OPEN BIOLOGY

royalsocietypublishing.org/journal/rsob

Review

**Cite this article:** Zhu F, Li D, Chen K. 2019 Structures and functions of invertebrate glycosylation. *Open Biol.* **9**: 180232.
http://dx.doi.org/10.1098/rsob.180232

Received: 19 November 2018
Accepted: 8 January 2019

**Subject Area:**
molecular biology

**Keywords:**
glycosylation, invertebrate, glycoconjugates, structure and function, glycan profiling

**Author for correspondence:**
Keping Chen
e-mail: kpchen@ujs.edu.cn

THE ROYAL SOCIETY PUBLISHING

# Structures and functions of invertebrate glycosylation

Feifei Zhu[1,2], Dong Li[1] and Keping Chen[1]

[1]Institute of Life Sciences, and [2]School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, People's Republic of China

FZ, 0000-0003-4152-0857; KC, 0000-0001-5254-2299

Glycosylation refers to the covalent attachment of sugar residues to a protein or lipid, and the biological importance of this modification has been widely recognized. While glycosylation in mammals is being extensively investigated, lower level animals such as invertebrates have not been adequately interrogated for their glycosylation. The rich diversity of invertebrate species, the increased database of sequenced invertebrate genomes and the time and cost efficiency of raising and experimenting on these species have enabled a handful of the species to become excellent model organisms, which have been successfully used as tools for probing various biologically interesting problems. Investigation on invertebrate glycosylation, especially on model organisms, not only expands the structural and functional knowledgebase, but also can facilitate deeper understanding on the biological functions of glycosylation in higher organisms. Here, we reviewed the research advances in invertebrate glycosylation, including N- and O-glycosylation, glycosphingolipids and glycosaminoglycans. The aspects of glycan biosynthesis, structures and functions are discussed, with a focus on the model organisms *Drosophila* and *Caenorhabditis*. Analytical strategies for the glycans and glycoconjugates are also summarized.

## 1. Introduction

Glycosylation is a posttranslational modification that ubiquitously occurs in eukaryotes. Compared to higher organisms such as mammals whose glycobiology is being extensively studied, invertebrate glycobiology is frequently neglected and the investigation is limited, fragmentary and unsystematic, partly due to perceived lack of importance compared to that of vertebrates. It is estimated that over 97% of earth's animal species are invertebrates; however, only a limited number of species have been studied with respect to their molecular biology. So far, many invertebrate glycomic studies have focused on recombinant glycoproteins, for example expressed using the baculovirus system. Recent years, increasing numbers of studies have been focusing on the glycomes originally derived from invertebrate species.

*Drosophila* and *Caenorhabditis* are by far the most well-studied invertebrates in glycobiology. As multicellular organisms, they serve as better models compared to lower eukaryotes such as yeast for the investigation of glycosylation and glycoengineering. Some invertebrates can have similar biological activities to those of higher organisms, and yet do not pose safety and ethical issues typically associated with vertebrate models for research purposes. Additionally, the time and cost input for establishing an invertebrate model can be greatly reduced. As a result, the knowledge base for invertebrate glycobiology is continually expanding. Studying invertebrate glycosylation, especially on a model organism, often sheds light on biological functions of the glycoconjugates and assists understanding of glycobiology and targeted glycoengineering in both invertebrates and vertebrates.

This review intends to discuss and summarize the knowledge and research advances related to invertebrate glycosylation, with a focus on the Arthropoda

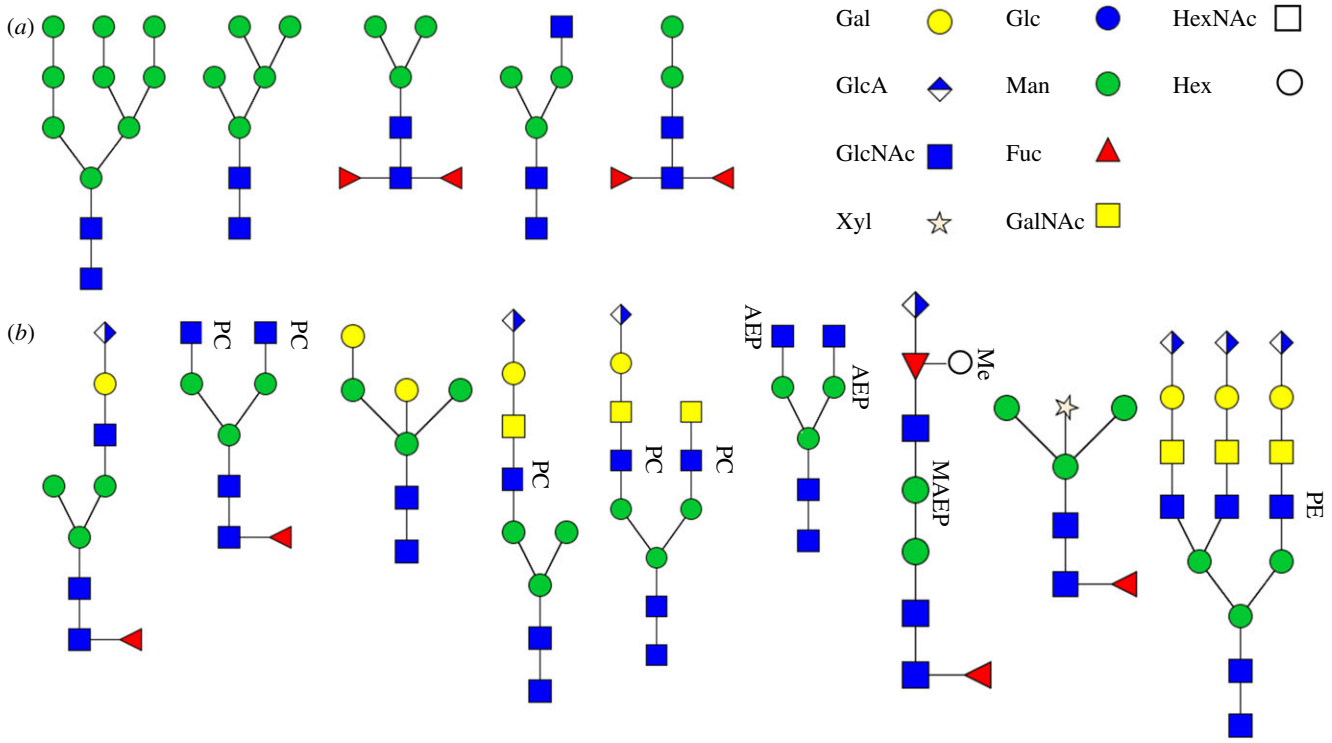


**Figure 1.** (*a*) Examples of *N*-glycans identified in both *Caenorhabditis* and *Drosophila*, whose structures are drawn from the annotated glycan database UnicarbKB. (*b*) Examples of novel *N*-glycans found in invertebrate species, from left to right: *Aedes aegyptii, Pristionchus pacificus, C. elegans, T. ni, L. dispar, Locusta migratoria, V. rubella, Schistosoma mansoni, honeybee royal jelly* [5–11]. Glycans are shown according to the nomenclature of the Consortium for Functional Glycomics. PC, phosphorylcholine; MAEP, methylaminoethylphosphonate; AEP, aminoethyl phosphonate; PE, phosphoethanolamine.

model insect *Drosophila* and the nematode worm *Caenorhabditis*. The review aims to focus on the biosynthesis, structures and functions of the glycans and glycoconjugates found in invertebrates, including protein *N*- and *O*-glycans, glycosphingolipids (GSLs) and glycosaminoglycans (GAGs). Meanwhile, analytical strategies for glycan and glycoconjugate analysis will be discussed.

## 2. N-glycosylation in invertebrates

N-glycosylation refers to the attachment of a glycan to the asparagine side chain of a protein. This modification occurs almost exclusively when the asparagine residue is followed by the XT/S sequon regardless of species type, where X refers to any amino acid residue except proline. The highly conserved sequon directs the biosynthesis of *N*-glycans to a protein in the endoplasmic reticulum (ER) and Golgi compartments and ensures that the glycan decoration follows a core machinery. Cells have complex and exquisite machinery for protein glycosylation. Inside animal cells, protein N-glycosylation initiates in the ER, where the carbohydrate moiety $Glc_3Man_9GlcNAc_2$ is synthesized through the action of a series of glycosyltransferases anchored in the ER and then attached to a newly translated protein [1]. The terminal glucoses and mannoses are subsequently removed by their respective glycosidases, and the remaining glycan–protein complex is carried onto the Golgi apparatus for further glycomic editing by different glycosyltransferases and glycosidases, producing variably branched and extended glycan structures [1].

However, the N-glycosylation synthesis routes in invertebrates remain ambiguous and controversial. It is reported that some insects, such as the silkworm *Bombyx mori*, possess a similar glycan synthesis route to that of the mammals [2]. But unlike vertebrates, the existence of an active *N*-acetylglucosaminidase in the insect cells interrupts the biosynthesis of the complex and hybrid-type *N*-glycans, resulting in almost exclusively high mannose and paucimannose type *N*-glycans [3]. Nevertheless, several species have been found to produce hybrid- and complex-type glycans. *Drosophila*, one of the best studied model insect in glycobiology, possesses similar N-glycosylation machinery to that of vertebrates and produces hybrid- and complex-type *N*-glycans that were once thought absent in insects [4]. However, the relative amount of these complex- and hybrid-type *N*-glycans tends to be low, usually below 10% of the entire *N*-glycan pool. Another well-studied invertebrate, the nematode *Caenorhabditis*, also contains a nearly contiguous series of *N*-glycans (figure 1) [12]. *Caenorhabditis* was also found to contain fuco-paucimannosidic and bisecting fucose–galactose branched glycans that are unique to this nematode [5,6,12].

