Expanding the Chondroitin Sulfate Glycoproteome — But How Far?

Fredrik Noborn[1]*, Mahnaz Nikpour[1], Andrea Persson[1], Jonas Nilsson[1,2] and Göran Larson[1]

[1] Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden, [2] Proteomics Core Facility, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

Chondroitin sulfate proteoglycans (CSPGs) are found at cell surfaces and in connective tissues, where they interact with a multitude of proteins involved in various pathophysiological processes. From a methodological perspective, the identification of CSPGs is challenging, as the identification requires the combined sequencing of specific core proteins, together with the characterization of the CS polysaccharide modification(s). According to the current notion of CSPGs, they are often considered in relation to a functional role in which a given proteoglycan regulates a specific function in cellular physiology. Recent advances in glycoproteomic methods have, however, enabled the identification of numerous novel chondroitin sulfate core proteins, and their glycosaminoglycan attachment sites, in humans and in various animal models. In addition, these methods have revealed unexpected structural complexity even in the linkage regions. These findings indicate that the number and structural complexity of CSPGs are much greater than previously perceived. In light of these findings, the prospect of finding additional CSPGs, using improved methods for structural and functional characterizations, and studying novel sample matrices in humans and in animal models is discussed. Further, as many of the novel CSPGs are found in low abundance and with not yet assigned functions, these findings may challenge the traditional notion of defining proteoglycans. Therefore, the concept of proteoglycans is considered, discussing whether "a proteoglycan" should be defined mainly on the basis of an assigned function or on the structural evidence of its existence.

Keywords: proteoglycans, glycosaminoglycans, chondroitin sulfate, core proteins, glycoproteomics, attachment site

INTRODUCTION

The concept of chondroitin sulfate proteoglycans (CSPGs) as discrete molecular entities first emerged during the late 1950s. At that time, chondroitin sulfate (CS) and protein complexes had been identified in hyaline cartilage, but the nature of the complexes remained elusive, and it was unclear whether the CS to protein association involved a covalent bond or not (Shatton and Schubert, 1954). In a pioneering study by Helen Muir, she showed that the CS chains were indeed covalently linked to the protein counterpart (Muir, 1958). A few years later, it was found that the CS polysaccharides are attached to serine residues of the core proteins via a tetrasaccharide linkage region, composed of Glucuronic acid (GlcA)—Galactose (Gal)—Galactose (Gal)—Xylose (Xyl) (Lindahl and Roden, 1966; Roden and Smith, 1966). Since then, an increasing number of different

CSPGs have been identified, each distinguished by the primary sequence of the core protein and with different numbers of CS chains attached (Olson et al., 2006; Iozzo and Schaefer, 2015; Pomin and Mulloy, 2018; Toledo et al., 2020).

CSPGs are important components in connective tissue and fine-tunes a wide range of cellular processes, including neural development, growth factor signaling and inflammation (Hatano and Watanabe, 2020; Hussein et al., 2020). However, the structural identification of CSPGs is often difficult, as the identification requires the combined sequencing of specific core proteins together with the structural verification of any potential CS polysaccharides. Consequently, studies on identifying novel CSPGs, using earlier established biochemical techniques, were mostly focused on the characterization of single core proteins in a defined cellular or physiological context (Krusius and Ruoslahti, 1986; Fisher et al., 1989). Such techniques are typically based on different read out assays following enzymatic depolymerization with bacterial lyases and/or site directed mutagenesis of cloned proteins.

The lack of suitable analytical methods for large-scale analyses of CSPGs in biological samples has for a long time limited the ability to identify both novel CSPGs and assess the degree of heterogeneity of the CS glycoproteome across various systems.

This mini review will specifically focus on the recent advances in glycoproteomics to identify and characterize CSPGs in complex sample mixtures in humans and animal model systems. These methods include trypsin digestion, enrichment of acidic glycopeptides by strong anion-exchange chromatography (SAX), and incubation with chondroitinase ABC to reduce the length and complexity of the CS chains. The samples are then analyzed with reversed phase nano-liquid chromatography-tandem mass spectrometry (nLC-MS/MS) and evaluated by glycopeptide search algorithms, resulting in the discovery and characterization of several novel CSPGs both in vertebrates and invertebrates. In light of these findings, the structural and conceptual insights that can be provided by such attachment site-specific analysis of the CS glycoproteome are discussed. Further, as many of the novel CSPGs are found in low abundance and with no yet assigned functions, these findings may challenge the traditional notion of proteoglycans, in which the described proteoglycans often have an assigned function. Finally, the concept of the CS glycoproteome is discussed in relation to whether the glycoproteomic space should be regarded as a static or a dynamic entity.

A GLYCOPROTEOMIC APPROACH TO IDENTIFY NOVEL PROTEOGLYCANS

In glycobiology, mass spectrometry-based strategies for investigating protein glycosylation (glycoproteomics) have become an increasingly important tool (Nilsson et al., 2009; Thaysen-Andersen et al., 2016; Narimatsu et al., 2018). Such strategies, well-covered in excellent reviews, are typically based on enrichment of glycopeptides and subsequent analysis with nLC-MS/MS to provide site-specific information of N- and O-glycans (Nilsson, 2016; Darula and Medzihradszky, 2018;

Madsen et al., 2020; Chernykh et al., 2021). To further develop this concept, we established a glycoproteomic protocol for global characterization of CS-glycopeptides in human urine and cerebrospinal fluid (CSF) (Noborn et al., 2015; Noborn et al., 2021). At first, bikunin (also known as protein AMBP) was used as a model CSPG since it was relatively well characterized, available in large quantities from human urine and also used in some countries as a pharmaceutical agent to treat acute pancreatitis (Ly et al., 2011; Lord et al., 2020). Human bikunin has a single CS chain of 27–39 monosaccharides attached to the N-terminal end (Ser-10) of the core protein (Lord et al., 2013). We incubated pharmaceutical grade bikunin with chondroitinase ABC that generated free disaccharides and a residual hexasaccharide structure still attached to the core protein. Digestion with trypsin generated a defined CS-glycopeptide suitable in size for nLC-MS/MS analysis. The analysis enabled the identification of several specific glycosidic and bikunin peptide fragments, serving as a proof-of-concept that site-specific analysis of CSPGs is indeed a feasible strategy. We then enriched trypsin-digested CSPGs from human urine and CSF and incubated the enriched samples with chondroitinase ABC, and thereafter analyzed the resulting CS-glycopeptides by nLC-MS/MS. Generated data were evaluated through proteomic software with adjustments to allow for glycopeptide identification, enabling the identification of 13 novel human CSPGs in addition to 13 already established CSPGs. In **Tables 1A,B**, the novel proteoglycans identified in humans and different model systems through glycoproteomics are shown. Interestingly, several of the novel human CSPGs were traditionally defined as prohormones, which was surprising, as prohormones are not typically regarded to belong to the proteoglycan superfamily (Noborn et al., 2015). This approach demonstrates the structural and conceptual insights that can be provided by global attachment site-specific analysis of the CS glycoproteome.

