Protein glycosylation is commonly classified based on the chemistry of the protein-glycan linkage, in microbes this involve either *N*-linked, *O*-linked, or *S*-linked glycans. Microbial protein-attached glycans range from small, mono- and disaccharides to large and complex glycans, consisting of multiple repeating carbohydrate units and extensive branching. The regulation, enzymology and biology of microbial glycosylations are extremely diverse and beyond the scope of this review (for current in-depth review on various microbial protein glycosylation systems readers are referred to [4••,5•,6••,7–10]).

Nevertheless, the general concepts of microbial pgl may be summarized as follows. Yeast, bacteria and Archaea share an evolutionary related canonical *N*-linked pgl systems that occur in a conserved fashion [11]; Lipid-linked glycans are added by a membrane bound oligosaccharyltransferase (N-OST) to the consensus sequence (sequon) N-X-S/T (X ≠ P) on protein targets. The membrane bound N-OST responsible for the "*en bloc*" transfers of glycan to protein targets in bacteria (PglB) or Archaea (AglB) are homologs to the yeast N-OST catalytic subunit Stt3. However, recent work indicates that a second class of N-OST might be operating in Archea [12]. Interestingly, some bacterial N-OST exhibit increased site specificity and require an extended target sequon compared to the yeast and Arcaeal homologs [13,14]. Although the majority of pgl systems described in Archaea are *N*-linked only some bacterial pgl systems belong to the canonical *N*-linked system. Alternatively, bacterial *N*-linked glycosylation can also result from the action of non-canonical HMW1C-like (after the

nontypeable Haemophilus influenza high molecular weight adhesin) cytosolic glycosyltransferases (N-GT). Although HMW1C-like N-GTs transfer mono- and di-hexoses preferably onto the canonical N-X-S/T sequon they exhibit highly relaxed site specificity and are capable of glycosylating Asn in a range of tri-peptide sequences [15,16].

In contrast to *N*-linked glycosylation, *O*-linked glycosylation commonly does not occur in a sequon but rather on hydroxyl containing amino acids in regions of low structural rigidity. A notable exception is the broad spectrum *O*-linked protein glycosylation system operating in all classes of bacteroidetes targeting the sequon D-S/T-A/L/V/I/M/T [17]. Moreover and in contrast to the *N*-linked pgl system, *O*-linked pgl systems can potentially modify more than a single type of hydroxyl bearing amino acid. In baker's yeast, *O*-glycosylation biosynthesis is limited to the *O*-linked mannose (O-Man) type, where glycoprotein transit through the secretory pathway leads to the assembly of linear O-$Man_{2\text{-}5}$ glycans on Ser/Thr residues [8,9]. A second type of O-Man glycosylation, found on nuclear and cytoplasmic yeast proteins has also been reported (described below), however, the enzymatic machinery responsible for this type of glycosylation has not been identified yet [18••].

In contrast to the relatively modest set of approx. 10 monosaccharide units employed to build mammalian *N*- and *O*-linked glycans [19,20••], prokaryotic glycans contain many unique and unusual monosaccharides exclusively found in these organisms [21]. Moreover, prokaryotic glycans can also be enzymatically modified by chemical groups ranging from methylation to attachment of amino acids. Furthermore, whereas some prokaryotic pgl systems express a limited number of glycoforms, other prokaryotes display extensive inter- and intra-strain glycan diversity. The presence of unusual and unique sugar components in addition to extensive microheterogeneity in bacterial and archaeal pgl systems generate challenges to MS-driven glycoproteomic studies. The same holds true in microbes with relatively simple glycosylations, for example, baker's yeast, where only two monosaccharides (Man and N-acetylglucosamine(GlcNAc)) are utilized to construct their entire glycome and heterogeneous glycoproteome. In the following sections, we discuss the basic concepts of mass spectrometry and how this technique has been utilized to address the analytical challenges of microbial glycoproteomics.

MS-driven microbial glycoproteomics

Bottom-up workflows are commonly adopted in glycoproteomics, enabling separation and ionization of (glyco) peptides by reversed-phase liquid chromatography (LC) and electrospray ionization (ESI), respectively. Protein- or peptide-level enrichment strategies (discussed below), are typically included and aim to increase the relative abundance of glycopeptides before MS analysis to overcome the inherent suppression of glycopeptide signals relative to unmodified forms in complex biological samples [20••]. During LC-ESI MS/MS, glycopeptide precursor ions are usually characterized by two distinct fragmentation modes: 1) collision-driven, typically collision induced dissociation (CID) or higher-energy collisional dissociation (HCD) and 2) radical-driven, either electron capture dissociation (ECD) or electron transfer dissociation (ETD) together referred to as ExD (Figure 1).

The combination of low (Figure 1a and b) and higher (Figure 1c) normalized collision energy (NCE) fragmentation, commonly the CID/HCD combination or a stepped NCE strategy has proven effective for the characterization of glycan structures and sequence identification of prokaryotic broad-spectrum glycoproteomes as well as eukaryotic *O*-linked glycoproteomes [20••,22]. Although higher NCE (HCD) commonly results in simultaneous glycan and peptide backbone fragmentation (Figure 1c) facilitating the identification of glycopeptide, the labile glycosidic linkages are rarely retained and hence only glycosylation sites in sequons are identified by higher NCE MS. The possibility of using collision fragmentation (CID/HCD) for glycosite identification has been previously suggested [20••,23–25]. However, the prerequisite is that glycan-retaining peptide fragments are produced in sufficient quantities to be detected. Primarily this means that the glycopeptides targeted for fragmentation need to carry a 'mobile' proton [26] and to apply collisional energy in the optimal range to generate peptide fragments retaining the glycan. Different peptide bonds have different activation barriers for dissociation, and thus the optimal collision energy will depend upon peptide sequence [26,27]. Using CID/HCD for glycosite identification therefore requires the analysis of sufficiently charged glycopeptides using the correct dissociation energy.

ExD is preferentially employed for glycosite identification (Figure 1d) [20••,28]. Since glycosidic bonds commonly exhibit lower activation barriers for dissociation, collision activation typically fragments glycosidic bonds over peptide bonds. The advantage of ExD is that it avoids the preferential fragmentation of labile glycosidic bonds because this dissociation process does not involve energy redistribution [29], the glycan therefore remains attached to the modified amino acid. This is the main reason why ExD has been employed for glycosite identification, although the same principle can be employed for glycopeptide identification. However, ExD is inherently inefficient compared to collisional fragmentation.

Although high resolution precursor and fragment detection is commonly sufficient to determine glycan sequence and composition, investigating secondary glycan