Biochimica et Biophysica Acta 1850 (2015) 33–42

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Advances in mass spectrometry driven O-glycoproteomics

Steven B. Levery [1], Catharina Steentoft, Adnan Halim, Yoshiki Narimatsu, Henrik Clausen, Sergey Y. Vakhrushev *

Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form 24 September 2014
Accepted 25 September 2014
Available online 2 October 2014

Keywords:
Orbitrap
ETD
SimpleCell
Glycopeptide
Lectin
Site-specific

ABSTRACT

Background: Global analyses of proteins and their modifications by mass spectrometry are essential tools in cell biology and biomedical research. Analyses of glycoproteins represent particular challenges and we are only at the beginnings of the glycoproteomic era. Some of the challenges have been overcome with N-glycoproteins and proteome-wide analysis of N-glycosylation sites is accomplishable today but only by sacrificing information of structures at individual glycosites. More recently advances in analysis of O-glycoproteins have been made and proteome-wide analysis of O-glycosylation sites is becoming available as well.
Scope of review: Here we discuss the challenges of analysis of O-glycans and new O-glycoproteomics strategies focusing on O-GalNAc and O-Man glycoproteomes.
Major conclusions: A variety of strategies are now available for proteome-wide analysis of O-glycosylation sites enabling functional studies. However, further developments are still needed for complete analysis of glycan structures at individual sites for both N- and O-glycoproteomics strategies.
General significance: The advances in O-glycoproteomics have led to identification of new biological functions of O-glycosylation and a new understanding of the importance of where O-glycans are positioned on proteins.

1. Introduction

Independent of other mechanisms for variational control of protein structure – e.g., alternative splicing and mRNA editing – post-(or co-) translational modifications (PTMs) provide for virtually unlimited expansion of the structural space occupied by all the proteins of an organism (proteome). This expansion, in turn, adds manifold layers of potentially dynamic complexity to an organism's global molecular interaction network (interactome). Of the more than 400 known PTMs [1], protein glycosylation is one of the most abundant, and certainly the most structurally diverse. Lowe and Marth [2] estimated that genes involved in glycosylation occupy at a minimum of 1% of mammalian genomes, which supports the idea that widespread persistence and diversity of glycosylation must underpin processes of major functional – and therefore adaptive – importance. This is not to say that each and every residue at every site of glycosylation carries a unique functional consequence, any more than every single nucleotide polymorphism (SNP) or amino acid substitution does; nevertheless, in full analogy, a significant number are known to have consequences, either individually or in aggregate. This serves to highlight an important purpose for mapping as many glycosylation sites and structural variants as possible — to facilitate discovery of novel biological functions against the background of variational noise. Currently, for example, a role of site-specific α-GalNAc ("mucin-type") O-glycosylation as a regulator of protein function is a dynamic area of investigation; most recently this led to discovery of a mechanism for co-regulation of proprotein convertase processing of proteins [3–6]. A major challenge in the field has been to fully characterize the human GalNAc O-glycoproteome in order to uncover other site-specific biological functions of this abundant PTM. Recent progress in the field of mass spectrometry driven site-specific O-glycoproteomics – greatly facilitated by high throughput genome editing technologies [7], along with extensive application of electron transfer dissociation (ETD) fragmentation – is emphasized in this review. Throughout, glycosylation in human cells will be referred to, unless otherwise indicated.

2. Common problems complicate structural and functional analysis of O-glycosylation

Glycosylation pathways themselves are highly diverse, including both N-linked (to Asn residues) and O-linked (mainly to Ser/Thr, but also possible to Tyr and HyL residues) biosynthesis. N-linked glycosylation is unique in that it is initiated by en bloc transfer of a pre-assembled complex glycan from a membrane-bound lipid phosphate glycan to the acceptor Asn residue; transfer requires, almost exclusively, recognition

* Corresponding author at: Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. Tel.: +45 35327797; fax: +45 35367980.
E-mail address: seva@sund.ku.dk (S.Y. Vakhrushev).
[1] A near final draft of this review prepared by Steven before his sudden passing has been edited and completed by the rest of the authors. In memory of a great friend and scientist.

http://dx.doi.org/10.1016/j.bbagen.2014.09.026
0304-4165/

of a tripeptide consensus sequence, Asn-Xxx-Ser/Thr (Xxx ≠ Pro), and is known to occur co- as well as post-translationally. This process, including the entire structure of the prefabricated donor glycan, is conserved throughout eukaryotes; considerable structural remodeling occurs post-transfer, and this processing diverges between the eukaryote kingdoms and their various orders. Further details are outside the scope of this review, but are well covered elsewhere [8].

O-glycosylation, in contrast, is initiated by transfer of one of six different monosaccharides besides α-GalNAc: these are β-GlcNAc, α-Man, α-Fuc, β-Xyl, β-Glc, and β-Gal (Fig. 1). Among these, the GalNAc-type O-linked glycosylation pathway, found on Ser, Thr, or Tyr residues, is distinct from the others by the following characteristics: (i) initiation is controlled by an extended family of up to 20 polypeptide

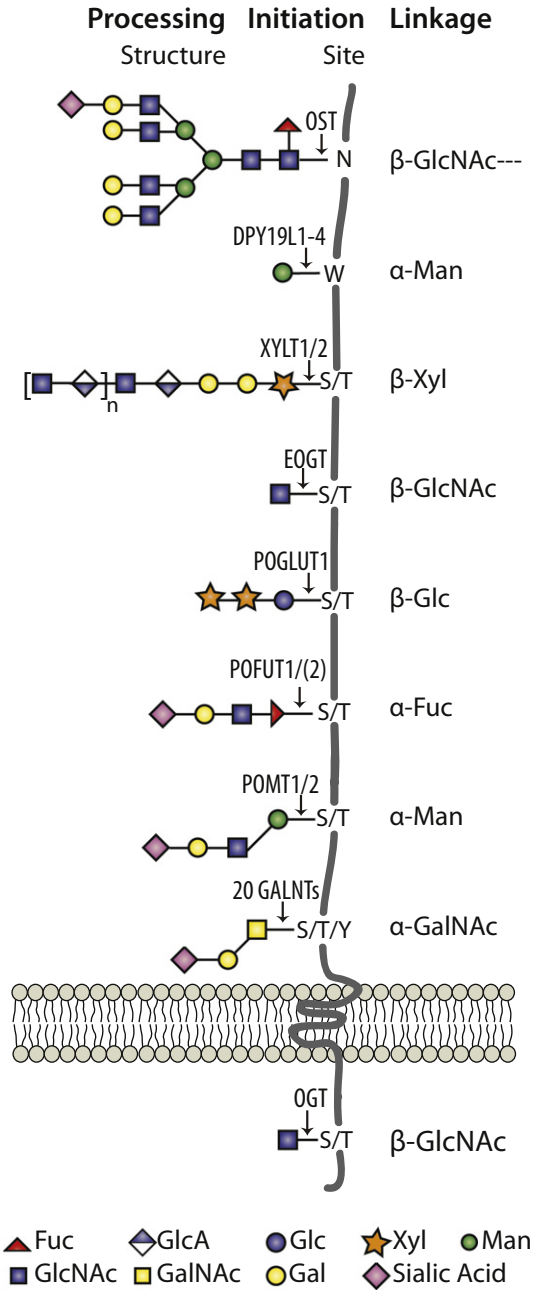

Fig. 1. Schematic depiction of mammalian protein glycosylation. Arrows indicate genes or gene complexes (OST) encoding key glycosyltransferases involved in biosynthetic initiation for each glycoconjugate. The linkage of the first mono- or oligosaccharide (—) is depicted as well as examples of processed mature structures. For the sake of simplicity all classes of glycosylation are shown on the same transmembrane protein.

GalNAc-transferases that have distinct but partly overlapping acceptor substrate specificities and tissue expression patterns — other types of O-glycosylation are controlled by only one or two isoenzymes (or, in the case of O-GlcNAc, by two enzymes fully distinct with respect to structure, compartmental localization, and function [9]); (ii) potential for generating the most differentially regulated glycoproteome; and (iii) lack of any clearly definable peptide consensus acceptor peptide sequence motif (this appears to be shared by O-Man). These factors along with the general problems; that glycosylation is a non-template driven process (not coded for directly by genes, but metabolic products generated by the orchestrated activity of hundreds of enzymes); that carbohydrate residues are not confined to strict linear sequences, but can vary with respect to linkage positions, anomeric configuration, and branching; and that glycans present uniquely difficult analytical challenges, with respect to both structures and sites of attachment on proteins; have contributed to special difficulties faced in determining specific functions for O-glycosylation [10,11].