A similar strategy was developed for complex proteoglycans of the extracellular matrix (hyalectans) from human and bovine sources (Klein et al., 2018). In addition to CS chains, hyalectan proteoglycans are also rich N- and mucin-type O-glycosylations that decorate large parts of the proteins. The abundant glycosylation impairs effective peptide identification and reduces the overall sequence coverage achieved by mass-spectrometric based methods. By using a combination of efficient enrichment procedures for glycosylated peptides and advanced MS/MS-software analysis, the authors were able to improve the peptide sequence coverage of these complex proteoglycans and identify a number of new attachment sites, including both N-, mucin type O- and CS-glycosylations.

GLYCOPROTEOMICS: NOVEL CSPGs LEADING TO NOVEL HYPOTHESES

Although glycoproteomic approaches assist in identifying novel CSPGs, the functional relevance of a CS modification on cellular physiology is often unclear and cannot be inferred from only the CS structure or from the core protein alone.

TABLE 1 | A. Novel human proteoglycans identified through glycoproteomics.

Human

Protein name	Uniprot ID	Modification[a]	Site(s)[b]	References
Brain-specific angiogenesis inhibitor 2	O60241	CS	266*	Noborn et al., 2015
CD99L2 protein	H0Y4H3	CS	141*	Noborn et al., 2015
Cholecystokinin	P06307	CS	31	Noborn et al., 2015
Collagen and calcium-binding EGF domain containing protein 1	Q6UXH8	CS	385	Noborn et al., 2015
Dermcidin	P81605	CS	30*	Noborn et al., 2015
Laminin subunit alpha-4	Q16363	CS	40*	Noborn et al., 2015
Laminin subunit gamma 2	Q13753	CS	803*	Toledo et al., 2020
Matrix-remodeling associated protein 5	Q9NR99	CS	702*	Noborn et al., 2015
Meprin A subunit alpha	Q16819	CS	631*	Nasir et al., 2016
Natriuretic peptides B	P16860	CS	41*	Toledo et al., 2020
Neurexin-1	Q9ULB1	HS	1,355	Zhang et al., 2018
Neurexin-2	Q9P2S2	HS	1,400	Zhang et al., 2018
Neurexin-3	Q9Y4C0	HS	1,315	Zhang et al., 2018
Neuropeptide W	Q8N729	CS	133*	Noborn et al., 2015
Neuroserpin	Q99574	CS	403	Noborn et al., 2015
Nidogen-2	Q14112	CS	452*, 358*	Toledo et al., 2020
Osteopontin	P10451	CS	234*, 308*	Noborn et al., 2015
Plexin domain-containing protein 1	Q8IUK5	CS	33	Noborn et al., 2015
Retinoic acid responder protein	P49788	CS	40	Nasir et al., 2016
Secretogranin-1	P05060	CS	93*, 239*	Noborn et al., 2015
Secretogranin-3	Q8WXD2	CS	37	Noborn et al., 2015
Sushi repeat-containing protein SRPX	P78539	CS	34*	Toledo et al., 2020

B. Novel proteoglycans identified in different model systems through glycoproteomics.

Protein name	Uniprot ID/Accession number	Modification[b]	Site(s)[b]	References
Caenorabditis elegans				
CLE-1A protein/CPG-10	Q9U9K7	Chn	581	Noborn et al., 2018
COLlagen/CPG-11	Q22651	Chn	336*	Noborn et al., 2018
C-type lectin domain containing protein 180/CPG-12	Q19970	Chn	275, 290	Noborn et al., 2018
Dauer Up-Regulated, isoform b/CPG-13	Q7JLY2	Chn	368	Noborn et al., 2018
High Incidence of Males, isoform b/CPG-14	NCBINP_001024582.1	Chn	4,850*	Noborn et al., 2018
LiPocalin-Related protein/CPG-15	Q23163	Chn	172*	Noborn et al., 2018
FiBrilliN homolog/CPG-16	Q23587	Chn	1,079*	Noborn et al., 2018
Papilin/CPG-17	O76840	Chn	775*	Noborn et al., 2018
Protein C45E5.4/CPG-18	Q18642	Chn	38*	Noborn et al., 2018
Protein C06G1.2/CPG-19	Q17742	Chn	35*	Noborn et al., 2018
Protein K08B4.2/CPG-20	Q9TYY6	Chn	167*	Noborn et al., 2018
Protein R17.3/CPG-21	O18003	Chn	111*	Noborn et al., 2018
Protein T10E9.3/CPG-22	O01603	Chn	382, 399	Noborn et al., 2018
Protein Y39B6B.y/CPG-23	NCBI PIR T45051	Chn	313*	Noborn et al., 2018
Protein Y41D4B.26/CPG-24	Q8WSN8	Chn	194	Noborn et al., 2018
Danio rerio				
Uncharacterized protein si:ch73-306e8.2 isoform X1	XP_00133717.3	CS	103*	Delbaere et al., 2020
Drosophila melanogaster				
Protein Windpipe	Q9W266	CS	282, 335, 337	Takemura et al., 2020
Rat INS-1 832/13, murine MIN-6 and NIT-1 cell lines				
Chromogranin-A	P10354	HS	433	Nikpour et al., 2021
Immunoglobulin superfamily member 10	Q3V1M1	CS	679	Nikpour et al., 2021
Islet Amyloid Polypeptide	P12969	CS	28	Nikpour et al., 2021
Secretogranin-1	P35314	HS	236	Nikpour et al., 2021

[a]Glycosaminoglycan modifications: CS, chondroitin sulfate; HS, heparan sulfate; or Chn, chondroitin.
[b]Glycosaminoglycan attachment sites denote serine amino acid numbers, determined unambiguously by nLC-MS/MS or ambiguously for glycopeptides containing two or more serine residues (*). Numbers relate to the corresponding protein ID.

However, the identification of novel CSPGs, as well as defining novel attachment site modifications of established CSPGs, generate novel hypotheses that can be experimentally tested. Here we exemplify with a few cases of how such global or targeted glycoproteomic approaches can be combined with hypothesis-driven research to investigate structure-function relationships of CSPGs.

We recently identified the Windpipe protein as a novel CSPGs in *Drosophila melanogaster* and found that the core protein carried three separate CS chains in its extracellular domain (Takemura et al., 2020). Genetic engineering combined with morphological analyses demonstrated that Windpipe inhibited Hedgehog (Hh) signaling in a CS dependent manner. Interestingly, Windpipe-overexpression resulted in reduced wing size compared with control flies. However, the wing size was restored in genetically modified flies in which all the three CS attachment serine residues were substituted with alanine residues (thereby precluding CS modification). This study thus demonstrates a novel role of a specific CSPG in regulating Hh signaling and illustrate the potential of combining glycoproteomics with molecular and cell biological studies in a well-defined model system.