Recently, dipteran species, specifically mosquito larvae, were found to contain sulfated and glucuronylated antennae [13], indicating that insect glycans can have higher levels of structure complexity and variation than previously expected. In addition to sulfated and glucuronylated, core difucosylated and zwitterion phosphorylcholine and phosphoethanolamine-modified *N*-glycans were also identified in a handful of invertebrate species such as *Trichoplusia ni* and *Lymantria dispar* (figure 1) [7]. The mollusc *Volvarina rubella* was also found to contain novel *N*-glycans with phosphonate and phosphorylcholine modifications in addition to Fuc and GlcA modifications (figure 1) [8,9]. Additionally, xylosylated glycans and triantennary phosphoethanolamine-modified glucuronylated glycans have also been identified from *Schistosoma* and honeybee royal jelly, respectively (figure 1) [10,11]. These findings have vastly expanded the current insect glycan repertoire and enabled a fresh look at invertebrate glycan structures and their functions.

The majority of membrane and secreted proteins are cotranslationally N-glycosylated and are involved in a broad range of biological activities. *N*-Glycoproteins are found to be involved predominantly in cell–cell adhesion, body growth, embryonic development and organ development [14]. Cell glycan biosynthesis is facilitated by a few hundred enzymes, including glycosyltransferases, glycosidases and enzymes related to sugar modification, metabolism and transport, etc. [15]. The glycoenzyme set involved in protein N-glycosylation appears to be different between different species orders, as evidenced by highly conserved N-glycoproteomes within their respective phyla and different *N*-glycan antennal modifications between evolutionarily distant species [14,16]. Extensive studies have shown that mutations of any of the glycoenzymes are likely to cause serious morphological and developmental defect or even death in many invertebrate species. The *Drosophila alg5* gene, which codes for an enzyme involved in the early steps of protein N-glycosylation, is essential for the correct epidermal differentiation during *Drosophila* late embryogenesis [17]. Mutation of the *Drosophila* gene *xit*, which encodes an enzyme involved in the addition of the terminal glucose to the *N*-glycan precursor, impairs cell intercalation in the lateral epidermis during germband extension and apical constriction of mesoderm precursor cells [18]. Thus, identification of invertebrate genes encoding the glycoenzymes involved in glycan biosynthesis seems highly necessary in order to facilitate deeper understanding of invertebrate glycosylation and to precisely engineer desired glycan structures using invertebrate models. In fact, genetically engineered mutants of *Caenorhabditis* and *Drosophila* have been established in order to reveal the biological functions of specific glycoenzymes such as fucosyltransferases [5], *N*-acetylglucosaminyltransferases [19], hexosaminidases [20,21] and other glycoenzymes [22,23].

## 3. Profiling of glycans and glycoconjugates in invertebrates

The structural diversity and heterogeneity of glycoconjugates is attributed mainly to the attached glycans. Unlike proteins and nucleotides, glycan structures can be highly branched, and the monosaccharide units composing the glycans are often isomeric, leading to increased structural complexity. Additional structural complexity arises from the different glycosidic linkages between the monosaccharides as well as further decoration of the monosaccharides by sulfation and phosphoethanolamine. Bacteria and archaea have the broadest diversity of monosaccharides, with a total of approximately a hundred different types of monosaccharides [24]. Plants are the next, and then eukaryotic animals. There are 11 monosaccharides typically found in invertebrate glycans: glucose, mannose, galactose, *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), xylose, fucose, *N*-acetylneuraminic acid (NeuAc), *N*-glycoylneuraminic acid (NeuGc), glucuronic acid (GlcA) and iduronic acid. Further modifications of the monosaccharides such as methylation, sulfation and zwitterionic modification including phosphorylcholine, phosphoethanolamine and aminoethyl phosphate have also been identified in invertebrates.

Because of the high structural complexity, structural analysis of glycans and glycoconjugates remains very challenging. The main workhorse in today's glycomic and glycoproteomic analysis is the mass spectrometry (MS)-based approach due to its high-sensitivity and high-throughput capabilities. Auxiliary methods, such as glycosidase digestion, and less frequently nuclear magnetic resonance spectroscopy, could also be used as complementary tools to confirm carbohydrate structures. In MS-based approaches, improving glycan detection sensitivity and assigning MS spectra are among the most challenging steps. Here, in this part, we will briefly discuss large-scale mapping of glycans and glycoconjugates in invertebrates by MS-based techniques.

Currently, large-scale identification of N-glycosylation sites have been quite routine. More than a thousand insect proteins have been found to be N-glycosylated. Zielinska *et al*. [25] developed a filter-aided deamidation method which takes advantage of the enzymatic deglycosylation reaction that turns the asparagine residue into aspartic acid. When the reaction is carried out in $^{18}O$ water, a fixed mass shift of 2.989 occurs, which can be readily detected by mass spectrometry and analysed automatically using software platforms such as MaxQuant [26]. Using this method, the N-glycoproteomes across seven evolutionarily distant species were mapped, including *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Danio rerio* [14]. *N*-Glycoproteins are found to be almost exclusively located outside the cell including the cell membrane, or in anticipated intracellular organelles such as ER and Golgi [14]. Analyses of N-Glycoproteome orthologues within and across different phyla indicate that each phylum has characteristic N-glycoproteomes that are distinct from species in other phyla [14]. Therefore, it seems necessary, not just out of curiosity, to investigate the N-glycoproteomes across different species, especially in insects that currently have a limited N-glycoproteome revealed, in order to fully understand the functions of protein glycosylation.

Although the filter-aided deamidation method can perform large-scale identification of protein glycosylation sites, it cannot provide information on the specific glycan structures attached to the glycosylation site. Typically, there are two means to obtain glycan structural information. The first way is to isolate the glycans from proteins and analyse the glycome separately. Currently, automated and high-throughput glycomic profiling is under rapid development and has achieved great progress. In typical high-throughput glycomic analysis, glycans are cleaved from proteins either by PNGase or by other chemical methods, enriched and then labelled before LC–MS analysis [27]. Glycan structures can be deciphered by software such as MultiGlycan [28], SimGlycan [29] and GlycoWorkBench [30], which has greatly promoted large-scale structural analysis of glycans with various labelling techniques. A handful of invertebrate species, such as *Drosophila* [31,32] and *Caenorhabditis* [33,34] have had their glycome analysed, which has significantly renewed the understanding of structural and functional glycobiology in invertebrates. That said, in-depth analysis of the N-glycomes, especially finding novel glycan structures, can still be quite challenging and far from routine. A combination of exoglycosidase digestion, offline LC separation and purification, as well as MALDI-TOF MS/MS or LC/MS/MS analysis is usually needed in order to reveal new glycan structures [19].

The other way to obtain glycan structural information is to study the glycopeptides with the glycans attached, which is even more challenging compared to glycomic study alone, because the former involves peptides and glycan identification

simultaneously. Developing robust and reliable pipelines for large-scale profiling of glycopeptides is still under way. The most challenging steps during the pipeline development, other than developing high-sensitivity MS detection methods for glycopeptides, perhaps is automated assignment of MS and MS/MS spectra to specific glycopeptides. Currently, software such as GlycoMaster DB [35], Byonic [36] and ArMone [37] has been developed to perform automated large-scale analysis on intact glycopeptides based on MS fragmentation datasets. So far, the glycopeptide identification pipelines have been quite successfully applied in studying large glycoproteomes for mammalian tissues and organs; however, very few reports have been focused on large-scale profiling of invertebrate glycoproteomes.

## 4. O-glycosylation in invertebrates

Depending on the first monosaccharides linked to the protein, O-glycosylation can be further divided into O-GalNAcylation, O-mannosylation, O-fucosylation and so on. Current knowledge on invertebrate O-glycosylation is still very rudimentary. In contrast with N-glycosylation, there is no consensus sequence for O-glycosylation. Serine and threonine are the most common accepting amino acids for these modifications. Tyrosine, hydroxylysine and hydroxyproline are also possible residues for O-glycosylation. In addition, *O*-glycan biosynthesis is carried out by the addition of monosaccharide one after another, which is different from the case of N-glycosylation. O-GalNAcylation is one of the most common forms of O-glycosylation, and was first identified in mucin, a heavily glycosylated protein 40% of whose molecular weight is occupied by glycans. Therefore, O-GalNAcylated glycans are also called mucin-type glycans. Approximately 90% of all *O*-glycans in *Drosophila* belong to the mucin type [38]. In mucin-type O-glycosylation, a GalNAc residue is directly linked to serine or threonine for the initiation of protein glycosylation. In invertebrates, mucin-type glycans are found extensively in secreted proteins or the extracellular part of membrane proteins. Currently, except mucin-type *O*-glycans, other types of *O*-glycans in invertebrates have not been investigated in detail. There are eight common O-glycosylation cores found in mammals [39]; however, few of these have been found to be present in invertebrates. Examples of mucin-type *O*-glycans for the model organisms *Caenorhabditis* and *Drosophila* are shown in figure 2.