With respect to analytical approaches relying on mass spectrometry – in other words, most methodologies relevant to current glycoproteomics, especially high throughput strategies – the structural (and therefore physico-chemical) diversity complicates all phases including enrichment strategies, liquid chromatographic separation, ion yields and fragmentation, and software processing and subsequent interpretation of LC–MS data taking into account the presence of PTMs [12,13]. Heterogeneity with respect to both site occupation and glycan structure is particularly difficult for currently available data analysis software to handle, especially in the absence of a straightforward peptide consensus site to constrain allowable glycosylation sites. Software searches based on matching peptide precursor m/z in the MS1 profile with their product ions in MS2 spectra must include distinct entries for each potential PTM carrying a unique m/z increment; as the number of distinct variable modifications increases, it eventually outstrips practical limits on computational power and processing time, and some programs (such as Proteome Discoverer) place a hard limit on this number.

Another factor contributing to ambiguity in glycan structure determinations by mass spectrometry, besides those mentioned already is the problem of isomeric residues, which are indistinguishable by m/z alone, regardless of instrumental resolution or accuracy. Thus, the m/z increments for GlcNAc and GalNAc are identical, and in the absence of other qualifying information must strictly be referred to as "HexNAc". Similarly, all "Hex" residues, e.g., Gal, Glc, and Man, are identical to the mass spectrometer in the absence of other information. This could be provided by the logic of biology and biochemistry, if the system is well-characterized, but may be impossible to provide de novo in a less characterized system. As a result, development of instrumental methodologies based on e.g. MSn fragmentation modes, is in great demand in the field of glycoproteomics.

With respect to fragmentation in MS2, a further problem is that under low energy collision induced dissociation (CID) conditions,-especially such as found in ion trap mass analyzers, sugar residue loss is the dominant mode for glycopeptides [13,14], resulting in virtually no peptide-specific fragmentation; identifying peptide sequences, let alone glycosites, is a challenging task impossible under these conditions. Under higher energy CID (HCD) conditions, there is uncontrolled and non-uniform loss of sugar residues; this is superior in the sense that with peptides carrying low-to-moderate glycosylation, energies can be readily adjusted high enough to efficiently cleave all sugar residues, with enough left over to provide good fragmentation coverage of the peptide. There are three disadvantages rendering HCD less optimal for O-glycoproteomics in particular: (i) in most cases site information is effectively lost along with the glycan; although some peptide fragments may retain the sugar component, these are generally either absent or too few to constrain one or more glycosites to one or more specific peptide residues where potential sites are present in excess (often the case); and although it is possible to optimize the yield of glycosylated

fragments by careful adjustment of the collision energy, this is practical only for single glycopeptide analytes introduced, e.g., via direct infusion, but not practical in the context of an LC–MS analysis; (ii) the HCD fragmentation of highly glycosylated peptides may again be dominated by sugar loss and provide little peptide sequence-specific fragments; and (iii) an extra provision must be made to match the *m/z* of the O-glycosylated precursor in MS1 with the resulting set of mainly unglycosylated fragment masses in MS2; this can be partially solved by multiple PTM subtraction algorithms, but this requires an extra step of calculation, and multiplies the number of data sets that have to be included in the subsequent search process; and is again ineffective for highly glycosylated peptides, or cases where high degrees of glycan complexity and heterogeneity must be accounted for. As will be outlined in the following section, an effective solution to these limitations is provided by fragmentation by electron capture dissociation (ECD), or the analogous electron transfer dissociation (ETD) mode.

Other problems include suppression of O-glycopeptide ionization by the presence in complex mixtures of unglycosylated peptides [12], which necessitates some form of O-glycopeptide enrichment prior to LC–MS analysis, which will in turn again be complicated by glycan structural diversity as already noted above. There is furthermore a lack of endo-O-glycosidase enzymes with broad specificity that can be used to release O-glycans en bloc for structural profiling. One O-glycosidase is known [15], but is effective only for a single O-glycan structure (core1, T, Galβ1-3GalNAcα1-O-S/T), and moreover inherently does not create an opportunity for concomitant chemical labeling of the former O-glycosite. Chemical methods for de-O-glycosylation (β-elimination) adapted for use with concomitant or subsequent peptide site labeling have been applied with some success. Non-reductive β-elimination employing NH_3 or NH_2-R reagents has been utilized for chemical release of O-GalNAc with concomitant aminylation of the modified residue [16,17]; the tag is relatively stable upon CID-fragmentation which allows the formerly O-glycosylated residue to be identified. An additional method, based on non-reductive β-elimination followed by Michael addition utilizing dithiothreitol (DTT) or biotin-penthylamine as nucleophiles, which confers a significant advantage by allowing affinity enrichment of tagged peptides, has been developed for high throughput identification of O-GlcNAc modifications [18]. Possible pitfalls of these methods involve peptide degradation and artifactual β-elimination from phosphorylated, and to some extent unmodified, Ser, Thr, and alkylated Cys residues. In principle, these methods can also be used for site identification in O-GalNAc modified peptides, but have not yet been applied to studies at a proteome-wide scale. In general, in the past, O-glycoprotein analysis has relied almost exclusively on strategies decoupling the problems of glycan structure determination from determination of the glycosites; once a protein has been identified, the O-glycans are chemically removed with concomitant stabilization (e.g., by β-elimination in the presence of a reducing agent), with less attention paid to site determination. Global analysis of O-glycan expression, or O-glycomics [19,20], has also been carried out using β-elimination from total cell or tissue protein, again at the expense of information about O-glycosites or even which specific proteins are O-glycosylated. As a result, we know hundreds of GalNAc-type O-glycan structures in detail (since the process can be scaled up sufficiently that other techniques such as NMR and methylation linkage analysis can be applied), but, until recently, the number of known O-glycosites/O-glycoproteins, on a proteome-wide scale, could be said to be negligible, except for a relative small number of well-studied proteins purified and examined intensively one at a time.

3. Current state-of-the-art solutions for O-glycoproteomics

3.1. Application of ETD-MS2 fragmentation

As mass spectrometry-based high throughput PTM proteomics has advanced, classical approaches to O-glycoprotein characterization have given way to bottom-up (and to some extent top-down) shotgun methods for site-specific analysis of PTM features. Many PTMs are sufficiently stable for HCD-MS2 to be an effective mode for fragmentation (e.g. phosphorylation, acetylation), and sufficiently homogeneous that enrichment strategies, although diverse, need to include only one or two straightforward steps. As emphasized in the previous section, glycans are too labile, and in most cases too diverse, for such straightforward enrichment and LC–MS solutions to be applied on a proteome-wide scale, although some relatively simple higher throughput site-specific strategies have been effectively applied to subsets of the glycoproteome, such as the case of N-glycoproteomics [21], or with methods developed recently for identifying sialylated O-glycopeptides/O-glycoproteins [22,23]. ECD- and ETD-MS2 were proposed some time ago as an alternative technique for peptide fragmentation, where peptide backbone cleavage occurs in N–C_α bond, resulting in c and z types of fragment ions [24–28]. While the fragmentation mechanism is still under debate and several hypothesis have been put forward [29], it is more or less clear that ECD and ETD induce peptide bond cleavage with minimal loss of PTM and therefore are very well-suited for analysis of phosphopeptides and glycopeptides [30]. ECD is mainly implemented into FT-ICR MS, while ETD is part of OrbiTrap FTMS (for review see [29,30]). In spite of the clear advantages of ECD and ETD they have their own limitations related to attaining the most effective charge state, which in turn limits the size and state of glycopeptides that can be sequenced, especially in a low abundance, high throughput setting. Another limitation is so-called "proline effect", where the peptide bond N-terminal to Pro residues is not fragmented due to the cyclic secondary amine structure of this amino acid [29].

Application of ETD-MS2 in O-glycoproteomics first made impact with O-GlcNAc glycoproteomes, facilitated by the fact that the bulk of this PTM (cytosolic type) comes without further extension in nature; it is inherently simple and homogeneous, which makes it amenable to O-glycopeptide enrichment based on affinity to a single lectin (WGA) [31], and glycopeptide sequencing with O-glycosite identification based on ETD-MS2 [32]. It should be noted that more recent studies have shown that the cytosolic O-GlcNAc residues can in fact be modified by 6-phosphate, although the extent to which this occur is still largely unknown [33,34].

Early advances in high throughput site-specific O-GalNAc proteomics (O-glycoproteome DB, http://www.glycoproteomics.somee.com) were also made using a combination of lectin affinity and complementary HCD and ETD-MS2 modes of fragmentation [35]; it was noted, however, that ETD-MS2 has preferences for low m/z precursors [36] and therefore the effectiveness could still be limited by the size of the O-glycan and/or the number of O-glycosites on a given peptide. This could be ameliorated by application of glycosidase trimming, artificially reducing the size and complexity of the O-glycans, thereby essentially recapitulating the favorable situation with O-GlcNAc. This was one of the main lessons we incorporated into developing the so-called SimpleCell strategy as it applied to mass spectrometry-based high throughput O-glycoproteomics (Fig. 2) [37].