Several protein hormones are stored as amyloid-like aggregates in the secretory granules of endocrine cells (Maji et al., 2009). Most of the protein hormones require mild acidic pH and the addition of low molecular weight heparin or CS for their aggregation *in vitro*. Our finding that several prohormones carry CS chains lead us to hypothesize that the CS side chains may facilitate self-assembly of the prohormones (Noborn et al., 2015). Indeed, binding studies showed that CS promoted the assembly of the chromogranin A core protein under acidic condition, giving a possible explanation to previous observations that chromogranin A has an inherent property to assemble in the acidic milieu of secretory granules. Whether the CS side chains of other prohormones (e.g., secretogranin-1 and 3) may also influence the assembly of their respective core protein remains to be determined.

Furthermore, an affinity-enriched glycoproteomic approach was recently employed to explore the involvement of CSPGs in the pathogenesis of pregnancy-associated malaria. The disease is caused by the parasite *Plasmodium falciparum* and has potentially very severe clinical outcome for both mother and child (Toledo et al., 2020; Tomlinson et al., 2020). The parasite induces expression of the malarial protein VAR2CSA on the surface of infected erythrocytes, which enables their binding to structural variants of CS in the placental intervillous space of pregnant women (Kane and Taylor-Robinson, 2011). This interaction is dependent on the size and sulfation-type of the polysaccharide, as longer chains with higher degree of C4-*O*- sulfation of GalNAc residues display stronger binding (Sugiura et al., 2016; Ma et al., 2021). A VAR2CSA-affinity column was recently used to enrich for CSPGs, capable of binding the VAR2CSA-protein, in human placenta. The enriched fraction was then analyzed through a CS glycoproteomic workflow, showing that a collection of different core proteins, rather than a single core protein, carried CS chains that were capable of binding the VAR2CSA-protein (Toledo et al., 2020). This indicates that several different CSPGs may serve as attachment factors for malaria-infected erythrocytes, which

provides new insight into the disease etiology. Notably, CS and heparan sulfate (HS) also constitute attachment factors for many enveloped and non-enveloped viruses (Olofsson and Bergström, 2005). Interestingly, recent studies show that cell surface HS act as a co-receptor to ACE2 and is essential for SARS-CoV-2 infection (Clausen et al., 2020; Mycroft-West et al., 2020). To our knowledge, however, the information of a potential HS core protein (or group of core proteins) involved in such processes is scarce (Daly et al., 2020). A glycoproteomic enrichment strategy, similar to the VAR2CSA-CS enrichment approach, may provide information on this issue.

PROTEOGLYCAN LINKAGE REGION COMPLEXITY

The CS linkage region has previously been perceived as a relatively uniform entity with limited structural variability. MS/MS-analysis of chondroitinase ABC-digested bikunin described a defined hexasaccharide structure with *O*-sulfations of the GalNAc residue and of the outer Gal residue (Ly et al., 2011). In addition to sulfation, other linkage region modifications have been reported, e.g., sialylation and phosphorylation (Sugahara et al., 1988; Kitagawa et al., 2008; Lu et al., 2010; Wen et al., 2014). However, the reports have been limited to only a few separate experimental systems and information on potential combinations of linkage region modifications is still scarce. To this aim, our glycoproteomic analysis of human bikunin and other proteoglycans from urine, CSF, and plasma have revealed an unexpected linkage region complexity, with different combinations of sulfation, phosphorylation and sialylation (Gomez Toledo et al., 2015; Noborn et al., 2015; Nasir et al., 2016; Nilsson et al., 2017). Furthermore, the structural variability of human bikunin was further increased by large variations in the mucin type *O*-linked glycosylation found nearby the CS site (Gomez Toledo et al., 2015). Moreover, an unexpected fucose (deoxy-hexose) modification was found on the xylose residue of the linkage region of human bikunin and decorin and on two novel human CSPGs; retinoic acid responder protein 1 and meprin A (Gomez Toledo et al., 2015; Nasir et al., 2016). This was surprising as fucosylated CS chains have previously only been described in sea cucumbers, a member of the phylum of Echinodermata (Myron et al., 2014; Dwivedi and Pomin, 2020). Notably, this fucose modification in human CS was located on xylose (and not on the CS chain), i.e., close to the protein component. This position is similar to the fucose modifications on *N*-glycans, which also may occur at the innermost GlcNAc residue linked to Asn in the consensus amino acid sequence (core fucosylation) (Plomp et al., 2017; Schneider et al., 2017; Wang and Ravetch, 2019). This kind of *N*-glycan fucosylation occurs both in invertebrate and vertebrate species and modifies the functional effects of the corresponding proteins, e.g., by regulating the inflammatory immune responses by modifying the *N*-glycans of the Fc chains of IgG1 molecules. This is of major importance when designing and introducing monoclonal antibodies for immune therapy, since the afucosylated variants show an increased antibody-dependent cellular cytotoxicity (ADCC) (Pereira et al., 2018). What functional effects the

core fucosylation of xylose of CSPG may have remains to be determined.

Furthermore, we have observed a non-canonical linkage region trisaccharide (GlcA-Gal-Xyl-*O*-) on glycopeptides of bikunin from urine of healthy donors. This trisaccharide thus lacks one Gal residue of the traditional linkage region, and results in a residual pentasaccharide structure upon chondroitinase ABC digestion (Persson et al., 2019). Interestingly, mutations in the *B3GALT6* gene, which codes for galactosyltransferase II (β3GalT6) is associated with Ehlers-Danlos–like syndromes, characterized by a spectrum of skeletal and connective tissue "linkeropathy" disorders (Malfait et al., 2013). The β3GalT6 enzyme is responsible for adding the second Gal residue in the GAG linkage region and mutations in the gene results in reduced GAG biosynthesis and appearance of the trisaccharide linkage region in a zebrafish *b3galt6* knock-out model (Delbaere et al., 2020). Whether patients with *B3GALT6* mutations also have increased levels of the non-canonical trisaccharide linkage region of bikunin in their urine remains to be determined. Taken together, at least 14 different variants of the bikunin linkage region (including canonical and non-canonical sequences) have now been described, thus demonstrating a much greater variability than previously appreciated (Gomez Toledo et al., 2015; Noborn et al., 2015; Nilsson et al., 2017; Persson et al., 2019). Xylose phosphorylation has been shown to function as a molecular switch to regulate the proteoglycan biosynthesis (Wen et al., 2014). The removal of xylose phosphorylation by 2-phosphoxylose phosphatase is considered a prerequisite for the polymerization of the linkage regions into longer chains (Koike et al., 2014). However, the influence of the other linkage modifications (or the combinations thereof) on the downstream GAG biosynthesis is yet unknown. Nevertheless, knowledge of linkage region complexity provides a theoretical framework for future functional-structural studies of CS biosynthesis. Moreover, analysis of extended site-specific CS chains, in theory achieved by using partial chondroitinase ABC digestion, may assist in further understanding of how these modifications affect the biosynthesis and thus the final structures of CS. In addition, this may provide information on how certain core protein determinants (e.g., amino acid sequences) may influence the CS structures. The concept of core protein specific structures have been difficult to address since most established proteoglycan workflows involve the separation of the CS polysaccharides and the core proteins. Ideally, site-specific analysis of full-length CS structures is desirable, as this type of information is likely required to fully explore the structural-functional basis of CSPGs.