In invertebrates, various UDP-GalNAc transferases initiate O-GalNAcylation by modifying the Ser/Thr residues with a GalNAc [40]. Following extension of the glycan chain results in several different core structures. For the core-1 structure, β1–3 galactosyltransferase adds an additional galactose to the GalNAc residue. Additional glucuronic acid or GlcNAc residues can be found to further modify the glycan chain (figure 2). Core-1 type and extended core-1 mucin-type glycan structures have been identified in *Caenorhabditis* and *Drosophila* (figure 2). *Caenorhabditis* also synthesizes core 1 mucin-type glycans substituted on Gal and/or GalNAc by Glc residues, similar to those in vertebrates [41]. So far, sulfated *O*-glycans have been identified in *Drosophila* but not *Caenorhabditis* [5], and phosphoethanolamine modification has been reported in the insect *Vespula germanica* [42], suggesting the potential of broad-spectrum modifications to invertebrate O-glycosylation that are waiting to be discovered.

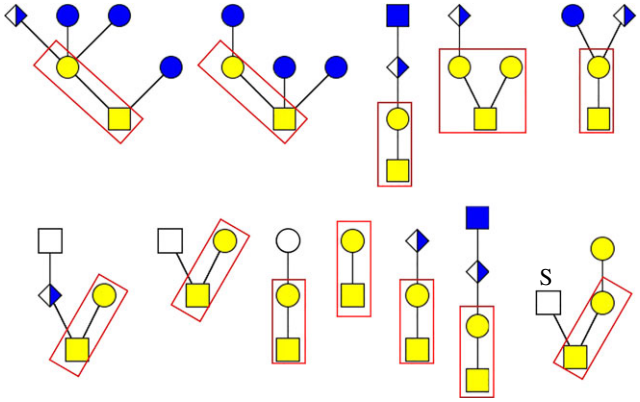


**Figure 2.** Examples of mucin-type *O*-glycans in *Caenorhabditis* (upper row, from annotated glycan database UnicarbKB) and *Drosophila* (lower row) [38]. Red rectangles denote common *O*-glycan cores.

Mucin-type *O*-glycan modification is critical for the development and function of multicellular organisms. In *Drosophila*, the *pgant* gene family, which is responsible for encoding the GalNAc transferases, is essential for viability of the insect. O-GalNAcylation regulates essential developmental programmes and modulates trafficking through the secretory pathway, and mutation or silencing of glycoenzymes such as the *pgant* genes results in malfunctions in cell adhesion and Golgi trafficking [43]. However, O-glycosylation functions in other invertebrate species have not been fully elucidated and need further investigation.

O-fucosylation and O-glucosylation are typical modifications to protein epidermal growth factor (EGF)-like domains and thrombospondin type 1 (TSP1) domains. The O-fucosylation site of the EGF domain is usually flanked by cysteine residues that form disulfide bridges, which can be recognized by O-fucosyltransferase (OFUT)1, whereas in TSP1 domains another O-fucosyltransferase, OFUT2, recognizes and directs the transfer of fucose to the protein. In *Drosophila*, mutations to OFUT1 lead to multiple organ defects and lethality [44,45]. The addition of a GlcNAc residue to the fucose is accomplished by Fringe, an *N*-acetylglucosaminyltransferase in *Drosophila* that is involved in many functions such as eye development [46] and adult eclosion and survival [47]. In O-glucosylation, transfer of glucose to the serine/threonine residue is achieved by a single glucosyltransferase named Rumi [48], and the addition of xylose to the glucose is accomplished by the enzyme Shams [49]. In *Drosophila* protein Notch, 22 sites were found to be O-fucosylated and 18 were O-glucosylated [50]. O-glucosylation and O-fucosylation function cooperatively and play important roles in Notch transport and signalling in *Drosophila* [51].

*O*-Mannose glycans constitute less than 1% of the total *Drosophila* glycan pool [38]. O-mannosylation has been gaining interest due to its conserved process across most eukaryotes and defects in this modification give rise to human diseases [52]. The transfer of mannose to serine or threonine is accomplished by the enzyme family *O*-mannosyltransferases. In *Drosophila*, O-mannosylation occurs in the protein known as dystroglycan. The *Drosophila* homologues of two *O*-mannosyltransferase genes, POMT1 and POMT2, function in association with each other to maintain normal muscle development [53]. Another type of O-mannosylation, such as for the cadherin superfamily, depends on the TMTC-type mannosyltransferases for O-mannosylation [54];

however, their mannosyltransferase activity in *Caenorhabditis* and *Drosophila* is yet to be proven. In *Drosophila*, POMT mutation results in rotated abdomen, defective synaptic transmission and muscle dystrophy [53,55]. In humans, mutations to these two transferases cause autosomal recessive disorder, which leads to malfunction in the brain, muscle and eye [56]. A summary of the main functions of invertebrate glycans is presented in table 1.

Compared to *N*-glycan analysis, *O*-glycan poses more analytical challenges, partly because there are currently no universal enzymes available to cleave off *O*-glycans from the proteins. General strategies to removed *O*-glycans from proteins are through chemical methods such as alkaline β-elimination or hydrazine hydrolysis. In addition, O-linked glycans do not have a common core structure, and instead can have more than eight types of core structures. Furthermore, unlike *N*-glycans which appear to be synthesized following predefined structural antennae, *O*-glycans seem to branch more irregularly, making it difficult to define the specific structure. Often times, glycosidases are used in addition to mass spectrometry to determine the definite structures.

## 5. *O*-GlcNAc modification

O-linked GlcNAc modification refers to the addition of a single GlcNAc residue to the serine or threonine residue of a protein. This modification is reversible and highly dynamic in that GlcNAc is added and removed regularly depending on the cellular environment by two unique enzymes, the O-linked GlcNAc transferase (OGT) and the O-linked GlcNAcase (OGA). In contrast with other types of O-glycosylation, O-GlcNAcylation can actually occur in the nucleus and cytosol rather than in the ER and Golgi apparatus; therefore, O-GlcNAcylation is in nature more akin to protein phosphorylation than typical O-glycosylation. This modification is particularly heavily present on proteins involved in signalling, stress response and energy metabolism such as nuclear pore proteins, phosphatases, metabolic enzymes, etc. [68]. O-GlcNAcylation regulates protein translation, stability and turnover, and has been demonstrated to be engaged in neurodegenerative diseases, diabetes and cancer [57–59]. Alteration of *O*-GlcNAc profiles of several proteins in the pancreatic β-cell has been reported to associate with the upregulation of insulin secretion from the pancreas [60]. There is also an extracellular form of O-GlcNAcylation on EGF repeats mediated by the EOGT enzyme in the ER [69].

O-GlcNAc modification is ubiquitous and essential in multicellular organisms. Mutations of genes relating to the GlcNAc modification will cause severe growth phenotypes or even death. The deletion of OGT or OGA from *C. elegans* results in up- and down-regulation of hundreds of transcripts, which is likely due to the misguided *O*-GlcNAc modification of RNA polymerase II as well as the basal transcription complex, indicating the role of O-GlcNAcylation in regulating transcription [70]. The disruption of the OGT gene in *C. elegans* has been shown to induce metabolic disorder and reduced lifespan [71]. A similar metabolic disorder was also observed in the species with OGA gene disruption, but interestingly its lifespan was extended. In *Drosophila*, the extent of protein O-GlcNAcylation was found to increase with the developmental stage [72]. Shortened liftspan was observed in an OGT gene

**Table 1.** Invertebrate glycosylation types and functions.

| type | exemplary structure | main known functions | references |
|---|---|---|---|
| *N*-glycan |  | cell–cell adhesion, body growth, embryonic development and organ development | [14,17,18] |
| *O*-glycan | mucin-type *O*-glycan |  | cell adhesion, Golgi trafficking and organ development | [38,43] |
| | *O*-fucose glycan |  | Notch transport and signalling | [44–47,49–51] |
| | *O*-glucose glycan |  | | |
| | *O*-mannose glycan |  | growth and embryonic/organ development | [53,55,56] |
| | *O*-GlcNAc glycan |  | regulating transcription, protein functions and metabolism | [57–60] |
| glycosphingolipid glycan |  | host–pathogen interactions, cell recognition, development and modulating transmembrane signalling | [61–63] |
| glycosaminoglycan |  | host–pathogen interactions, embryogenesis, regenerative and developmental roles | [64–67] |