3.2. Enrichment methods

Four basic classes of enrichment protocols have been used for glycoproteins and proteolytically generated glycopeptides: (i) methods based on chemical or enzymatic oxidation of glycans; (ii) methods based on click chemistry requiring metabolic labeling with alkyne- or azide-modified sugars; (iii) methods based on charge affinity; and (iv) methods based on lectin affinity.

(i) Periodate has been known for decades as the reagent of choice for oxidation of glycans [38]; periodate has been incorporated into glycoproteomic strategies utilizing it as a general carbohydrate oxidant. In this case the oxidized carbohydrate can be -captured on an immobilizing solid matrix, e.g., hydrazide beads, and

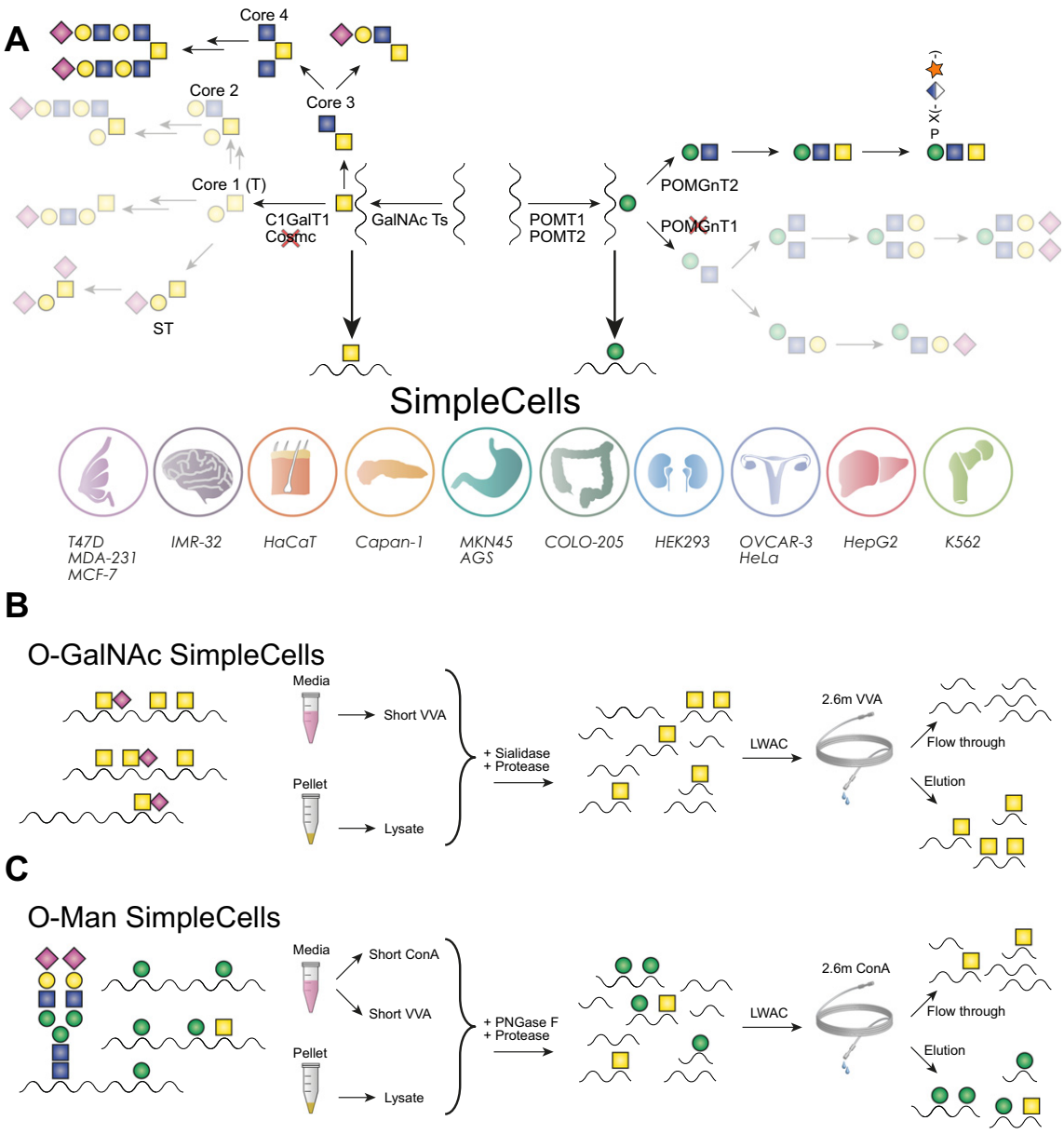

Fig. 2. Schematic depiction of SimpleCell strategy. (A) COSMC and POMGNT1 knockout resulting in simplified O-GalNAc (in some cells NeuAcα2-6GalNAc, also known as STn) and O-Man glycosylations. (B) Enrichment strategy of O-GalNAc glycopeptides by *Vicia villosa* agglutinin (VVA). Cell media of O-GalNAc SimpleCells are pre-enriched on a short VVA column. O-GalNAc proteins from media and pellet are protease digested, sialidase treated (for STn positive cells) and enriched using a 2.6 m VVA lectin column. (C) Enrichment strategy of O-Man glycopeptides by *Concanavalin A* (ConA). Cell media of O-Man SimpleCells are pre-enriched on a short VVA or ConA column. O-Man proteins from media and pellet are protease digested, PNGase F treated and enriched using a 2.6 m ConA lectin column.

later eluted by acid treatment or in the case of N-glycans by enzymatic release with PNGase F [22,39]. For N-glycans the latter allows concomitant labeling of the former N-glycosites when performed in $H_2^{18}O$ [21,40], but recent studies have also identified $H_2^{18}O$ incorporation as a consequence of natural deamidation which may result in misinterpreting N-glycosite occupancies [41–43]. The Larsen group treated *Escherichia coli* extracts with PNGase F in $H_2^{18}O$ and found canonical N-glycosylation sites where Asparagine was converted to ^{18}O-Aspartic acid, clearly demonstrating the natural deamidation process since *E. coli* does not have an N-glycosylation machinery [41].

For the hydrazide capture procedure, washing off unbound proteins followed by on-bead proteolytic digestion allows for independent identification of the captured glycoproteins via the proteolytically released fraction and in the case of N-glycoproteins also subsequent isolation of their formerly N-glycosylated peptides by enzymatic PNGase F release. Of note, the PNGase F enzyme is inactive on N-linked glycans harboring a α1-3 linked fucose at the chitobiose core [44], a modification typically found on N-glycans from plants and invertebrates, and is thus not suited for general N-glycosite studies in these organisms.

Under mild conditions periodate can be applied for specific oxidation of sialic acids, thus establishing a method for characterizing the "sialome", including both N- and O-glycoproteins. Mild periodate oxidation introduces aldehydes specifically on sialic acids, and, importantly, does not oxidize the underlying glycans, thereby allowing N- and O-glycoproteins, with intact glycans, to be immobilized on a solid phase. The terminal sialic acids are labile to mild acid treatment [45], relative to most other glycosidically linked residues, allowing release of immobilized glycopeptides from, e.g., hydrazide beads,

with retention of the N- or O-glycan core. The obvious disadvantages are that all information about sialic acids is lost, and that unsialylated fractions of the glycoproteome remain uncaptured.

Galactose Oxidase (GO) may be used for oxidation of terminal Gal or GalNAc residues of N- and O-linked glycans [46,47]. The enzymatic reaction converts the C-6 primary alcohol into an aldehyde, thus making the oxidized Gal or GalNAc residue amenable to conjugation with compounds containing e.g. hydrazide or aminooxy functionalities. Although the concept of utilizing GO oxidized Gal or GalNAc for enrichment of N-linked glycoproteins/glycopeptides by biotin-tagging or through solid-phase chemistry has been demonstrated [48,49], studies aimed at analyzing the unsialylated O-glycoproteome through the GO-based methodology have not been reported yet.