Further development of the glycoproteomic concept has enabled site-specific characterization of both CS and HS sites (Noborn et al., 2016). To this aim, perlecan (also known as basement membrane specific proteoglycan) was chosen as a model proteoglycan since it is known to be substituted with both CS and HS chains (Iozzo, 2005; Noborn et al., 2015). A trypsin-digested perlecan sample, derived from Engelbreth-Holm-Swarm mouse sarcoma, was enriched on a SAX column and digested with heparinase or chondroitinase ABC, generating residual tetrasaccharide and hexasaccharide structures still attached to the peptide. This allowed for the identification of a glycopeptide derived from the N-terminal domain of perlecan, which encompassed the three previously known HS sites (Iozzo and Sanderson, 2011). In addition, the combination of chondroitinase and heparinase digestion revealed a "hybrid site" in the C-terminal domain, carrying either an HS or a CS chain (Noborn et al., 2016). This was surprising since it demonstrated that a single GAG attachment site is capable of carrying either HS or CS chains, thus, revealing a less appreciated level of proteoglycan heterogeneity. The identification of such "hybrid sites" is likely to be biologically important as HS and CS often display opposite effects on cellular physiology (Coles et al., 2011; Doody et al., 2015). Furthermore, similar glycoproteomic protocols have been used to identify novel HSPGs, and to investigate the functional role of HSPGs in neuronal synapse formation (Zhang et al., 2018; Montoliu-Gaya et al., 2021; Nikpour et al., 2021).

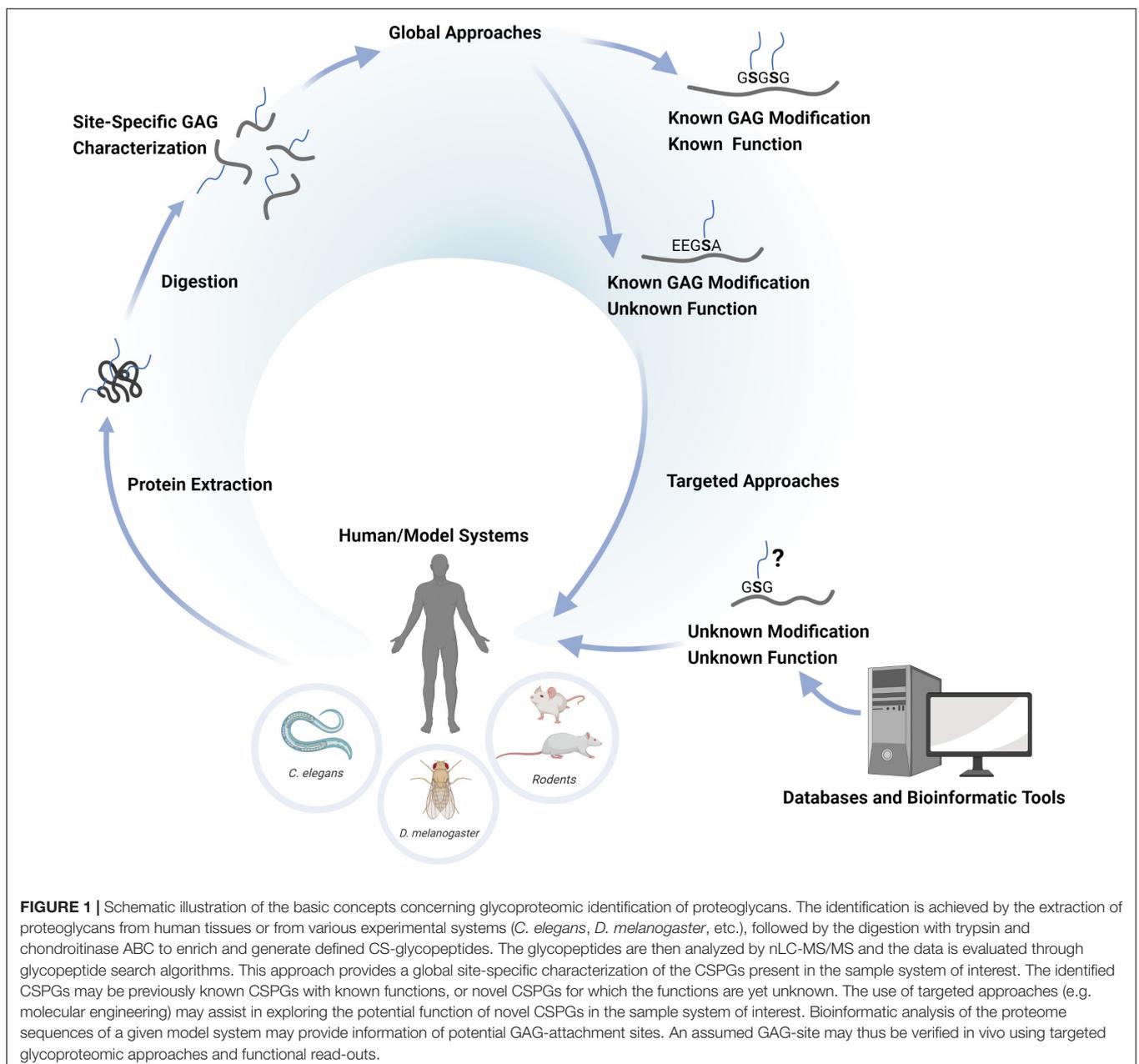# EXPANDING THE PROTEOGLYCAN GLYCOPROTEOME — BUT HOW FAR?

Glycosylation is one of the most ubiquitous forms of post-translational modification and is expected to be present on more than half of all mammalian proteins (Apweiler et al., 1999; Schjoldager et al., 2020). However, despite its high abundance, GAG modifications are considered to occur only on a very minor proportion of the 20,000–25,000 proteins encoded by the human genome. Although glycoproteomic studies of various sample matrices have significantly increased our structural knowledge of proteoglycans, fewer than 80 core proteins carrying CS have so far been identified in humans (Noborn et al., 2015; Nasir et al., 2016; Toledo et al., 2020). A list of all CSPGs known to date in humans was recently published by Toledo et al. (2020). The number of human core proteins carrying HS are even less, and so far only a little more than 20 HSPGs have been established (Zhang et al., 2018; Nikpour et al., 2021). Moreover, while proteoglycans in vertebrates have been the focus of several structural studies, the information about proteoglycans in invertebrates is scarce and the reports are restricted to only a few species (Noborn and Larson, 2021). Thus, only a limited number of CSPGs have been identified in *Drosophila melanogaster*, which is surprising since it is one of the most studied invertebrates (Momota et al., 2011; Zhu et al., 2019; Takemura et al., 2020). Future studies employing glycoproteomic approaches will likely identify novel CSPGs of invertebrates and further expand our knowledge of the proteoglycan glycoproteome in different species.