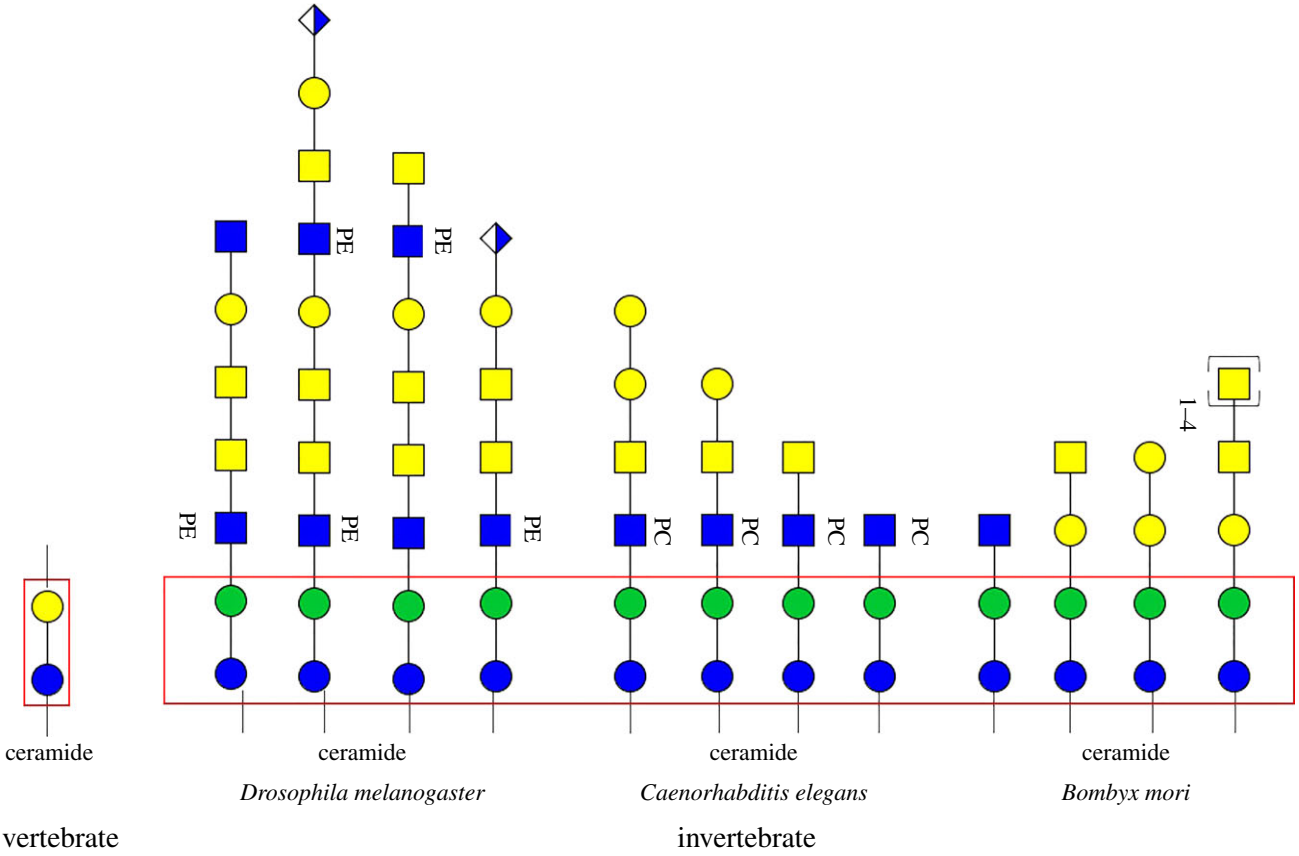


**Figure 3.** Exemplary GSL structures in invertebrate species *D. melanogaster* [82,85], *C. elegans* [86] and *B. mori* [87]. Red rectangles denote GSL cores. PC, phosphorylcholine; PE, phosphoethanolamine.

mutant of *Drosophila*, which survived through the larval stages but died in the pupal stages [73]. Because the OGT gene is highly conserved throughout species, the OGT gene in humans has been transgenically introduced for the rescue of OGT null *Drosophila*.

Most current studies on O-GlcNAcylation have been focused on the functional aspects; profiling of *O*-GlcNAc modified proteins has been reported in mammals but seldom in invertebrates. As a structurally simple modification, *O*-GlcNAc modification is even more difficult to detect and determine. First, as mentioned above, *O*-GlcNAc modification is dynamically cycling on and off the modified peptide depending on the cell environment, which is easily deglycosylated by the glycosidases that exist in the cell during cell lysis. To prevent autodeglycosylation, an *O*-GlcNAc glycosidase inhibitor, such as PUGNAc, is added during cell lysis and sample preparation. Another issue is associated with O-GlcNAcylation detection, such as by mass spectrometry, the most popular technique currently used for *O*-GlcNAc analysis. The attached *O*-GlcNAc is sensitive to the electrospray ionization process and readily falls off from the peptide backbone. Last but not least is the lack of efficient and widely applicable enrichment methods for O-GlcNAcylated peptides prior to MS detection. The traditional glycan enrichment method lectin affinity is not quite applicable in the case of O-GlcNAcylation, because when only a single GlcNAc residue is attached to the peptide, the interaction between O-GlcNAcylated peptide and the lectin is very weak. Nevertheless, by taking advantage of this weak interaction, Vosseller *et al.* [74] developed a lectin weak affinity chromatography method for targeted enrichment of O-GlcNAcylated peptides. Using a similar method, more than 1700 O-GlcNAcylated peptides have been identified in mouse synaptic membrane [75]. O-GlcNAcylated peptides can also be enriched via immunoprecipitation using antibodies specific to *O*-GlcNAc, such as RL2 [76], CTD110.6 [77], 18B10.C7 [78], etc. Click chemistry-based method has also become quite popular in recent years for the enrichment of O-GlcNAcylated peptides [79]. In this method, *O*-GlcNAc residues are either enzymatically grafted with an azide-tagged GalNAz or metabolically incorporated with an azide-tagged GlcNAc residue, and labelled with biotin and subsequently enriched by avidin or streptavidin immobilized solid matrix [80].

## 6. Glycosphingolipids in invertebrates

GSL is a type of glycolipid commonly found in animals. In GSLs, glycans are covalently linked to a ceramide lipid moiety that is composed of a long-chain alcohol known as sphingosine in amide linkage to a fatty acid. GSLs are structurally diverse mainly due to variable sugar modifications to the ceramide, and the glycan moiety of invertebrate GSLs differs notably from those of vertebrates. Vertebrates have a GSL core disaccharide Gal(β1–4)Glcβ linked to the ceramide, whereas for invertebrates the most common core disaccharide is Man(β1–4)Glc [81–83], except for the mollusc *Aplysia kurodai*, whose GSL core is the same as in vertebrates [84]. In addition, the ceramide moiety of invertebrate GSLs is found to be different from those of vertebrates, in that the sphingosine chain is shorter for invertebrates [81].

Further extension of the glycan core structure is species-specific. The highly diversified and individually tailored GSL structures indicate their important roles in developmental or tissue-restricted functions. In *Drosophila*, the core

mactosyl Man(β1–4)Glc structure can be further modified with a GalNAc(β1–4)GlcNAc(β1–3) residue, and the terminal Gal can be further capped with GlcA (figure 3) [82,85]. Phosphoethanolamine is present as a typical modification to dipteran glycolipids and aminoethylphosphonate to those of molluscs [8,13,84]. In *Caenorhabtidis*, the GSLs were reported to consist of the core structure GlcNAc(β1–3)Man(β1–4)Glc(β1)Cer, similar to that in arthropods (figure 3) [86]. Also, phosphorylcholine is a known antenna component of nematode glycolipids (figure 3) [88,89]. Recently, GSLs of the lepidopteran species *B. mori* were investigated, which contain the same conserved core structure, but novel extensions were revealed (figure 3) [87]. The major GSL components for *Manduca sexta* were identified as mactosyl ceramide [81]. Gangliosides, GSLs modified with sialic acid residues, have not been reported for invertebrates to date.

In invertebrates, GSLs play important roles in host–pathogen interactions, cell recognition and body development [61,90]. Elimination of the *egh* and *brn* genes, which encode mannosyltransferases and N-acetylglucosaminyltransferases that are responsible for the early-step biosynthesis of GSLs, is lethal to *Drosophila* [62]. Mutation of either of the two genes caused overproliferation of neural cells and enlarged peripheral nerves, phenotypes similar to human neurofibromatosis diseases [63]. The results indicate that, like vertebrates, GSLs play pivotal roles in cell recognition and modulating transmembrane signalling.

Although invertebrate GSLs are no longer thought to be GSL-free, profiling of GSL structures have been limited to only a few invertebrate species and the molecular details associated with GSLs are largely unknown. Glycolipid profiling is gaining growing interest motivated by the important biological roles of the glycan head groups. The analysis of GSLs involves determination of both the ceramide and glycan moieties, both of which have high structural diversity. In general, glycolipids are extracted from tissues or body fluids by chloroform–methanol extraction. Separation and analysis of GSLs relies largely on a combination of techniques, such as thin-layer chromatography [87], gas chromatography [62], nuclear magnetic resonance [62,87] mass spectrometry [91], etc., and mostly count on manual annotation and interpretation of the GSL data. High-throughput workflows are still under development. A database and software for the MS analysis of ganglioside and sulfate-modified GSLs are under development for automated interpretation of MS data of the GSLs [92].

## 7. Glycosaminoglycans in invertebrates

GAG is a linear polysaccharide consisting of repeating disaccharide units covalently linked to a protein (proteoglycan). The most abundant cell surface GAG structural subtypes include heparan sulfate characterized by disaccharide unit GlcA(β1–4)GlcNAc(α1–4) and chondroitin sulfate characterized by GlcA(β1–3)GalNAc(β1–4). Variable degrees of sulfation and GlcA epimerization (to iduronic acid) may occur on the GAG backbone. While vertebrates tend to have additional GAG types such as dermatan sulfate and hyaluronic acid, most invertebrates were reported to contain only heparan sulfate and chondroitin sulfate with or without sulfate.