(ii) Metabolic labeling with GalNAz and GlcNAz has been used for characterization of O-GalNAc and O-GlcNAc proteomes, respectively. The Bertozzi group originally developed a metabolic labeling strategy for O-GalNAc glycosylation using the bioorthogonal chemical reporter peracetylated N-azidoacetylgalactosamine (Ac4GalNAz), an azido analog of GalNAc azide [50,51]. GalNAz was shown to be metabolized by cultured cells and incorporated into the core position of O-linked glycans [50] and later it was shown that several GalNAc-Ts functioned with UDP-GalNAz in vitro [52]. The azide has unique chemical reactivity with phosphine probes in the Staudinger ligation reaction. O-glycoproteins modified with GalNAz can therefore be tagged for enrichment for glycoproteome analysis, but care should be taken by the finding that UDP-GalNAz in cells are epimerized to UDP-GlcNAz by the C4-Glc/GlcNAc epimerase and hence labeling of O-GlcNAc glycosylation will also occur [53]. This has been exploited to probe cytosolic O-GlcNAc glycoproteins [53], although broader glycoproteomics studies have not been reported yet. Instead the O-GlcNAc glycoproteome has been probed by a related strategy using enzymatic in vitro incorporation of GalNAz into O-GlcNAc glycosites by a mutant β4galactosyltransferase (GalT1Y289L) and subsequent tagging with biotin using Biotin-PEG-PC-Alkyne. This enables capture of O-GlcNAc glycopeptides with avidin and using this strategy Alfaro et al. [54] identified the largest O-GlcNAc glycopeptide data set from mouse brain so far using a combination of CID, HCD and ETD.

(iii) Titanium dioxide (TiO_2) beads, which are effective for enrichment of peptides carrying acidic PTMs, such as phosphoryl, have also proved useful for enrichment of sialylated glycopeptides [55]. The mechanism of sialic acid binding has been suggested to resemble a multidentate interaction, involving carboxyl- and hydroxyl groups of sialic acid in engagement with the TiO_2 surface. Typically, the methodology is utilized to enrich sialylated species and identify formerly N-glycosylated peptides following PNGase F digestion [56,57]. Theoretically, this strategy may also be used to capture sialylated O-glycopeptides but TiO_2 based O-glycoproteomic studies have not been reported yet. As with mild periodate oxidation, the main weakness is failure to capture unsialylated fractions of the glycoproteome.

(iv) Lectin weak affinity chromatography (LWAC) was originally used effectively for enrichment of O-GlcNAc and O-GalNAc glycoproteomes using WGA and Jacalin, respectively [31,32,35,58,59]. This was a second lesson incorporated into the workflow for enrichment of SimpleCell O-glycoproteomes [37]; engineering cells to produce homogeneous O-glycosylation with only the innermost GalNAc residues would be efficiently enriched by LWAC using immobilized *Vicia villosa* agglutinin (VVA). We have compared the efficiency of PNA enrichment with wild type cells that produce homogeneous core1 based O-glycoproteomes such as K562 [37] and more recently Chinese hamster ovary (CHO) cells [60], and in general it appears that VVA enrichment is more sensitive and provides deeper O-glycoproteomes. The SimpleCell strategy was also recently expanded to O-mannosylation demonstrating that enrichment of O-Man SimpleCell O-glycoproteomes can be carried out by LWAC with immobilized Concanavalin A (ConA) [61]. LWAC enrichment can be carried out at both the protein and peptide level [37,62]. Adding enrichment at the protein level was found to be both necessary and sufficient for analysis of secretome fractions of SimpleCell O-glycoproteomes [62,63].

4. Genetic engineering of cell lines for glycan structure simplification — the SimpleCell strategy

A straightforward solution to the multiple analytical challenges associated with glycan size, complexity, and physico-chemical diversity is to reduce or eliminate all of these factors through structural simplification to small, homogeneous, residual units, e.g., monosaccharides. While this carries the obvious disadvantage of eliminating site-specific information about complex glycosylation, it greatly facilitates all other aspects of bottom-up protein glycosite analysis, including enrichment, LC–MS/MS protocols, and data processing and interpretation. This is clearly the way to proceed if the goal is to widely (deeply) expand the protein "glycosylation space", that is, to find out – and find out quickly – where saccharides are located on a proteome-wide basis. Historically, in the case of GlcNAc O-glycoproteomics, the inherent lack of complexity led immediately to accelerated expansion of the GlcNAc O-glycoproteome, lending credence to the suggestion that trimming GalNAc-type O-glycans to a single residue would achieve similar ends [32,35]. Early approaches to GalNAc O-glycosite simplification through glycosidase or chemical treatments were indeed successful [35], along with improved enrichment strategies and the use of ETD-MS2 fragmentation, in expanding the GalNAc O-glycoproteome. However, a true breakthrough, taking O-glycoproteomics beyond incremental expansions, occurred with the introduction of SimpleCells [37], in which gene editing was used to generate cell lines essentially lacking the machinery for extending mucin-type O-glycosylation beyond the initiating O-GalNAc residue, except for sialylation (easily removed, where it occurs, by neuraminidase treatment prior to enrichment).

The SimpleCell strategy relies on precise nuclease mediated gene editing to knockout glycosyltransferase genes responsible for glycan elongation such that glycosylation is simplified and present in a more homogeneous form that enable efficient lectin capture (for a review see [64]). The O-GalNAc SimpleCell strategy relies on knockout of *COSMC*, an obligate chaperone of C1GalT1, the β3Gal-transferase responsible for Core 1 (Galβ1-3GalNAcα1-O-Ser/Thr) elongation of O-linked GalNAc (Fig. 2A). Originally Zinc finger nuclease (ZFN) mediated gene editing was used and later both TALENs and CRISPR/Cas9 mediated editing have become available [7,37,64]. In the absence of functional COSMC protein, only truncated O-linked Tn (GalNAcα1-O-Ser/Thr) and/or STn (NeuAcα2-6GalNAcα1-O-Ser/Thr) structures are formed in most cell types (a common alternative elongation pathway, core3, is restricted to gastrointestinal tissues and largely absent in cancer cell lines) [65]. SimpleCells may express Tn, a mixture of Tn and STn, or exclusively of STn, which appear to depend at least partly on expression of the ST6GalNAc-I sialyltransferase. After neuraminidase treatment if required (depending on STn expression), protease-generated GalNAc-glycopeptides are efficiently enriched using *V. villosa* agglutinin (VVA) LWAC. Secretome analysis was facilitated by an additional short capture VVA chromatography step at the protein level prior to proteolysis to isolate endogenously produced O-glycoproteins from media components [62,63]. Additional reduction in sample complexity was achieved by use of isoelectric focusing (IEF) prior to nLC–MS (for current workflow see Fig. 3). Further depth of coverage was shown to be increased substantially by the use proteases other than trypsin, e.g., chymotrypsin [63]. Using a combination of high resolution