The enzymatic transfer of a Xyl residue to certain Ser residues in the proteoglycan core protein initiates GAG biosynthesis (Briggs and Hohenester, 2018). The Xyl-modified Ser residue is typically followed by a Gly residue (-SG-) and is associated with a cluster of acidic residues in close proximity, indicating that certain motifs of the amino acid sequence influence the initiation process (Esko and Zhang, 1996). Such motifs may assist in the prediction of potential GAG sites of core proteins and provides insights into the potential space of the GAG glycoproteome. When we recently mapped the chondroitin (Chn) glycoproteome in *C. elegans*, this resulted in the identification of 15 novel core proteins in addition to the 9 previously established core

proteins (Olson et al., 2006; Noborn et al., 2018). Bioinformatic analysis of protein sequences in the regions of the glycosylated Ser residues showed a highly stringent attachment motif E/D-G/A-S-G. A search in the Swiss-Prot database using this motif retrieved 19 additional potential CPGs, indicating that possibly additional CPGs are yet to be found in the nematode (Noborn et al., 2018). However, it is unclear why these assumed CPGs escaped detection by the glycoproteomic analysis. This may simply relate to the sensitivity of the assay and that the glycopeptides corresponding to these potential CPGs were below the detection threshold. On the other hand, it may also suggest that some motifs are only occasionally occupied, or substituted with HS instead of Chn. Nevertheless, this demonstrates the ambiguity of using only attachment motifs as a strategy to identify

proteoglycans, and that glycoproteomic analysis is needed to determine the GAG glycoproteome in any given sample system. The basic strategies for proteoglycan identification and the different levels of structural- and functional understanding of their roles in any biological system is schematically illustrated in **Figure 1**.

The prohormone chromogranin-A (CgA) is one of the main proteins in secretory granules of endocrine cells and the precursor for several bioactive peptides (Kim et al., 2001). CgA has previously been established as a CSPG and carries a CS chain in the C-terminal end (Ser-424) of the human core protein (Gowda et al., 1990; Noborn et al., 2015). Our recent glycoproteomic analysis of cultured insulin-secreting cells, demonstrate that CgA in rodents, but not in humans, also carries

FIGURE 1 | Schematic illustration of the basic concepts concerning glycoproteomic identification of proteoglycans. The identification is achieved by the extraction of proteoglycans from human tissues or from various experimental systems (*C. elegans*, *D. melanogaster*, etc.), followed by the digestion with trypsin and chondroitinase ABC to enrich and generate defined CS-glycopeptides. The glycopeptides are then analyzed by nLC-MS/MS and the data is evaluated through glycopeptide search algorithms. This approach provides a global site-specific characterization of the CSPGs present in the sample system of interest. The identified CSPGs may be previously known CSPGs with known functions, or novel CSPGs for which the functions are yet unknown. The use of targeted approaches (e.g. molecular engineering) may assist in exploring the potential function of novel CSPGs in the sample system of interest. Bioinformatic analysis of the proteome sequences of a given model system may provide information of potential GAG-attachment sites. An assumed GAG-site may thus be verified in vivo using targeted glycoproteomic approaches and functional read-outs.

HS at the same C-terminal site, making this another "hybrid site" (Noborn et al., 2016; Nikpour et al., 2021). Although the glycosite is highly conserved between species (Noborn et al., 2015), this highlights the importance of cell-targeted mapping and that GAG modifications found in one cell type or tissue or species may not simply be inferred to another sample system. These findings also support the view that the GAG glycoproteome should not be regarded as a static entity, but rather a dynamic system that changes with the cellular physiology. A key issue for the future is to comprehend the biological factors that eventually will govern the expansion of the GAG glycoproteome.

DISCUSSION

As with any research area, the characterization of any particular biomolecule is critical for the understanding of its role in biological and pathological processes. The identification and structural characterization of proteoglycans is no exception. According to our current understanding of CSPGs, these are often regarded in a functional context, in which any given proteoglycan regulate a specific function in cellular pathophysiology. For example, decorin, a small leucine-rich extracellular matrix proteoglycan, is well-studied for its role as a suppressor of tumor cell growth (Neill et al., 2016; Appunni et al., 2019). Decorin binds and antagonizes various receptor tyrosine kinases, thereby inhibiting downstream oncogenic signaling and reducing tumor growth (Reszegi et al., 2020; Hu X. et al., 2021). Apart from their involvement in cancer, CSPGs are also in focus of other research areas, including neurogenesis and spinal cord injury (Maeda, 2015). Following damage to the central nervous system, axons fail to regenerate due to the formation of a glial scar, which is composed of extracellular matrix components including CSPGs (Hussein et al., 2020). The production of several CSPG family members is differentially regulated in the glial scar as neurocan, brevican and versican are increased, while aggrecan is reduced (Mukhamedshina et al., 2019). How the individual CSPGs contribute to the pathogenesis is yet unclear, but their relevance is illustrated by the interest of using chondroitinase ABC as a therapeutic agent (Bradbury et al., 2002; Hu et al., 2018; Hu J. et al., 2021). Although detailed structural-functional understanding of PGs many times remains unknown, the examples given in this review point to essential roles of CSPGs in various pathophysiological conditions and demonstrate the significant progress that has been made in the field in recent years.

The recent advances in glycoproteomics have enabled the identification of several novel CS core proteins in humans and in various animal models. With such techniques, additional CS core proteins and novel GAG modifications will likely be discovered with no yet assigned function/s. Thus, one should consider whether "a proteoglycan" should be defined mainly on the basis of an assigned function or on structural evidence of its existence. Importantly, glycoproteomic strategies have the potential of finding many novel proteoglycans and provide global structural information that may contribute to our conceptual understanding of the complex family of proteoglycans. However, viewing proteoglycans from a "structure only" perspective will meet its limitations. The development of even more advanced glycoproteomics strategies will presumably identify proteoglycans at low abundances and/or at low levels of occupancy. This suggests that caution is needed in the interpretation of the data as minute amounts of any given proteoglycan may have little or no relevance for our understanding of any given biological system. Therefore, using glycoproteomics in a clearly defined biological or pathological context will likely be a necessary future strategy, since it will set constraints on the interpretation of the structural information derived. This will likely assist in determining which proteoglycans are relevant for understanding the underlying biological mechanisms in any given system. Nonetheless, glycoproteomic strategies will surely assist in further expanding the knowledge of proteoglycan core proteins and their functions. In the perspective of the present pandemic of COVID-19, the importance of proteoglycans as attachment factors/receptors or co-receptors of pathogenic viruses cannot be underestimated and detailed structural information of the host GAGs and their core protein structures will most likely be warranted to understand the molecular details of virus infections as well as to design novel anti-viral drugs.