More than 20 chondroitin sulfate proteoglycans have been found in the model organism *Caenorhabditis*, indicating previously underestimated GAG structural and functional heterogeneity in invertebrates [33]. Chondroitin chains in *Caenorhabditis* were once thought not sulfated due to the lack of relative sulfotransferases and epimerase for modifying the sugar residues [93,94]; however, a recent study found an active chondroitin sulfotransferase in this species and its sulfated chondroitin chains were revealed, albeit at a low level [95]. In addition, the *Caenorhabditis* proteoglycan core proteins were found to be different from those found in vertebrates or the invertebrate *Drosophila* [93,96]. The common tetrasaccharide core linking the repeating disaccharides and the serine residue of the proteoglycan is reported as GlcA(β1–3)Gal(β1–3)Gal(β1–4)Xyl for chondroitin and heparan [97], whereby the relevant enzymes synthesizing the core are encoded by genes defective in *sqv* mutants [98]. Recently, a novel GAG tetrasaccharide core with additional galactose and phosphorylcholine modifications was reported for the parasitic nematode *Oesophagostomum dentatum* [99].

Degrees of sulfation on the GAG disaccharide backbone are highly conserved within a class but significantly different between classes of invertebrates [94], indicating evolutionarily distinct functionalities. For example, *Drosophila* chondroitin sulfate is composed of 71% HexA-GalNAc and 29% HexA-GalNAc (4-*O*-sulfate) [100], whereas that of the Chelicerata species *Trachypleus tridentatus* is composed of 46% HexA-GalNAc (4-*O*-sulfate) and 54% GlcA(3-*O*-sulfate)-GalNAc(4-*O*-sulfate) [101]. GAGs are present covalently linked to proteins via type-specific linkages forming proteoglycans. As for *Caenorhabditis*, the GAGs in *Drosophila* are based on the same canonical GlcA-Gal-Gal-Xyl core for attachment to proteins; some of the relevant enzymes have been characterized, such as oxt [102] and GalT7 [103].

GAGs are ubiquitously found on the surface and extracellular matrix of mammalian cells, interacting with various ligands and playing crucial roles in many pathophysiological processes. The chondroitin sulfate GAGs are structural constituents of complex matrices such as cartilage, brain, intervertebral discs, tendons and corneas. Genetic studies on the model organism *Drosophila* showed that heparan sulfate GAGs act as core receptors for many growth factors, and participate in the generation and long-range maintenance of gradients for morphogens during embryogenesis and regenerative processes [64]. Knockdown of GAGs in *Drosophila* reduces the binding of α C protein, a virulence determinant of group B streptococcus, resulting in longer host survival [65]. The results indicate that host cell surface GAGs are vital during pathogen invasion and that interfering with the binding of this sugar may protect the host against infection. Mutation of a heparin sulfate proteoglycan homologue in *Drosophila* leads to cell cycle arrest of neuroblasts in the larval brain [66,67]. The *Drosophila ttv* gene, which encodes an enzyme responsible for adding monosaccharide to the GAG backbone, is a homologue of the mammalian Ext class of tumour suppressor genes that cause human bone dysplasia [104,105].

Typical GAG and proteoglycan analyses use bottom-up approaches. The proteoglycans are extracted by strong denaturing agents and purified by ion exchange or size exclusion chromatography. The GAG moiety can be isolated from the protein via β-elimination or hydrazinolysis, and its disaccharide unit can be degraded by bacterial lyases such as heparinase or chondroitinase, and further derivatized before analysis by gas chromatography, liquid chromatography and/or mass spectrometry [106–108]. However, only compositional information of the disaccharide building blocks can be derived

from the bottom-up approach, and no sequence information can be obtained. Top-down mass spectrometry analysis of GAGs is possible for the direct structure and site determination of intact proteoglycan and GAG structures, but may require pure proteoglycans, which is rather difficult [109]. Additional difficulties arise from the assignment of the GAG MS/MS spectrum for top-down analysis due to the variation of sulfation sites between the disaccharide units. Software programs are being developed for automated annotation of the GAG fragmentation spectra to assist structure elucidation of GAG compositions and sequences [110].

## 8. Concluding remarks

Glycosylation confers heterogeneity on glycoconjugates and finely tunes their structures and functions [111,112]. In vertebrates, aberrant glycosylation can be indicative of various disease states [113,114]. In invertebrates, disturbance of glycosylation pathways has been demonstrated to cause serious defects, such as abnormal metamorphosis and even mortality. The investigation of invertebrate glycosylation often provides new insight into mechanisms underlying physical/neurological impairments in vertebrates and helps to establish novel therapeutic treatment strategies.

It is also found that invertebrate glycomic profiles can change upon alteration of physiological, pathological or developmental stages. A mutant *Caenorhabditis* that is resistant to bacterial infection is observed to be deficient in many *N*- and *O*-glycans compared with its wild-type [115]. Similarly, *Caenorhabditis bre-1* mutant, which encodes an enzyme that catalyses biosynthesis of GDP-mannose to GDP fucose, was found deficient in fucosylated glycoconjugates and resistant to the toxin *Bacillus thuringiensis* [22]. Our laboratory also found that the N-glycomic profiles of *B. mori* alter after BmNPV viral infection (F Zhu, D Li, D Song, P Lv, Yao Q, Chen K 2019, unpublished data). In addition, the monosaccharide profiles of *B. mori* nervous system change with different development stages [116]. Therefore, it seems necessary, not just out of curiosity, to profile the glycoproteomes across different development stages of species, which will contribute to deeper understanding of the functional roles of glycosylation.

Data accessibility. This article has no additional data.

Authors' contributions. F.Z. wrote the manuscript, D.L. formatted and edited the manuscript and K.C. revised the manuscript. All authors gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. Work in the authors' laboratory has been funded by the National Natural Science Foundation of China (31702186 and 31872425), Natural Science Foundation of Jiangsu Province China (BK20160509) and China Postdoctoral Science Foundation (2016M601725).

## References

1. Helenius A. 2001 Intracellular functions of N-linked glycans. *Science* **291**, 2364–2369. (doi:10.1126/science.291.5512.2364)
2. Geisler C, Jarvis DL. 2010 Identification of genes encoding N-glycan processing β-N-acetylglucosaminidases in *Trichoplusia ni* and *Bombyx mori*: implications for glycoengineering of baculovirus expression systems. *Biotechnol. Prog.* **26**, 34–44. (doi:10.1002/btpr.298)
3. Nomura T *et al.* 2015 Improvement of glycosylation structure by suppression of β-N-acetylglucosaminidases in silkworm. *J. Biosci. Bioeng.* **119**, 131–136. (doi:10.1016/j.jbiosc.2014.07.012)
4. Hagen KGT, Zhang L, Tian E, Zhang Y. 2009 Glycobiology on the fly: developmental and mechanistic insights from *Drosophila*. *Glycobiology* **19**, 102–111. (doi:10.1093/glycob/cwn096)
5. Yan S, Brecker L, Jin C, Titz A, Dragosits M, Karlsson NG, Jantsch V, Wilson IBH, Paschinger K. 2015 Bisecting galactose as a feature of N-glycans of wild-type and mutant *Caenorhabditis elegans*. *Mol. Cell. Proteomics* **14**, 2111–2125. (doi:10.1074/mcp.M115.049817)
6. Yan S, Wilson IBH, Paschinger K. 2015 Comparison of RP-HPLC modes to analyse the N-glycome of the free-living nematode *Pristionchus pacificus*. *Electrophoresis* **36**, 1314–1329. (doi:10.1002/elps.201400528)
7. Stanton R, Hykollari A, Eckmair B, Malzl D, Dragosits M, Palmberger D, Wang P, Wilson IBH, Paschinger K. 2017 The underestimated N-glycomes of lepidopteran species. *Biochim. Biophys. Acta* **1861**, 699–714. (doi:10.1016/j.bbagen.2017.01.009)
8. Paschinger K, Wilson IBH. 2016 Analysis of zwitterionic and anionic N-linked glycans from invertebrates and protists by mass spectrometry. *Glycoconjugate J.* **33**, 273–283. (doi:10.1007/s10719-016-9650-x)
9. Eckmair B, Jin C, Abed-Navandi D, Paschinger K. 2016 Multistep fractionation and mass spectrometry reveal zwitterionic and anionic modifications of the N- and O-glycans of a marine snail. *Mol. Cell Proteomics* **15**, 573–597. (doi:10.1074/mcp.M115.051573)
10. Hykollari A, Malzl D, Eckmair B, Vanbeselaere J, Scheidl P, Karlsson NG, Wilson IBH, Paschinger K. 2018 Isomeric separation and recognition of anionic and zwitterionic N-glycans from royal jelly glycoproteins. *Mol. Cell Proteomics* **17**, 2177–2196. (doi:10.1074/mcp.RA117.000462)
11. Khoo K-H, Huang H-H, Lee K-M. 2001 Characteristic structural features of schistosome cercarial N-glycans: expression of Lewis X and core xylosylation. *Glycobiology* **11**, 149–163. (doi:10.1093/glycob/11.2.149)
12. Cipollo JF, Costello CE, Hirschberg CB. 2002 The fine structure of *Caenorhabditis elegans* N-glycans. *J. Biol. Chem.* **277**, 49 143–49 157. (doi:10.1074/jbc.M208020200)
13. Kurz S, Aoki K, Jin C, Karlsson NG, Tiemeyer M, Wilson IBH, Paschinger K. 2015 Targeted release and fractionation reveal glucuronylated and sulphated N- and O-glycans in larvae of dipteran insects. *J. Proteomics* **126**, 172–188. (doi:10.1016/j.jprot.2015.05.030)
14. Zielinska DF, Gnad F, Schropp K, Wiśniewski JR, Mann M. 2012 Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell.* **46**, 542–548. (doi:10.1016/j.molcel.2012.04.031)
15. Neelamegham S, Mahal LK. 2016 Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure. *Curr. Opin. Struct. Biol.* **40**, 145–152. (doi:10.1016/j.sbi.2016.09.013)
16. Gagneux P, Aebi M, Varki A. 2017 Evolution of glycan diversity. In *Essentials of glycobiology* (eds A Varki, RD Cummings, JD Esko), pp. 2015–2017, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
17. Shaik KS, Pabst M, Schwarz H, Altmann F, Moussian B. 2011 The Alg5 ortholog Wollknäuel is essential for correct epidermal differentiation during *Drosophila* late embryogenesis. *Glycobiology* **21**, 743–756. (doi:10.1093/glycob/cwq213)
18. Zhang Y, Kong D, Reichl L, Vogt N, Wolf F, Grasshans J. 2014 The glucosyltransferase Xiantuan of the endoplasmic reticulum specifically affects E-Cadherin expression and is required for gastrulation movements in *Drosophila*. *Dev. Biol.* **390**, 208–220. (doi:10.1016/j.ydbio.2014.03.007)
19. Yan S, Wang H, Schachter H, Jin C, Wilson IBH, Paschinger K. 2018 Ablation of N-acetylglucosaminyltransferases in *Caenorhabditis*