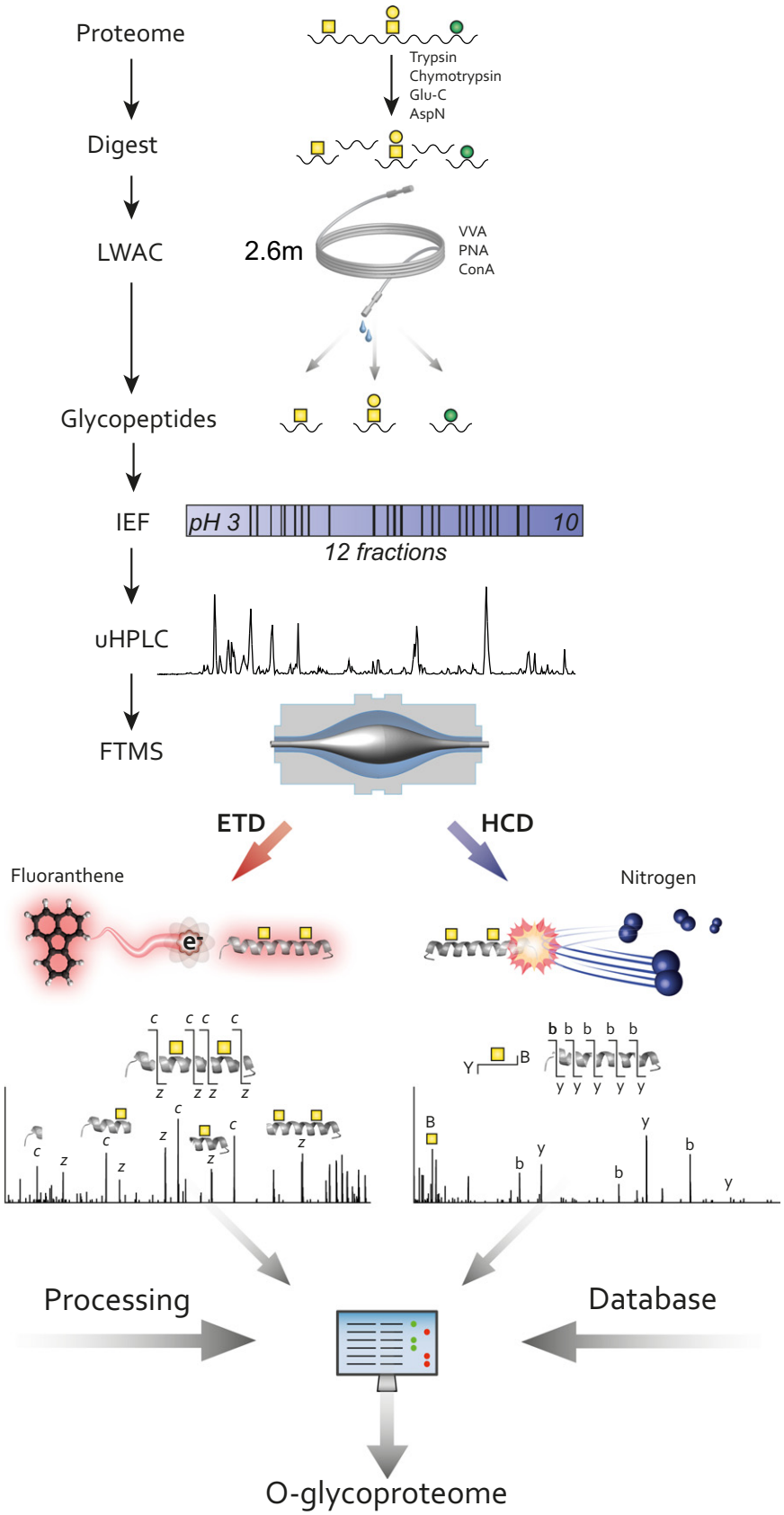

Fig. 3. Outline of glycoproteomics workflow. Proteolytic digest initially treated by neuraminidase was submitted for O-glycopeptide enrichment by LWAC on a long column of immobilized VVA, PNA or ConA. The most abundant for O-glycopeptides LWAC fractions were further fractionated by isoelectric focusing (IEF). 12 IEF fractions were then submitted to nLC–MS (OrbiTrap) followed by repetitive top 5 HCD MS2 and top 5 ETD MS2 analysis.

(Orbitrap-FT) MS1 with data-dependent acquisition (DDA) of HCD-MS2 and ETD-MS2 spectra, both at relatively high (Orbitrap-FT) spectral resolution, we have to date reported more than 3000 O-glycosites distributed over 600 O-glycoproteins from 12 cancer cell lines, providing a "first global view" of the human GalNAc O-glycoproteome representing an almost 10-fold expansion over that previously known [66]. The data set is considered highly reliable, since spectra associated with each site were reviewed by inspection (notably, a procedure that we have now abandoned, due to the high volume of new data being produced). It has to be mentioned that all of these results were produced using an early generation Orbitrap XL instrument, and with faster and more sensitive instruments, such as the Velos Pro and Fusion models an enormous increase in throughput and depth is achievable (unpublished).

5. O-GalNAc glycosylation of Tyr

Along with expansion of the O-glycoproteome an interesting new development was the discovery of GalNAc attachment to tyrosine residues. It was first reported as a modification of the amyloid precursor protein (APP) from cerebrospinal fluid [67], and immediately thereafter identified in NUCB2 from colon cancer (COLO205) SimpleCells and, most likely, in WT Jurkat cells, which are a spontaneous *cosmc* knockout cell line [37]. The APP modification was found on Tyr-681, within a well-studied region close to the α-secretase cleavage site [67]. As often happens, once a new structural feature has been identified, more instances can be found by specifically searching for it; thus, 5 additional GalNAc-O-Tyr sites were subsequently confirmed [63], and the total number identified from SimpleCells was recently expanded to 23, distributed among 19 proteins from 11 of the 12 cell lines studied [66]. We have currently identified a total of more than 50 Tyr glycosites (unpublished). Although the mass increment is the same for O-GlcNAc, the possibility that these could be other than O-GalNAc sites is reduced by the finding of a complex O-glycan attached to Tyr-681 in APP, along with the same arguments as those generally validating the SimpleCell results: by the enrichment with highly specific VVA, and their observation in the extracellular domains of membrane-associated and/or secreted proteins processed through the ER-Golgi pathway. Recently, 3 more potential GalNAc-O-Tyr sites were observed in the glycoproteome of the murine synaptosome [68], although the authors were more reluctant to identify the HexNAc residue unambiguously. In this case, similar to the SimpleCell findings, only single HexNAc modifications were observed, and WGA, a less specific lectin, was used for enrichment, justifying some hesitancy; on the other hand, if the sugar were instead identified as O-GlcNAc, this would already constitute an additional novelty.

6. Dissection of the function of polypeptide GalNAc-transferases using isogenic SimpleCells

The O-GalNAc glycoproteome is orchestrated by the large polypeptide GalNAc-transferase (GalNAc-T) enzyme family with up to 20 members and the role of each isoform is poorly understood [10] (Fig. 1). Most of the current knowledge of these isoenzymes is based on in vitro enzyme assays with short peptide substrates [10,69,70]. Several knockout animal models with deficiency in GalNAc-T isoforms have been made [69,71–73], but these have not yet provided broad insight into the specific contribution of individual isoforms to the O-glycoproteomes, which is likely due to the complexity of performing O-glycoproteomics studies with heterogeneous glycoproteomes. Genetically engineered isogenic cell systems based on SimpleCells with and without individual GalNAc-T isoforms are now opening for discovery of non-redundant functions of this family of isoenzymes [62]. In a first preliminary study of HepG2 liver cells, comparing sets of identified GalNAc-glycopeptides in isogenic cells with and without GALNT2 led to discovery of unique functions of the GalNAc-T2 isoform [62]. It was demonstrated that e.g. the major lipoprotein, ApoC-III, and the lipase inhibitor ANGPTL3 are unique substrates for GalNAc-T2, and thus expected not to be O-glycoproteins in individuals without a functional *GALNT2* gene [62]. This is important since there is strong evidence that *GALNT2* plays a role in dyslipidemia [74], and in preliminary studies we have confirmed that APOC-III serves as a biomarker for deficiency in *GALNT2* (unpublished). We previously demonstrated that another isoform, GalNAc-T3, serves an essential non-redundant role in glycosylation of FGF23 and deficiency in *GALNT3* cause familial tumoral calcinosis due to loss of function of FGF23 [3]. Moreover, recently a unique function of the *GALNT11* gene in regulation of Notch signaling and heart heterotaxy was reported [75]. The access to isogenic cell systems with and without GalNAc-T isoforms amenable for high throughput O-glycoproteomics is likely going to dramatically change the perception of this type of O-glycosylation and serve as a discovery platform for new biological functions of site-specific glycosylation.

7. The human O-Man glycoproteome

We originally proposed that the SimpleCell strategy could be adapted for O-glycoproteomes other than O-GalNAc [37]. O-mannosylation (O-Man) is one type of protein O-glycosylation that for long remained elusive in mammals [76]. For years, α-dystroglycan (αDG) remained the only well-characterized glycoprotein with respect to O-Man glycans and their attachment sites. What started as the discovery of mannitol containing sugars released by reductive β-elimination from rat brain proteoglycans [77], later also shown to be a modification of αDG from bovine peripheral nerve [78], culminated recently in a series of reports describing the site-specific location of O-Man glycans on αDG [79–82]. Additional O-Man modified proteins were identified during the past ten years, including light chain IgG2 [83], receptor-type tyrosine-protein phosphatase β (RPTPβ) [84], CD24 [85], neurofascin 186 [86] and the lecticans [87]. Applying the SimpleCell strategy to O-Man glycosylation (Fig. 2B) has now brought a marked expansion to the human O-Man glycoproteome. The major O-Man elongation pathway is dependent on the POMGnT1 enzyme, which adds a GlcNAc residue linked β-1,2 to the initially transferred α-Man residue. Knockout of the POMGNT1 gene would therefore result in truncation in the majority of O-mannosylated proteins to single O-Man residues. Another elongation pathway by POMGnT2 (GTDC2) has so far only been found on specific sites in dystroglycan [88,89]. Elongation of O-Man glycans by the POMGnT2 pathway gives rise to a specific glycan epitope (GalNAcβ3GlcNAcβ4Man-*O*-Thr) which, following phosphorylation by SGK196 on the 6-position of O-Man [88] and further elongation by repeating (-3Xyl-α1,3GlcAβ1-) disaccharide units, ultimately results in a glycan polymer that is necessary for laminin binding [90,91]. Disruption of O-Man biosynthetic pathways and defective αDG glycosylation result in a broad spectrum of muscular dystrophies [92].

The O-Man SimpleCell O-glycoproteomics strategy was recently brought to realization in a breast cancer cell line, requiring only minimal changes to the analytical workflow established for GalNAc O-glycoproteomics (substitution of ConA for VVA in the LWAC step) (Fig. 2B) [61]. The strategy enabled identification of the large family of cadherins and protocadherins as major carriers of O-Man glycosylation [61]. Plexins, the macrophage-stimulating protein receptor (MST1R) and the hepatocyte growth factor receptor (HGFR), two proteins that share domain folds with the plexin superfamily, together members of the protein disulfide-isomerase (PDI) family, were some of the additional O-Man glycoproteins identified through the SimpleCell approach; worth mentioning is also KIAA1549, an uncharacterized transmembrane protein, containing the hitherto largest number of O-Man modifications for a single protein (64 glycosites in total).