AUTHOR CONTRIBUTIONS

FN and GL did the writing and provided the main concepts discussed in this mini review. MN, AP, and JN assisted in the organization and editing of the publication and contributed to the table and the figure. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Appunni, S., Anand, V., Khandelwal, M., Gupta, N., Rubens, M., and Sharma, A. (2019). Small Leucine Rich Proteoglycans (decorin, biglycan and lumican) in cancer. *Clin. Chim. Acta* 491, 1–7. doi: 10.1016/j.cca.2019.01.003

Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* 1473, 4–8. doi: 10.1016/s0304-4165(99)00165-8

Bradbury, E. J., Moon, L. D., Popat, R. J., King, V. R., Bennett, G. S., Patel, P. N., et al. (2002). Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature* 416, 636–640.

Briggs, D. C., and Hohenester, E. (2018). Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure* 26, 801–809.e3.

Chernykh, A., Kawahara, R., and Thaysen-Andersen, M. (2021). Towards structure-focused glycoproteomics. *Biochem. Soc. Trans.* 49, 161–186. doi: 10.1042/bst20200222

Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* 183, 1043–1057.e15.

Coles, C. H., Shen, Y., Tenney, A. P., Siebold, C., Sutton, G. C., Lu, W., et al. (2011). Proteoglycan-specific molecular switch for RPTPsigma clustering and neuronal extension. *Science* 332, 484–488. doi: 10.1126/science.1200840

Daly, J. L., Simonetti, B., Klein, K., Chen, K. E., Williamson, M. K., Anton-Plagaro, C., et al. (2020). Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* 370, 861–865.

Darula, Z., and Medzihradszky, K. F. (2018). Analysis of Mammalian O-Glycopeptides-We Have Made a Good Start, but There is a Long Way to Go. *Mol. Cell Proteomics* 17, 2–17. doi: 10.1074/mcp.mr117.000126

Delbaere, S., De Clercq, A., Mizumoto, S., Noborn, F., Bek, J. W., Alluyn, L., et al. (2020). b3galt6 Knock-Out Zebrafish Recapitulate β3GalT6-Deficiency Disorders in Human and Reveal a Trisaccharide Proteoglycan Linkage Region. *Front. Cell Dev. Biol.* 8:597857. doi: 10.3389/fcell.2020.597857

Doody, K. M., Stanford, S. M., Sacchetti, C., Svensson, M. N., Coles, C. H., Mitakidis, N., et al. (2015). Targeting phosphatase-dependent proteoglycan switch for rheumatoid arthritis therapy. *Sci. Transl. Med.* 7:288ra276.

Dwivedi, R., and Pomin, V. H. (2020). Marine Antithrombotics. *Mar. Drugs* 18:514. doi: 10.3390/md18100514

Esko, J. D., and Zhang, L. (1996). Influence of core protein sequence on glycosaminoglycan assembly. *Curr. Opin. Struct. Biol.* 6, 663–670. doi: 10.1016/s0959-440x(96)80034-0

Fisher, L. W., Termine, J. D., and Young, M. F. (1989). Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. *J. Biol. Chem.* 264, 4571–4576. doi: 10.1016/s0021-9258(18)83781-4

Gomez Toledo, A., Nilsson, J., Noborn, F., Sihlbom, C., and Larson, G. (2015). Positive Mode LC-MS/MS Analysis of Chondroitin Sulfate Modified Glycopeptides Derived from Light and Heavy Chains of The Human Inter-alpha-Trypsin Inhibitor Complex. *Mol. Cell Proteomics* 14, 3118–3131. doi: 10.1074/mcp.m115.051136

Gowda, D. C., Hogue-Angeletti, R., Margolis, R. K., and Margolis, R. U. (1990). Chromaffin granule and PC12 cell chondroitin sulfate proteoglycans and their relation to chromogranin A. *Arch. Biochem. Biophys.* 281, 219–224. doi: 10.1016/0003-9861(90)90435-2

Hatano, S., and Watanabe, H. (2020). Regulation of Macrophage and Dendritic Cell Function by Chondroitin Sulfate in Innate to Antigen-Specific Adaptive Immunity. *Front. Immunol.* 11:232. doi: 10.3389/fimmu.2020.00232

Hu, H. Z., Granger, N., Pai, S. B., Bellamkonda, R. V., and Jeffery, N. D. (2018). Therapeutic efficacy of microtube-embedded chondroitinase ABC in a canine clinical model of spinal cord injury. *Brain* 141, 1017–1027. doi: 10.1093/brain/awy007

Hu, J., Rodemer, W., Zhang, G., Jin, L. Q., Li, S., and Selzer, M. E. (2021). Chondroitinase ABC Promotes Axon Regeneration and Reduces Retrograde Apoptosis Signaling in Lamprey. *Front. Cell Dev. Biol.* 9:653638. doi: 10.3389/fcell.2021.653638

Hu, X., Villodre, E. S., Larson, R., Rahal, O. M., Wang, X., Gong, Y., et al. (2021). Decorin-mediated suppression of tumorigenesis, invasion, and metastasis in inflammatory breast cancer. *Commun. Biol.* 4:72. doi: 10.1201/b14586-5

Hussein, R. K., Mencio, C. P., Katagiri, Y., Brake, A. M., and Geller, H. M. (2020). Role of Chondroitin Sulfation Following Spinal Cord Injury. *Front. Cell Neurosci.* 14:208. doi: 10.3389/fncel.2020.00208

Iozzo, R. V. (2005). Basement membrane proteoglycans: from cellar to ceiling. *Nat. Rev. Mol. Cell. Biol.* 6, 646–656. doi: 10.1038/nrm1702

Iozzo, R. V., and Sanderson, R. D. (2011). Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J. Cell Mol. Med.* 15, 1013–1031. doi: 10.1111/j.1582-4934.2010.01236.x

Iozzo, R. V., and 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

Kane, E. G., and Taylor-Robinson, A. W. (2011). Prospects and Pitfalls of Pregnancy-Associated Malaria Vaccination Based on the Natural Immune Response to Plasmodium falciparum VAR2CSA-Expressing Parasites. *Malar. Res. Treat.* 2011:764545.