induces expression of unusual intersected and bisected N-glycans. *Biochim. Biophys. Acta Gen. Subj.* **1862**, 2191–2203. (doi:10.1016/j.bbagen.2018.07.002)

20. Yan S *et al.* 2012 Galactosylated fucose epitopes in nematodes: increased expression in a *Caenorhabditis* mutant associated with altered lectin sensitivity and occurrence in parasitic species. *J. Biol. Chem.* **287**, 28 276–28 290. (doi:10.1074/jbc.M112.353128)
21. Paschinger K, Hackl M, Gutternigg M, Kretschmer-Lubich D, Stemmer U, Jantsch V, Lochnit G, Wilson IBH. 2006 A deletion in the Golgi $\alpha$-mannosidase II gene of *Caenorhabditis elegans* results in unexpected non-wild-type N-glycan structures. *J. Biol. Chem.* **281**, 28 265–28 277. (doi:10.1074/jbc.M602878200)
22. Barrows BD, Haslam SM, Bischof LJ, Morris HR, Dell A, Aroian RV. 2007 Resistance to *Bacillus thuringiensis* toxin in *Caenorhabditis elegans* from loss of fucose. *J. Biol. Chem.* **282**, 3302–3311. (doi:10.1074/jbc.M606621200)
23. Geisler C, Kotu V, Sharrow M, Rendić D, Pöltl G, Tiemeyer M, Wilson IBH, Jarvis DL. 2012 The *Drosophila* neurally altered carbohydrate mutant has a defective Golgi GDP-fucose transporter. *J. Biol. Chem.* **287**, 29 599–29 609. (doi:10.1074/jbc.M112.379313)
24. Seeberger PH. 2017 Monosaccharide diversity. In *Essentials of glycobiology* (eds A Varki, RD Cummings, JD Esko), pp. 2015–2017, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
25. Zielinska DF, Gnad F, Wiśniewski JR, Mann M. 2010 Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907. (doi:10.1016/j.cell.2010.04.012)
26. Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M, Cox J. 2015 Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453–1456. (doi:10.1002/pmic.201400449)
27. Lauc G, Wuhrer M (eds) 2017 *High-throughput glycomics and glycoproteomics: methods and protocols, methods in molecular biology*, vol. 1503. Berlin, Germany: Springer Science + Business Media New York.
28. Hu Y, Zhou S, Yu C-Y, Tang H, Mechref Y. 2014 Automated annotation and quantitation of glycans by liquid chromatography/electrospray ionization mass spectrometric analysis using the MultiGlycan-ESI computational tool. *Rapid Commun. Mass Spectrom.* **29**, 135–142. (doi:10.1002/rcm.7093)
29. Meitei NS, Apte A, Snovida SI, Rogers JC, Saba J. 2015 Automating mass spectrometry-based quantitative glycomics using aminoxy tandem mass tag reagents with SimGlycan. *J. Proteomics* **127**, 211–222. (doi:10.1016/j.jprot.2015.05.015)
30. Damerell D, Ceroni A, Maass K, Ranzinger R, Dell A, Haslam SM. 2015 Annotation of glycomics MS and MS/MS spectra using the GlycoWorkbench software tool. In *Glycoinformatics* (eds T Lütteke, M Frank), pp. 3–15. New York, NY: Springer.
31. Katoh T, Tiemeyer M. 2013 The N's and O's of *Drosophila* glycoprotein glycobiology. *Glycoconjugate J.* **30**, 57–66. (doi:10.1007/s10719-012-9442-x)
32. Sharrow M, Aoki K, Baas S, Porterfield M, Tiemeyer M. 2010 Chapter 13—Genetic and structural analysis of the glycoprotein and glycolipid glycans of *Drosophila melanogaster*. In *Handbook of glycomics* (eds RD Cummings, JM Pierce), pp. 329–345. San Diego, CA: Academic Press.
33. Noborn F, Gomez Toledo A, Nasir W, Nilsson J, Dierker T, Kjellén L, Larson G. 2018 Expanding the chondroitin glycoproteome of *Caenorhabditis elegans*. *J. Biol. Chem.* **293**, 379–389. (doi:10.1074/jbc.M117.807800)
34. Paschinger K, Gutternigg M, Rendić D, Wilson IBH. 2008 The N-glycosylation pattern of *Caenorhabditis elegans*. *Carbohydr. Res.* **343**, 2041–2049. (doi:10.1016/j.carres.2007.12.018)
35. He L, Xin L, Shan B, Lajoie GA, Ma B. 2014 GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J. Proteome Res.* **13**, 3881–3895. (doi:10.1021/pr401115y)
36. Bern M, Kil YJ, Becker C. 2012 Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics* **40**, 13.20.11–13.20.14. (doi:10.1002/0471250953.bi1320s40)
37. Cheng K, Chen R, Seebun D, Ye M, Figeys D, Zou H. 2014 Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J. Proteomics* **110**, 145–154. (doi:10.1016/j.jprot.2014.08.006)
38. Aoki K, Porterfield M, Lee SS, Dong B, Nguyen K, McGlamry KH, Tiemeyer M. 2008 The diversity of O-linked glycans expressed during *Drosophila melanogaster* development reflects stage- and tissue-specific requirements for cell signaling. *J. Biol. Chem.* **283**, 30 385–30 400. (doi:10.1074/jbc.M804925200)
39. Patsos G, Corfield A. 2009 O-glycosylation: structural diversity and functions. In *The sugar code* (ed. HJ Gabius), pp. 111–137. Weinheim, Germany: Wiley-VCH.
40. Tran DT, Zhang L, Zhang Y, Tian E, Earl LA, Ten Hagen KG. 2012 Multiple members of the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase family are essential for viability in *Drosophila*. *J. Biol. Chem.* **287**, 5243–5252. (doi:10.1074/jbc.M111.306159)
41. Guérardel Y, Balanzino L, Maes E, Leroy Y, Coddeville B, Oriol R, Strecker G. 2001 The nematode *Caenorhabditis elegans* synthesizes unusual O-linked glycans: identification of glucose-substituted mucin-type O-glycans and short chondroitin-like oligosaccharides. *Biochem. J.* **357**, 167–182. (doi:10.1042/bj3570167)
42. Maes E, Garénaux E, Strecker G, Leroy Y, Wieruszeski J-M, Brassart C, Guérardel Y. 2005 Major O-glycans from the nest of *Vespula germanica* contain phospho-ethanolamine. *Carbohydr. Res.* **340**, 1852–1858. (doi:10.1016/j.carres.2005.05.008)
43. Walski T, De Schutter K, Van Damme EJM, Smagghe G. 2017 Diversity and functions of protein glycosylation in insects. *Insect Biochem. Mol. Biol.* **83**, 21–34. (doi:10.1016/j.ibmb.2017.02.005)
44. Okajima T, Irvine KD. 2002 Regulation of Notch signaling by O-linked fucose. *Cell* **111**, 893–904. (doi:10.1016/S0092-8674(02)01114-5)
45. Sasamura T *et al.* 2003 *Neurotic*, a novel maternal neurogenic gene, encodes an O-fucosyltransferase that is essential for Notch-Delta interactions. *Development* **130**, 4785–4795. (doi:10.1242/dev.00679)
46. Dominguez M, Ferres-Marco D, Gutierrez-Aviño FJ, Speicher SA, Beneyto M. 2003 Growth and specification of the eye are controlled independently by eyegone and eyeless in *Drosophila melanogaster*. *Nat. Genet.* **36**, 31–39. (doi:10.1038/ng1281)
47. Dönitz J, Schmitt-Engel C, Grossmann D, Gerischer L, Tech M, Schoppmeier M, Klingler M, Bucher G. 2015 iBeetle-Base: a database for RNAi phenotypes in the red flour beetle *Tribolium castaneum*. *Nucleic Acids Res.* **43**, D720–D725. (doi:10.1093/nar/gku1054)
48. Acar M, Jafar-Nejad H, Takeuchi H, Rajan A, Ibrani D, Rana NA, Pan H, Haltiwanger RS, Bellen HJ. 2008 Rumi, a CAP10 domain protein, is a glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* **132**, 247–258. (doi:10.1016/j.cell.2007.12.016)
49. Lee TV, Sethi MK, Leonardi J, Rana NA, Buettner FFR, Haltiwanger RS, Bakker H, Jafar-Nejad H. 2013 Negative regulation of Notch signaling by xylose. *PLoS Genet.* **9**, e1003547. (doi:10.1371/journal.pgen.1003547)
50. Harvey BM, Rana NA, Moss H, Leonardi J, Jafar-Nejad H, Haltiwanger RS. 2016 Mapping sites of O-glycosylation and fringe elongation on *Drosophila* Notch. *J. Biol. Chem.* **291**, 16 348–16 360. (doi:10.1074/jbc.M116.732537)
51. Ishio A *et al.* 2015 O-fucose monosaccharide of *Drosophila* Notch has a temperature-sensitive function and cooperates with O-glucose glycan in Notch transport and notch signaling activation. *J. Biol. Chem.* **290**, 505–519. (doi:10.1074/jbc.M114.616847)
52. Xu C, Ng DTW. 2015 O-mannosylation: the other glycan player of ER quality control. *Semin. Cell Dev. Biol.* **41**, 129–134. (doi:10.1016/j.semcdb.2015.01.014)
53. Ichimiya T, Manya H, Ohmae Y, Yoshida H, Takahashi K, Ueda R, Endo T, Nishihara S. 2004 The twisted abdomen phenotype of *Drosophila* POMT1 and POMT2 mutants coincides with their heterophilic protein O-mannosyltransferase activity. *J. Biol. Chem.* **279**, 42 638–42 647. (doi:10.1074/jbc.M404900200)
54. Larsen ISB *et al.* 2017 Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc. Natl Acad. Sci. USA* **114**, 11 163–11 168. (doi:10.1073/pnas.1708319114)
55. Wairkar YP, Fradkin LG, Noordermeer JN, DiAntonio A. 2008 Synaptic defects in a *Drosophila* model of congenital muscular dystrophy. *J. Neurosci.* **28**, 3781–3789. (doi:10.1523/jneurosci.0478-08.2008)
56. Nakamura N, Lyalin D, Panin VM. 2010 Protein O-mannosylation in animal development and