It is reasonable to believe that the 62 proteins identified to date represent the tip of the iceberg for the elusive O-Man glycoproteome;

further expansion could thus provide valuable insight into the function of the O-Man modification. This task is currently undertaken in our lab through the continuation of our O-Man SimpleCell program not only by employing cell lines from other tissues and organisms but also by individually targeting the POMT1 and POMT2 enzymes with ambitions to generate a differential map over their preferred protein substrates.

8. Perspectives

The challenge of high throughput proteome-wide identification of O-glycosites has been partly met by recent advances in tagging strategies, genetic engineering of cells, and ETD based mass spectrometry. Wider application of these strategies should greatly expand our knowledge of the different O-glycoproteomes and open for discovery of biological functions in health and disease. The O-GlcNAc, O-GalNAc, and O-Man glycoproteomes are open for discoveries using the outlined strategies and more O-glycoproteomes should be addressable by the SimpleCell strategy such as the O-Xyl proteoglycan Ome (Fig. 1). However, currently these strategies largely ignore glycan structures at specific glycosites. These obvious limitations are to some extent being addressed by alternative approaches targeting not the enrichment methods or MS hardware itself but rather the bioinformatic branch of glycoproteomics. Among many innovative ideas, one promising application includes the subtraction routine employed by several groups to reduce the computational load associated with multiple variables (glycoforms) [22,63,93]. In its most simple form, the glycoform mass is initially subtracted from the precursor ion mass, allowing proteome-wide database searches without multiple variable modifications, and subsequently paired back to same precursor ion to couple glycan structure to peptide backbone. Further bioinformatic developments in these directions could thus bring us one step closer to global glycoproteomics addressing also native O-glycans.

Additional technological and methodological developments certainly call for activity, which, when combined with challenging applications, may promote further advancement of glycoproteomics and pave the way to exciting discoveries. For example, we look forward to developments in ion-mobility utilities which certainly will allow us to address additional structural details (anomeric and epimeric configurations) that are currently out of reach for conventional glycoproteomics.

The vision must be to be able to address the more difficult problem of simultaneous identification of O-glycan sites and O-glycan structures on a proteome-wide level. It is our hypothesis that the combination of knowledge of O-glycosites with recent improvements in sample separation assisted by ultra-high pressure LC (uHPLC) and hydrophilic interaction liquid chromatography (HILIC), along with an entirely new configuration of the Orbitrap mass spectrometer ion optics, including a breakthrough redesign of the ETD module (Fusion), will enable the field to break the barrier and develop a method for analysis of intact O-glycoproteomes. A successful analytical strategy for analysis of native O-glycoproteins in complex biological samples will open the field widely for functional studies of O-glycoproteins.

Acknowledgments

This work was supported by The Novo Nordisk Foundation, The Danish Research Councils, a programme of excellence from the University of Copenhagen, The Mizutani Foundation, and The Danish National Research Foundation (DNRF107).

References

[1] P. Minguez, L. Parca, F. Diella, D.R. Mende, R. Kumar, M. Helmer-Citterich, A.C. Gavin, V. van Noort, P. Bork, Deciphering a global network of functionally associated post-translational modifications, Mol. Syst. Biol. 8 (2012) 599.
[2] J.B. Lowe, J.D. Marth, A genetic approach to mammalian glycan function, Annu. Rev. Biochem. 72 (2003) 643–691.
[3] K. Kato, C. Jeanneau, M.A. Tarp, A. Benet-Pages, B. Lorenz-Depiereux, E.P. Bennett, U. Mandel, T.M. Strom, H. Clausen, Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation, J. Biol. Chem. 281 (2006) 18370–18377.
[4] K.T. Schjoldager, H. Clausen, Site-specific protein O-glycosylation modulates proprotein processing — deciphering specific functions of the large polypeptide GalNAc-transferase gene family, Biochim. Biophys. Acta 1820 (2012) 2079–2094.
[5] K.T. Schjoldager, M.B. Vester-Christensen, E.P. Bennett, S.B. Levery, T. Schwientek, W. Yin, O. Blixt, H. Clausen, O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids, J. Biol. Chem. 285 (2010) 36293–36303.
[6] K.T. Schjoldager, M.B. Vester-Christensen, C.K. Goth, T.N. Petersen, S. Brunak, E.P. Bennett, S.B. Levery, H. Clausen, A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing, J. Biol. Chem. 286 (2011) 40122–40132.
[7] C. Steentoft, E.P. Bennett, K.T. Schjoldager, S.Y. Vakhrushev, H.H. Wandall, H. Clausen, Precision genome editing: a small revolution for glycobiology, Glycobiology 24 (2014) 663–680.
[8] R. Kornfeld, S. Kornfeld, Assembly of asparagine-linked oligosaccharides, Annu. Rev. Biochem. 54 (1985) 631–664.
[9] G.W. Hart, M.P. Housley, C. Slawson, Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins, Nature 446 (2007) 1017–1022.
[10] E.P. Bennett, U. Mandel, H. Clausen, T.A. Gerken, T.A. Fritz, L.A. Tabak, Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family, Glycobiology 22 (2012) 736–756.
[11] D.J. Gill, H. Clausen, F. Bard, Location, location, location: new insights into O-GalNAc protein glycosylation, Trends Cell Biol. 21 (2011) 149–158.
[12] P.H. Jensen, D. Kolarich, N.H. Packer, Mucin-type O-glycosylation—putting the pieces together, FEBS J. 277 (2010) 81–94.
[13] 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-glycoproteome, Biochim. Biophys. Acta 1844 (2014) 1437–1452.
[14] M. Wuhrer, M.I. Catalina, A.M. Deelder, C.H. Hokke, Glycoproteomics based on tandem mass spectrometry of glycopeptides, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 849 (2007) 115–128.
[15] R.A. O'Neill, Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis, J. Chromatogr. A 720 (1996) 201–215.
[16] F.G. Hanisch, M. Jovanovic, J. Peter-Katalinic, Glycoprotein identification and localization of O-glycosylation sites by mass spectrometric analysis of deglycosylated/alkylaminylated peptide fragments, Anal. Biochem. 290 (2001) 47–59.
[17] G.J. Rademaker, S.A. Pergantis, L. Blok-Tip, J.J. Langridge, A. Kleen, J.E. Thomas-Oates, Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity, Anal. Biochem. 257 (1998) 149–160.
[18] L. Wells, K. Vosseller, R.N. Cole, J.M. Cronshaw, M.J. Matunis, G.W. Hart, Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications, Mol. Cell. Proteomics 1 (2002) 791–804.
[19] G. Zauner, R.P. Kozak, R.A. Gardner, D.L. Fernandes, A.M. Deelder, M. Wuhrer, Protein O-glycosylation analysis, Biol. Chem. 393 (2012) 687–708.
[20] S.M. Haslam, S.J. North, A. Dell, Mass spectrometric analysis of N- and O-glycosylation of tissues and cells, Curr. Opin. Struct. Biol. 16 (2006) 584–591.
[21] D.F. Zielinska, F. Gnad, J.R. Wisniewski, M. Mann, Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints, Cell 141 (2010) 897–907.
[22] J. Nilsson, U. Ruetschi, A. Halim, C. Hesse, E. Carlsohn, G. Brinkmalm, G. Larson, Enrichment of glycopeptides for glycan structure and attachment site identification, Nat. Methods 6 (2009) 809–811.
[23] A. Halim, J. Nilsson, U. Ruetschi, C. Hesse, G. Larson, Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD, Mol. Cell. Proteomics 11 (2012) M111 (013649).
[24] N.L. Kelleher, R.A. Zubarev, K. Bush, B. Furie, B.C. Furie, F.W. McLafferty, C.T. Walsh, Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid, Anal. Chem. 71 (1999) 4250–4253.
[25] E. Mirgorodskaya, P. Roepstorff, R.A. Zubarev, Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer, Anal. Chem. 71 (1999) 4431–4436.
[26] J.E. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9528–9533.
[27] J. Wiesner, T. Premsler, A. Sickmann, Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications, Proteomics 8 (2008) 4466–4483.
[28] L.M. Mikesh, B. Ueberheide, A. Chi, J.J. Coon, J.E. Syka, J. Shabanowitz, D.F. Hunt, The utility of ETD mass spectrometry in proteomic analysis, Biochim. Biophys. Acta 1764 (2006) 1811–1822.
[29] K.O. Zhurov, L. Fornelli, M.D. Wodrich, U.A. Laskay, Y.O. Tsybin, Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis, Chem. Soc. Rev. 42 (2013) 5014–5030.
[30] Y. Mechref, Use of CID/ETD mass spectrometry to analyze glycopeptides, Curr. Protoc. Protein Sci. 68 (2012) 12.11.1–12.11.11.
[31] K. Vosseller, J.C. Trinidad, R.J. Chalkley, C.G. Specht, A. Thalhammer, A.J. Lynn, J.O. Snedecor, S. Guan, K.F. Medzihradszky, D.A. Maltby, R. Schoepfer, A.L. Burlingame,