Kim, T., Tao-Cheng, J. H., Eiden, L. E., and Loh, Y. P. (2001). Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis. *Cell* 106, 499–509. doi: 10.1016/s0092-8674(01)00459-7

Kitagawa, H., Tsutsumi, K., Ikegami-Kuzuhara, A., Nadanaka, S., Goto, F., Ogawa, T., et al. (2008). Sulfation of the galactose residues in the glycosaminoglycan-protein linkage region by recombinant human chondroitin 6-O-sulfotransferase-1. *J. Biol. Chem.* 283, 27438–27443. doi: 10.1074/jbc.m803279200

Klein, J. A., Meng, L., and Zaia, J. (2018). Deep Sequencing of Complex Proteoglycans: a Novel Strategy for High Coverage and Site-specific Identification of Glycosaminoglycan-linked Peptides. *Mol. Cell Proteomics* 17, 1578–1590. doi: 10.1074/mcp.ra118.000766

Koike, T., Izumikawa, T., Sato, B., and Kitagawa, H. (2014). Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* 289, 6695–6708. doi: 10.1074/jbc.m113.520536

Krusius, T., and Ruoslahti, E. (1986). Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA. *Proc. Natl. Acad. Sci. U. S. A.* 83, 7683–7687. doi: 10.1073/pnas.83.20.7683

Lindahl, U., and Roden, L. (1966). The chondroitin 4-sulfate-protein linkage. *J. Biol. Chem.* 241, 2113–2119. doi: 10.1016/s0021-9258(18)96674-3

Lord, M. S., Day, A. J., Youssef, P., Zhuo, L., Watanabe, H., Caterson, B., et al. (2013). Sulfation of the bikunin chondroitin sulfate chain determines heavy chain.hyaluronan complex formation. *J. Biol. Chem.* 288, 22930–22941. doi: 10.1074/jbc.m112.404186

Lord, M. S., Melrose, J., Day, A. J., and Whitelock, J. M. (2020). The Inter-alpha-Trypsin Inhibitor Family: versatile Molecules in Biology and Pathology. *J. Histochem. Cytochem.* 68, 907–927. doi: 10.1369/0022155420940067

Lu, H., Mcdowell, L. M., Studelska, D. R., and Zhang, L. (2010). Glycosaminoglycans in Human and Bovine Serum: detection of Twenty-Four Heparan Sulfate and Chondroitin Sulfate Motifs Including a Novel Sialic Acid-modified Chondroitin Sulfate Linkage Hexasaccharide. *Glycobiol. Insights* 2010, 13–28. doi: 10.4137/gbi.s4273

Ly, M., Leach, F. E. III., Laremore, T. N., Toida, T., Amster, I. J., and Linhardt, R. J. (2011). The proteoglycan bikunin has a defined sequence. *Nat. Chem. Biol.* 7, 827–833. doi: 10.1038/nchembio.673

Ma, R., Lian, T., Huang, R., Renn, J. P., Petersen, J. D., Zimmerberg, J., et al. (2021). Structural basis for placental malaria mediated by Plasmodium falciparum VAR2CSA. *Nat. Microbiol.* 6, 380–391. doi: 10.1038/s41564-020-00858-9

Madsen, T. D., Hansen, L. H., Hintze, J., Ye, Z., Jebari, S., Andersen, D. B., et al. (2020). An atlas of O-linked glycosylation on peptide hormones reveals diverse biological roles. *Nat. Commun.* 11:4033.

Maeda, N. (2015). Proteoglycans and neuronal migration in the cerebral cortex during development and disease. *Front. Neurosci.* 9:98. doi: 10.3389/fnins.2015.00098

Maji, S. K., Perrin, M. H., Sawaya, M. R., Jessberger, S., Vadodaria, K., Rissman, R. A., et al. (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* 325, 328–332. doi: 10.1126/science.1173155

Malfait, F., Kariminejad, A., Van Damme, T., Gauche, C., Syx, D., Merhi-Soussi, F., et al. (2013). Defective initiation of glycosaminoglycan synthesis due to B3GALT6 mutations causes a pleiotropic Ehlers-Danlos-syndrome-like connective tissue disorder. *Am. J. Hum. Genet.* 92, 935–945. doi: 10.1016/j.ajhg.2013.04.016

Momota, R., Naito, I., Ninomiya, Y., and Ohtsuka, A. (2011). Drosophila type XV/XVIII collagen, Mp, is involved in Wingless distribution. *Matrix Biol.* 30, 258–266. doi: 10.1016/j.matbio.2011.03.008

Montoliu-Gaya, L., Tietze, D., Kaminski, D., Mirgorodskaya, E., Tietze, A. A., and Sterky, F. H. (2021). CA10 regulates neurexin heparan sulfate addition via a direct binding in the secretory pathway. *EMBO Rep.* 22:e51349.

Muir, H. (1958). The nature of the link between protein and carbohydrate of a chondroitin sulphate complex from hyaline cartilage. *Biochem. J.* 69, 195–204. doi: 10.1042/bj0690195

Mukhamedshina, Y. O., Povysheva, T. V., Nikolenko, V. N., Kuznecov, M. S., Rizvanov, A. A., and Chelyshev, Y. A. (2019). Upregulation of proteoglycans in the perilesion perimeter in ventral horns after spinal cord injury. *Neurosci. Lett.* 704, 220–228. doi: 10.1016/j.neulet.2019.04.006

Mycroft-West, C. J., Su, D., Pagani, I., Rudd, T. R., Elli, S., Gandhi, N. S., et al. (2020). Heparin Inhibits Cellular Invasion by SARS-CoV-2: structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thromb. Haemost.* 120, 1700–1715. doi: 10.1055/s-0040-1721319

Myron, P., Siddiquee, S., and Al Azad, S. (2014). Fucosylated chondroitin sulfate diversity in sea cucumbers: a review. *Carbohydr. Polym.* 112, 173–178. doi: 10.1016/j.carbpol.2014.05.091

Narimatsu, H., Kaji, H., Vakhrushev, S. Y., Clausen, H., Zhang, H., Noro, E., et al. (2018). Current Technologies for Complex Glycoproteomics and Their Applications to Biology/Disease-Driven Glycoproteomics. *J. Proteome Res.* 17, 4097–4112. doi: 10.1021/acs.jproteome.8b00515

Nasir, W., Toledo, A. G., Noborn, F., Nilsson, J., Wang, M., Bandeira, N., et al. (2016). SweetNET: a Bioinformatics Workflow for Glycopeptide MS/MS Spectral Analysis. *J. Proteome Res.* 15, 2826–2840. doi: 10.1021/acs.jproteome.6b00417

Neill, T., Schaefer, L., and Iozzo, R. V. (2016). Decorin as a multivalent therapeutic agent against cancer. *Adv. Drug Deliv. Rev.* 97, 174–185. doi: 10.1016/j.addr.2015.10.016

Nikpour, M., Nilsson, J., Persson, A., Noborn, F., Vorontsov, E., and Larson, G. (2021). Proteoglycan profiling of human, rat and mouse insulin-secreting cells. *Glycobiology* doi: 10.1093/glycob/cwab035 [Epub Online ahead of print].