physiology: from human disorders to *Drosophila* phenotypes. *Semin. Cell Dev. Biol.* **21**, 622–630. (doi:10.1016/j.semcdb.2010.03.010)

57. Zhu Y, Shan X, Yuzwa SA, Vocadlo DJ. 2014 The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **289**, 34 472–34 481. (doi:10.1074/jbc.R114.601351)
58. Ma Z, Vosseller K. 2014 Cancer metabolism and elevated O-GlcNAc in oncogenic signaling. *J. Biol. Chem.* **289**, 34 457–34 465. (doi:10.1074/jbc.R114.577718)
59. Vaidyanathan K, Wells L. 2014 Multiple tissue-specific roles for the O-GlcNAc post-translational modification in the induction of and complications arising from type II diabetes. *J. Biol. Chem.* **289**, 34 466–34 471. (doi:10.1074/jbc.R114.591560)
60. Andrali SS, Qian Q, Özcan S. 2007 Glucose mediates the translocation of neurod1 by O-linked glycosylation. *J. Biol. Chem.* **282**, 15 589–15 596. (doi:10.1074/jbc.M701762200)
61. Hakomori S-i, Igarashi Y. 1995 Functional role of glycosphingolipids in cell recognition and signaling. *J. Biochem.* **118**, 1091–1103. (doi:10.1093/oxfordjournals.jbchem.a124992)
62. Chen Y-W, Pedersen JW, Wandall HH, Levery SB, Pizette S, Clausen H, Cohen SM. 2007 Glycosphingolipids with extended sugar chain have specialized functions in development and behavior of *Drosophila*. *Dev. Biol.* **306**, 736–749. (doi:10.1016/j.ydbio.2007.04.013)
63. Dahlgaard K, Jung A, Qvortrup K, Clausen H, Kjaerulff O, Wandall HH. 2012 Neurofibromatosis-like phenotype in *Drosophila* caused by lack of glucosylceramide extension. *Proc. Natl Acad. Sci. USA* **109**, 201115453.
64. Iozzo RV, Schaefer L. 2015 Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. *Matrix Biol.* **42**, 11–55. (doi:10.1016/j.matbio.2015.02.003)
65. Baron MJ, Wong SL, Nybakken K, Carey VJ, Madoff LC. 2009 Host glycosaminoglycan confers susceptibility to bacterial infection in *Drosophila melanogaster*. *Infect. Immun.* **77**, 860–866. (doi:10.1128/iai.00995-08)
66. Park Y, Rangel C, Reynolds MM, Caldwell MC, Johns M, Nayak M, Welsh CJR, McDermott S, Datta S. 2003 *Drosophila* Perlecan modulates FGF and Hedgehog signals to activate neural stem cell division. *Dev. Biol.* **253**, 247–257. (doi:10.1016/S0012-1606(02)00019-2)
67. Voigt A, Pflanz R, Schäfer U, Jäckle H. 2002 Perlecan participates in proliferation activation of quiescent *Drosophila* neuroblasts. *Dev. Dyn.* **224**, 403–412. (doi:10.1002/dvdy.10120)
68. Zachara N, Akimoto Y, Hart GW. 2017 The O-GlcNAc modification. In *Essentials of glycobiology* (eds A Varki, RD Cummings, JD Esko), pp. 2015–2017, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
69. Varshney S, Stanley P. 2017 EOGT and O-GlcNAc on secreted and membrane proteins. *Biochem. Soc. Trans.* **45**, 401–408. (doi:10.1042/BST20160165)
70. Hardivillé S, Hart GW. 2014 Nutrient regulation of signaling, transcription, and cell physiology by O-GlcNAcylation. *Cell Metab.* **20**, 208–213. (doi:10.1016/j.cmet.2014.07.014)
71. Rahman MM, Stuchlick O, El-Karim EG, Stuart R, Kipreos ET, Wells L. 2010 Intracellular protein glycosylation modulates insulin mediated lifespan in *C. elegans*. *Aging* **2**, 678–690. (doi:10.18632/aging.100208)
72. Mariappa D, Selvan N, Borodkin V, Alonso J, Ferenbach AT, Shepherd C, Navratilova IH, vanAalten DMF. 2015 A mutant O-GlcNAcase as a probe to reveal global dynamics of the *Drosophila* O-GlcNAc developmental proteome. *Biochem. J.* **470**, 255–262. (doi:10.1042/BJ20150610)
73. Ingham PW. 1984 A gene that regulates the bithorax complex differentially in larval and adult cells of *Drosophila*. *Cell* **37**, 815–823. (doi:10.1016/0092-8674(84)90416-1)
74. Vosseller K *et al.* 2006 O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* **5**, 923–934. (doi:10.1074/mcp.T500040-MCP200)
75. Trinidad JC, Barkan DT, Gulledge BF, Thalhammer A, Sali A, Schoepfer R, Burlingame AL. 2012 Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol. Cell. Proteomics* **11**, 215–229. (doi:10.1074/mcp.0112.018366)
76. Ohn T, Kedersha N, Hickman T, Tisdale S, Anderson P. 2008 A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. *Nat. Cell Biol.* **10**, 1224–1231. (doi:10.1038/ncb1783)
77. Zachara NE, Molina H, Wong KY, Pandey A, Hart GW. 2011 The dynamic stress-induced 'O-GlcNAc-ome' highlights functions for O-GlcNAc in regulating DNA damage/repair and other cellular pathways. *Amino Acids* **40**, 793–808. (doi:10.1007/s00726-010-0695-z)
78. Zhao P, Viner R, Teo CF, Boons G-J, Horn D, Wells L. 2011 Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome Res.* **10**, 4088–4104. (doi:10.1021/pr2002726)
79. Qin W, Lv P, Fan X, Quan B, Zhu Y, Qin K, Chen Y, Wang C, Chen X. 2017 Quantitative time-resolved chemoproteomics reveals that stable O-GlcNAc regulates box C/D snoRNP biogenesis. *Proc. Natl Acad. Sci. USA* **114**, E6749–E6758. (doi:10.1073/pnas.1702688114).
80. Vercoutter-Edouart A-S *et al.* 2014 Detection and identification of O-GlcNAcylated proteins by proteomic approaches. *Proteomics* **15**, 1039–1050. (doi:10.1002/pmic.201400326)
81. Abeytunga DTU, Oland L, Somogyi A, Polt R. 2008 Structural studies on the neutral glycosphingolipids of *Manduca sexta*. *Bioorg. Chem.* **36**, 70–76. (doi:10.1016/j.bioorg.2007.10.002)
82. Seppo A, Tiemeyer M. 2000 Function and structure of *Drosophila glycans*. *Glycobiology* **10**, 751–760. (doi:10.1093/glycob/10.8.751)
83. Dennis RD, Geyer R, Egge H, Menges H, Stirm S, Wiegandt H. 1985 Glycosphingolipids in insects. *Eur. J. Biochem.* **146**, 51–58. (doi:10.1111/j.1432-1033.1985.tb08618.x)
84. Satake M, Miyamoto E. 2012 A group of glycosphingolipids found in an invertebrate: their structures and biological significance. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **88**, 509–517. (doi:10.2183/pjab.88.509)
85. Seppo A, Moreland M, Schweingruber H, Tiemeyer M. 2000 Zwitterionic and acidic glycosphingolipids of the *Drosophila melanogaster* embryo. *Eur. J. Biochem.* **267**, 3549–3558. (doi:10.1046/j.1432-1327.2000.01383.x)
86. Gerdt S, Dennis RD, Borgonie G, Schnabel R, Geyer R. 2001 Isolation, characterization and immunolocalization of phosphorylcholine-substituted glycolipids in developmental stages of *Caenorhabditis elegans*. *Eur. J. Biochem.* **266**, 952–963. (doi:10.1046/j.1432-1327.1999.00937.x)
87. Itonori S, Hashimoto K, Nakagawa M, Harada M, Suzuki T, Kojima H, Ito M, Sugita M. 2018 Structural analysis of neutral glycosphingolipids from the silkworm *Bombyx mori* and the difference in ceramide composition between larvae and pupae. *J. Biochem.* **163**, 201–214. (doi:10.1093/jb/mvx072)
88. Lochnit G, Dennis RD, Ulmer AJ, Geyer R. 1998 Structural elucidation and monokine-inducing activity of two biologically active zwitterionic glycosphingolipids derived from the porcine parasitic nematode *Ascaris suum*. *J. Biol. Chem.* **273**, 466–474. (doi:10.1074/jbc.273.1.466)
89. Paschinger K, Gonzalez-Sapienza GG, Wilson IBH. 2012 Mass spectrometric analysis of the immunodominant glycan epitope of *Echinococcus granulosus* antigen Ag5. *Int. J. Parasitol.* **42**, 279–285. (doi:10.1016/j.ijpara.2012.01.002)
90. van Die I, van Stijn CMW, Geyer H, Geyer R. 2010 Chapter six—Structural and functional analysis of glycosphingolipids of *Schistosoma mansoni*. In *Methods in enzymology*, vol. 480 (ed. M Fukuda), pp. 117–140. New York, NY: Academic Press.
91. Vukelić Ž, Zamfir AD, Bindila L, Froesch M, Peter-Katalinić J, Usuki S, Yu RK. 2005 Screening and sequencing of complex sialylated and sulfated glycosphingolipid mixtures by negative ion electrospray Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **16**, 571–580. (doi:10.1016/j.jasms.2005.01.013)
92. Rožman M, Fabris D, Mrla T, Vukelić Ž. 2014 Database and data analysis application for structural characterization of gangliosides and sulfated glycosphingolipids by negative ion mass spectrometry. *Carbohydr. Res.* **400**, 1–8. (doi:10.1016/j.carres.2014.06.029)
93. Lawrence R, Olson SK, Steele RE, Wang LC, Warrior R, Cummings RD, Esko JD. 2008 Evolutionary differences in glycosaminoglycan fine structure detected by quantitative glycan reductive isotope labeling. *J. Biol. Chem.* **283**, 33 674–33 684. (doi:10.1074/jbc.M804288200)