O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry, Mol. Cell. Proteomics 5 (2006) 923–934.
[32] R.J. Chalkley, A. Thalhammer, R. Schoepfer, A.L. Burlingame, Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 8894–8899.
[33] M.E. Graham, M. Thaysen-Andersen, N. Bache, G.E. Craft, M.R. Larsen, N.H. Packer, P.J. Robinson, A novel post-translational modification in nerve terminals: O-linked N-acetylglucosamine phosphorylation, J. Proteome Res. 10 (2011) 2725–2733.
[34] H. Hahne, B. Kuster, Discovery of O-GlcNAc-6-phosphate modified proteins in large-scale phosphoproteomics data, Mol. Cell. Proteomics 11 (2012) 1063–1069.
[35] Z. Darula, K.F. Medzihradszky, Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum, Mol. Cell. Proteomics 8 (2009) 2515–2526.
[36] D.L. Swaney, G.C. McAlister, J.J. Coon, Decision tree-driven tandem mass spectrometry for shotgun proteomics, Nat. Methods 5 (2008) 959–964.
[37] C. Steentoft, S.Y. Vakhrushev, M.B. Vester-Christensen, K.T. Schjoldager, Y. Kong, E.P. Bennett, U. Mandel, H. Wandall, S.B. Levery, H. Clausen, Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines, Nat. Methods 8 (2011) 977–982.
[38] L. Van Lenten, G. Ashwell, Studies on the chemical and enzymatic modification of glycoproteins. A general method for the tritiation of sialic acid-containing glycoproteins, J. Biol. Chem. 246 (1971) 1889–1894.
[39] H. Zhang, X.J. Li, D.B. Martin, R. Aebersold, Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry, Nat. Biotechnol. 21 (2003) 660–666.
[40] J. Gonzalez, T. Takao, H. Hori, V. Besada, R. Rodriguez, G. Padron, Y. Shimonishi, A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: identification of the positions of carbohydrate-linked asparagine in recombinant alpha-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water, Anal. Biochem. 205 (1992) 151–158.
[41] G. Palmisano, M.N. Melo-Braga, K. Engholm-Keller, B.L. Parker, M.R. Larsen, Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses, J. Proteome Res. 11 (2012) 1949–1957.
[42] P.M. Angel, J.M. Lim, L. Wells, C. Bergmann, R. Orlando, A potential pitfall in 18O-based N-linked glycosylation site mapping, Rapid Commun. Mass Spectrom. 21 (2007) 674–682.
[43] Y. Du, F. Wang, K. May, W. Xu, H. Liu, Determination of deamidation artifacts introduced by sample preparation using 18O-labeling and tandem mass spectrometry analysis, Anal. Chem. 84 (2012) 6355–6360.
[44] V. Tretter, F. Altmann, L. Marz, Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1–3 to the asparagine-linked N-acetylglucosamine residue, Eur. J. Biochem. 199 (1991) 647–652.
[45] M.E. Rafelson Jr., H. Clauser, J. Legault-Demare, Removal of sialic acid from serum gonadotropin by acidic and enzymic hydrolysis, Biochim. Biophys. Acta 47 (1961) 406–407.
[46] J.A. Cooper, W. Smith, M. Bacila, H. Medina, Galactose oxidase from *Polyporus circinatus*, Fr, J. Biol. Chem. 234 (1959) 445–448.
[47] C.G. Gahmberg, S.I. Hakomori, External labeling of cell surface galactose and galactosamine in glycolipid and glycoprotein of human erythrocytes, J. Biol. Chem. 248 (1973) 4311–4317.
[48] T.N. Ramya, E. Weerapana, B.F. Cravatt, J.C. Paulson, Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans, Glycobiology 23 (2013) 211–221.
[49] H. Shimaoka, H. Kuramoto, J. Furukawa, Y. Miura, M. Kurogochi, Y. Kita, H. Hinou, Y. Shinohara, S. Nishimura, One-pot solid-phase glycoblotting and probing by transoximization for high-throughput glycomics and glycoproteomics, Chemistry 13 (2007) 1664–1673.
[50] H.C. Hang, C. Yu, D.L. Kato, C.R. Bertozzi, A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 14846–14851.
[51] J.A. Prescher, C.R. Bertozzi, Chemistry in living systems, Nat. Chem. Biol. 1 (2005) 13–21.
[52] H.C. Hang, C. Yu, M.R. Pratt, C.R. Bertozzi, Probing glycosyltransferase activities with the Staudinger ligation, J. Am. Chem. Soc. 126 (2004) 6–7.
[53] M. Boyce, I.S. Carrico, A.S. Ganguli, S.H. Yu, M.J. Hangauer, S.C. Hubbard, J.J. Kohler, C.R. Bertozzi, Metabolic cross-talk allows labeling of O-linked beta-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 3141–3146.
[54] J.F. Alfaro, C.X. Gong, M.E. Monroe, J.T. Aldrich, T.R. Clauss, S.O. Purvine, Z. Wang, D.G. Camp II, J. Shabanowitz, P. Stanley, G.W. Hart, D.F. Hunt, F. Yang, R.D. Smith, Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 7280–7285.
[55] M.R. Larsen, S.S. Jensen, L.A. Jakobsen, N.H. Heegaard, Exploring the sialiome using titanium dioxide chromatography and mass spectrometry, Mol. Cell. Proteomics 6 (2007) 1778–1787.
[56] G. Palmisano, S.E. Lendal, K. Engholm-Keller, R. Leth-Larsen, B.L. Parker, M.R. Larsen, Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry, Nat. Protoc. 5 (2010) 1974–1982.
[57] G. Palmisano, B.L. Parker, K. Engholm-Keller, S.E. Lendal, K. Kulej, M. Schulz, V. Schwammle, M.E. Graham, H. Saxtorph, S.J. Cordwell, M.R. Larsen, A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development, Mol. Cell. Proteomics 11 (2012) 1191–1202.
[58] Z. Darula, J. Sherman, K.F. Medzihradszky, How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides, Mol. Cell. Proteomics 11 (2012) O111 (016774).
[59] N.E. Zachara, K. Vosseller, G.W. Hart, Detection and analysis of proteins modified by O-linked N-acetylglucosamine, Curr. Protoc. Protein Sci. (2011) (95: II: 17.6: 17.6.1–17.6.33).
[60] Z. Yang, A. Halim, Y. Narimatsu, H.J. Joshi, C. Steentoft, K.T. Schjoldager, M.A. Schulz, N.R. Sealover, K.J. Kayser, E.P. Bennett, S.B. Levery, S.Y. Vakhrushev, H. Clausen, The GalNAc-type O-glycoproteome of CHO cells characterized by the SimpleCell strategy, Mol. Cell. Proteomics (2014) (pii: mcp.M114.041541).
[61] M.B. Vester-Christensen, A. Halim, H.J. Joshi, C. Steentoft, E.P. Bennett, S.B. Levery, S.Y. Vakhrushev, H. Clausen, Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 21018–21023.
[62] K.T. Schjoldager, S.Y. Vakhrushev, Y. Kong, C. Steentoft, A.S. Nudelman, N.B. Pedersen, H.H. Wandall, U. Mandel, E.P. Bennett, S.B. Levery, H. Clausen, Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 9893–9898.
[63] S.Y. Vakhrushev, C. Steentoft, M.B. Vester-Christensen, E.P. Bennett, H. Clausen, S.B. Levery, Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells, Mol. Cell. Proteomics 12 (2013) 932–944.
[64] C. Steentoft, E.P. Bennett, H. Clausen, Glycoengineering of human cell lines using zinc finger nuclease gene targeting: SimpleCells with homogeneous GalNAc O-glycosylation allow isolation of the O-glycoproteome by one-step lectin affinity chromatography, Methods Mol. Biol. 1022 (2013) 387–402.
[65] T. Iwai, T. Kudo, R. Kawamoto, T. Kubota, A. Togayachi, T. Hiruma, T. Okada, T. Kawamoto, K. Morozumi, H. Narimatsu, Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 4572–4577.
[66] C. Steentoft, S.Y. Vakhrushev, H.J. Joshi, Y. Kong, M.B. Vester-Christensen, K.T. Schjoldager, K. Lavrsen, S. Dabelsteen, N.B. Pedersen, L. Marcos-Silva, R. Gupta, E.P. Bennett, U. Mandel, S. Brunak, H.H. Wandall, S.B. Levery, H. Clausen, Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology, EMBO J. 32 (2013) 1478–1488.
[67] A. Halim, G. Brinkmalm, U. Ruetschi, A. Westman-Brinkmalm, E. Portelius, H. Zetterberg, K. Blennow, G. Larson, J. Nilsson, Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 11848–11853.
[68] J.C. Trinidad, R. Schoepfer, A.L. Burlingame, K.F. Medzihradszky, N- and O-glycosylation in the murine synaptosome, Mol. Cell. Proteomics 12 (2013) 3474–3488.
[69] T.A. Gerken, O. Jamison, C.L. Perrine, J.C. Collette, H. Moinova, L. Ravi, S.D. Markowitz, W. Shen, H. Patel, L.A. Tabak, Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases, J. Biol. Chem. 286 (2011) 14493–14507.
[70] Y. Kong, H.J. Joshi, K.T. Schjoldager, T.D. Madsen, T.A. Gerken, M.B. Vester-Christensen, H.H. Wandall, E.P. Bennett, S.B. Levery, S.Y. Vakhrushev, H. Clausen, Probing polypeptide GalNAc-transferase isoform substrate specificities by in vitro analysis, Glycobiology (2014).
[71] C.T. Esapa, R.A. Head, J. Jeyabalan, H. Evans, T.A. Hough, M.T. Cheeseman, E.G. McNally, A.J. Carr, G.P. Thomas, M.A. Brown, P.I. Croucher, S.D. Brown, R.D. Cox, R.V. Thakker, A mouse with an N-Ethyl-N-nitrosourea (ENU) Induced Trp589Arg Galnt3 mutation represents a model for hyperphosphataemic familial tumoural calcinosis, PLoS One 7 (2012) e43205.
[72] S. Ichikawa, A.H. Sorenson, A.M. Austin, D.S. Mackenzie, T.A. Fritz, A. Moh, S.L. Hui, M.J. Econs, Ablation of the Galnt3 gene leads to low-circulating intact fibroblast growth factor 23 (Fgf23) concentrations and hyperphosphatemia despite increased Fgf23 expression, Endocrinology 150 (2009) 2543–2550.
[73] L.A. Tabak, The role of mucin-type O-glycans in eukaryotic development, Semin. Cell Dev. Biol. 21 (2010) 616–621.
[74] C.J. Willer, S. Sanna, A.U. Jackson, A. Scuteri, L.L. Bonnycastle, R. Clarke, S.C. Heath, N.J. Timpson, S.S. Najjar, H.M. Stringham, J. Strait, W.L. Duren, A. Maschio, F. Busonero, A. Mulas, G. Albai, A.J. Swift, M.A. Morken, N. Narisu, D. Bennett, S. Parish, H. Shen, P. Galan, P. Meneton, S. Hercberg, D. Zelenika, W.M. Chen, Y. Li, L.J. Scott, P.A. Scheet, J. Sundvall, R.M. Watanabe, R. Nagaraja, S. Ebrahim, D.A. Lawlor, Y. Ben-Shlomo, G. Davey-Smith, A.R. Shuldiner, R. Collins, R.N. Bergman, M. Uda, J. Tuomilehto, A. Cao, F.S. Collins, E. Lakatta, G.M. Lathrop, M. Boehnke, D. Schlessinger, K.L. Mohlke, G.R. Abecasis, Newly identified loci that influence lipid concentrations and risk of coronary artery disease, Nat. Genet. 40 (2008) 161–169.
[75] M.T. Boskovski, S. Yuan, N.B. Pedersen, C.K. Goth, S. Makova, H. Clausen, M. Brueckner, M.K. Khokha, The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality, Nature 504 (2013) 456–459.
[76] J.L. Praissman, L. Wells, Mammalian O-mannosylation pathway: glycan structures, enzymes, and protein substrates, Biochemistry 53 (2014) 3066–3078.
[77] J. Finne, T. Krusius, R.K. Margolis, R.U. Margolis, Novel mannitol-containing oligosaccharides obtained by mild alkaline borohydride treatment of a chondroitin sulfate proteoglycan from brain, J. Biol. Chem. 254 (1979) 10295–10300.
[78] A. Chiba, K. Matsumura, H. Yamada, T. Inazu, T. Shimizu, S. Kusunoki, I. Kanazawa, A. Kobata, T. Endo, Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve alpha-dystroglycan. The role of a novel O-mannosyl-type oligosaccharide in the binding of alpha-dystroglycan with laminin, J. Biol. Chem. 272 (1997) 2156–2162.