Nilsson, J. (2016). Liquid chromatography-tandem mass spectrometry-based fragmentation analysis of glycopeptides. *Glycoconj. J.* 33, 261–272. doi: 10.1007/s10719-016-9649-3

Nilsson, J., Noborn, F., Gomez Toledo, A., Nasir, W., Sihlbom, C., and Larson, G. (2017). Characterization of Glycan Structures of Chondroitin Sulfate-Glycopeptides Facilitated by Sodium Ion-Pairing and Positive Mode LC-MS/MS. *J. Am. Soc. Mass Spectrom.* 28, 229–241. doi: 10.1007/s13361-016-1539-1

Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., et al. (2009). Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* 6, 809–811. doi: 10.1038/nmeth.1392

Noborn, F., Gomez Toledo, A., Green, A., Nasir, W., Sihlbom, C., Nilsson, J., et al. (2016). Site-specific identification of heparan and chondroitin sulfate glycosaminoglycans in hybrid proteoglycans. *Sci. Rep.* 6:34537.

Noborn, F., Gomez Toledo, A., Nasir, W., Nilsson, J., Dierker, T., Kjellen, L., et al. (2018). Expanding the chondroitin glycoproteome of Caenorhabditis elegans. *J. Biol. Chem.* 293, 379–389. doi: 10.1074/jbc.m117.807800

Noborn, F., Gomez Toledo, A., Sihlbom, C., Lengqvist, J., Fries, E., Kjellen, L., et al. (2015). Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. *Mol. Cell Proteomics* 14, 41–49. doi: 10.1074/mcp.m114.043703

Noborn, F., and Larson, G. (2021). Characterization of C. elegans Chondroitin Proteoglycans and Their Large Functional and Structural Heterogeneity; Evolutionary Aspects on Structural Differences Between Humans and the Nematode. *Adv. Exp. Med. Biol.* 21, 155–170. doi: 10.1007/5584_2020_485

Noborn, F. N., Persson, M., Sihlbom, A., Nilsson, C., and Larson, J. (2021). A Glycoproteomic Approach to Identify Novel Proteoglycans. *Methods Mol. Biol. Glycosaminoglycans, 978-1-0716-1397-9, 481765_2_En,* 2303, ed. K. Balagurunathan, et al.

Olofsson, S., and Bergström, T. (2005). Glycoconjugate glycans as viral receptors. *Ann. Med.* 37, 154–172. doi: 10.1080/07853890510007340

Olson, S. K., Bishop, J. R., Yates, J. R., Oegema, K., and Esko, J. D. (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

Pereira, N. A., Chan, K. F., Lin, P. C., and Song, Z. (2018). The "less-is-more" in therapeutic antibodies: Afucosylated anti-cancer antibodies with enhanced antibody-dependent cellular cytotoxicity. *MAbs* 10, 693–711. doi: 10.1080/19420862.2018.1466767

Persson, A., Nilsson, J., Vorontsov, E., Noborn, F., and Larson, G. (2019). Identification of a non-canonical chondroitin sulfate linkage region trisaccharide. *Glycobiology* 29, 366–371. doi: 10.1093/glycob/cwz014

Plomp, R., Ruhaak, L. R., Uh, H. W., Reiding, K. R., Selman, M., Houwing-Duistermaat, J. J., et al. (2017). Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. *Sci. Rep.* 7:12325.

Pomin, V. H., and Mulloy, B. (2018). Glycosaminoglycans and Proteoglycans. *Pharmaceuticals (Basel)* 11:27.

Reszegi, A., Horvath, Z., Feher, H., Wichmann, B., Tatrai, P., Kovalszky, I., et al. (2020). Protective Role of Decorin in Primary Hepatocellular Carcinoma. *Front. Oncol.* 10:645. doi: 10.3389/fonc.2020.00645

Roden, L., and Smith, R. (1966). Structure of the neutral trisaccharide of the chondroitin 4-sulfate-protein linkage region. *J. Biol. Chem.* 241, 5949–5954. doi: 10.1016/s0021-9258(18)96362-3

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., and Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* 21, 729–749. doi: 10.1038/s41580-020-00294-x

Schneider, M., Al-Shareffi, E., and Haltiwanger, R. S. (2017). Biological functions of fucose in mammals. *Glycobiology* 27, 601–618. doi: 10.1093/glycob/cwx034

Shatton, J., and Schubert, M. (1954). Isolation of a mucoprotein from cartilage. *J. Biol. Chem.* 211, 565–573. doi: 10.1016/s0021-9258(18)71147-2

Sugahara, K., Yamashina, I., De Waard, P., Van Halbeek, H., and Vliegenthart, J. F. (1988). Structural studies on sulfated glycopeptides from the carbohydrate-protein linkage region of chondroitin 4-sulfate proteoglycans of swarm rat chondrosarcoma. Demonstration of the structure Gal(4-O-sulfate)beta 1-3Gal beta 1-4XYL beta 1-O-Ser. *J. Biol. Chem.* 263, 10168–10174. doi: 10.1016/s0021-9258(19)81492-8

Sugiura, N., Clausen, T. M., Shioiri, T., Gustavsson, T., Watanabe, H., and Salanti, A. (2016). Molecular dissection of placental malaria protein VAR2CSA interaction with a chemo-enzymatically synthesized chondroitin sulfate library. *Glycoconj. J.* 33, 985–994. doi: 10.1007/s10719-016-9685-z

Takemura, M., Noborn, F., Nilsson, J., Bowden, N., Nakato, E., Baker, S., et al. (2020). Chondroitin sulfate proteoglycan Windpipe modulates Hedgehog signaling in Drosophila. *Mol. Biol. Cell* 31, 813–824. doi: 10.1091/mbc.e19-06-0327

Thaysen-Andersen, M., Packer, N. H., and Schulz, B. L. (2016). Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-glycoproteome and Its Regulation in Health and Disease. *Mol. Cell Proteomics* 15, 1773–1790. doi: 10.1074/mcp.o115.057638

Toledo, A. G., Pihl, J., Spliid, C. B., Persson, A., Nilsson, J., Pereira, M. A., et al. (2020). An Affinity-Chromatography and Glycoproteomics Workflow to Profile the Chondroitin Sulfate Proteoglycans that interact with malarial

VAR2CSA in the Placenta and in Cancer. *Glycobiology* 30, 989–1002. doi: 10. 1093/glycob/cwaa039

Tomlinson, A., Semblat, J. P., Gamain, B., and Chene, A. (2020). VAR2CSA-Mediated Host Defense Evasion of Plasmodium falciparum Infected Erythrocytes in Placental Malaria. *Front. Immunol.* 11:624126. doi: 10.3389/fimmu.2020.624126

Wang, T. T., and Ravetch, J. V. (2019). Functional diversification of IgGs through Fc glycosylation. *J. Clin. Invest.* 129, 3492–3498. doi: 10.1172/jci 130029

Wen, J., Xiao, J., Rahdar, M., Choudhury, B. P., Cui, J., Taylor, G. S., et al. (2014). Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15723–15728.

Zhang, P., Lu, H., Peixoto, R. T., Pines, M. K., Ge, Y., Oku, S., et al. (2018). Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships. *Cell* 174, 1450–1464.e23.

Zhu, F., Li, D., and Chen, K. (2019). Structures and functions of invertebrate glycosylation. *Open Biol.* 9:180232.