94. Yamada S, Sugahara K, Özbek S. 2011 Evolution of glycosaminoglycans: comparative biochemical study. *Commun. Integr. Biol.* **4**, 150–158. (doi:10.4161/cib.4.2.14547)
95. Dierker T, Shao C, Haitina T, Zaia J, Hinas A, Kjellén L. 2016 Nematodes join the family of chondroitin sulfate-synthesizing organisms: identification of an active chondroitin sulfotransferase in *Caenorhabditis elegans*. *Sci. Rep.* **6**, 34662. (doi:10.1038/srep34662)
96. Olson SK, Bishop JR, Yates JR, Oegema K, Esko JD. 2006 Identification of novel chondroitin proteoglycans in *Caenorhabditis elegans*: embryonic cell division depends on CPG-1 and CPG-2. *J. Cell Biol.* **173**, 985–994. (doi:10.1083/jcb.200603003)
97. Li L, Ly M, Linhardt RJ. 2012 Proteoglycan sequence. *Mol. Biosyst.* **8**, 1613–1625. (doi:10.1039/c2mb25021g)
98. Hwang H-Y, Olson SK, Brown JR, Esko JD, Horvitz HR. 2003 The *Caenorhabditis elegans* genes sqv-2 and sqv-6, which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase. *J. Biol. Chem.* **278**, 11 735–11 738. (doi:10.1074/jbc.C200518200)
99. Vanbeselaere J, Yan S, Joachim A, Paschinger K, Wilson IBH. 2018 The parasitic nematode *Oesophagostomum dentatum* synthesizes unusual glycosaminoglycan-like O-glycans. *Glycobiology* **28**, 474–481. (doi:10.1093/glycob/cwy045)
100. Toyoda H, Kinoshita-Toyoda A, Selleck SB. 2000 Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that tout-velu, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J. Biol. Chem.* **275**, 2269–2275. (doi:10.1074/jbc.275.4.2269)
101. Fongmoon D, Shetty AK, Basappa YS, Sugiura M, Kongtawelert P, Sugahara K. 2007 Chondroitinase-mediated degradation of rare 3-O-sulfated glucuronic acid in functional oversulfated chondroitin sulfate K and E. *J. Biol. Chem.* **282**, 36 895–36 904. (doi:10.1074/jbc.M707082200).
102. Wilson IBH. 2002 Functional characterization of *Drosophila melanogaster* peptide O-xylosyltransferase, the key enzyme for proteoglycan chain initiation and member of the core 2/I N-acetylglucosaminyltransferase family. *J. Biol. Chem.* **277**, 21 207–21 212. (doi:10.1074/jbc.M201634200)
103. Vadaie N, Hulinsky RS, Jarvis DL. 2002 Identification and characterization of a *Drosophila melanogaster* ortholog of human β1,4-galactosyltransferase VII. *Glycobiology* **12**, 589–597. (doi:10.1093/glycob/cwf074)
104. Ahn J, Lüdecke H-J, Lindow S, Horton WA, Lee B, Wagner MJ, Horsthemke B, Wells DE. 1995 Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nat. Genet.* **11**, 137. (doi:10.1038/ng1095-137)
105. Häcker U, Nybakken K, Perrimon N. 2005 Heparan sulphate proteoglycans: the sweet side of development. *Nat. Rev. Mol. Cell Biol.* **6**, 530. (doi:10.1038/nrm1681))
106. Langeslay DJ, Jones CJ, Beni S, Larive CK. 2012 Glycosaminoglycans: oligosaccharide analysis by liquid chromatography, capillary electrophoresis, and specific labeling. In *Proteoglycans. Methods in molecular biology (methods and protocols)*, vol. 836 (ed. F Rédini). Totowa, NJ: Humana Press.
107. Staples GO, Zaia J. 2011 Analysis of glycosaminoglycans using mass spectrometry. *Curr. Proteomics.* **8**, 325–336. (doi:10.2174/15701 64117982208)
108. Deakin JA, Lyon M. 2008 A simplified and sensitive fluorescent method for disaccharide analysis of both heparan sulfate and chondroitin/dermatan sulfates from biological samples. *Glycobiology* **18**, 483–491. (doi:10.1093/glycob/cwn028)
109. Ly M, Leach Iii FE, Laremore TN, Toida T, Amster IJ, Linhardt RJ. 2011 The proteoglycan bikunin has a defined sequence. *Nat. Chem. Biol.* **7**, 827. (doi:10.1038/nchembio.673)
110. Hogan JD, Klein JA, Wu J, Chopra P, Boons G-J, Carvalho L, Lin C, Zaia J. 2018 Software for peak finding and elemental composition assignment for glycosaminoglycan tandem mass spectra. *Mol. Cell. Proteomics* **17**, 1448–1456. (doi:10.1074/mcp.RA118.000590)
111. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. 2001 Glycosylation and the immune system. *Science* **291**, 2370–2376. (doi:10.1126/science.291.5512.2370)
112. Bertozzi CR, Kiessling LL. 2001 Chemical glycobiology. *Science* **291**, 2357–2364. (doi:10.1126/science.1059820)
113. Wada Y *et al.* 2007 Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17**, 411–422. (doi:10.1093/glycob/cwl086)
114. Stanta JL *et al.* 2010 Identification of N-glycosylation changes in the CSF and serum in patients with schizophrenia. *J. Proteome Res.* **9**, 4476–4489. (doi:10.1021/pr1002356)
115. Cipollo JF, Awad AM, Costello CE, Hirschberg CB. 2004 srf-3, a mutant of *Caenorhabditis elegans*, resistant to bacterial infection and to biofilm binding, is deficient in glycoconjugates. *J. Biol. Chem.* **279**, 52 893–52 903. (doi:10.1074/jbc.M409557200)
116. Soya S, Şahar U, Karaçalı S. 2016 Monosaccharide profiling of silkworm (*Bombyx mori* L.) nervous system during development and aging. *Invert. Neurosci.* **16**, 8. (doi:10.1007/s10158-016-0191-6)