[79] R. Harrison, P.G. Hitchen, M. Panico, H.R. Morris, D. Mekhaiel, R.J. Pleass, A. Dell, J.E. Hewitt, S.M. Haslam, Glycoproteomic characterization of recombinant mouse alpha-dystroglycan, Glycobiology 22 (2012) 662–675.
[80] J. Nilsson, J. Nilsson, G. Larson, A. Grahn, Characterization of site-specific O-glycan structures within the mucin-like domain of alpha-dystroglycan from human skeletal muscle, Glycobiology 20 (2010) 1160–1169.
[81] S.H. Stalnaker, K. Aoki, J.M. Lim, M. Porterfield, M. Liu, J.S. Satz, S. Buskirk, Y. Xiong, P. Zhang, K.P. Campbell, H. Hu, D. Live, M. Tiemeyer, L. Wells, Glycomic analyses of mouse models of congenital muscular dystrophy, J. Biol. Chem. 286 (2011) 21180–21190.
[82] S.H. Stalnaker, S. Hashmi, J.M. Lim, K. Aoki, M. Porterfield, G. Gutierrez-Sanchez, J. Wheeler, J.M. Ervasti, C. Bergmann, M. Tiemeyer, L. Wells, Site mapping and characterization of O-glycan structures on alpha-dystroglycan isolated from rabbit skeletal muscle, J. Biol. Chem. 285 (2010) 24882–24891.
[83] T. Martinez, D. Pace, L. Brady, M. Gerhart, A. Balland, Characterization of a novel modification on IgG2 light chain. Evidence for the presence of O-linked mannosylation, J. Chromatogr. A 1156 (2007) 183–187.
[84] K.L. Abbott, R.T. Matthews, M. Pierce, Receptor tyrosine phosphatase beta (RPTPbeta) activity and signaling are attenuated by glycosylation and subsequent cell surface galectin-1 binding, J. Biol. Chem. 283 (2008) 33026–33035.
[85] C. Bleckmann, H. Geyer, A. Lieberoth, F. Splittstoesser, Y. Liu, T. Feizi, M. Schachner, R. Kleene, V. Reinhold, R. Geyer, O-glycosylation pattern of CD24 from mouse brain, Biol. Chem. 390 (2009) 627–645.
[86] S. Pacharra, F.G. Hanisch, I. Breloy, Neurofascin 186 is O-mannosylated within and outside of the mucin domain, J. Proteome Res. 11 (2012) 3955–3964.
[87] S. Pacharra, F.G. Hanisch, M. Muhlenhoff, A. Faissner, U. Rauch, I. Breloy, The lecticans of mammalian brain perineural net are O-mannosylated, J. Proteome Res. 12 (2013) 1764–1771.
[88] T. Yoshida-Moriguchi, T. Willer, M.E. Anderson, D. Venzke, T. Whyte, F. Muntoni, H. Lee, S.F. Nelson, L. Yu, K.P. Campbell, SGK196 is a glycosylation-specific O-mannose kinase required for dystroglycan function, Science 341 (2013) 896–899.
[89] Y. Hara, M. Kanagawa, S. Kunz, T. Yoshida-Moriguchi, J.S. Satz, Y.M. Kobayashi, Z. Zhu, S.J. Burden, M.B. Oldstone, K.P. Campbell, Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 17426–17431.
[90] R. Barresi, K.P. Campbell, Dystroglycan: from biosynthesis to pathogenesis of human disease, J. Cell Sci. 119 (2006) 199–207.
[91] K. Inamori, T. Yoshida-Moriguchi, Y. Hara, M.E. Anderson, L. Yu, K.P. Campbell, Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE, Science 335 (2012) 93–96.
[92] L. Wells, The o-mannosylation pathway: glycosyltransferases and proteins implicated in congenital muscular dystrophy, J. Biol. Chem. 288 (2013) 6930–6935.
[93] S. Joenvaara, I. Ritamo, H. Peltoniemi, R. Renkonen, N-glycoproteomics — an automated workflow approach, Glycobiology 18 (2008) 339–349.