



Characterization of *C. elegans* Chondroitin Proteoglycans and Their Large Functional and Structural Heterogeneity; Evolutionary Aspects on Structural Differences Between Humans and the Nematode

Fredrik Noborn and Göran Larson

Abstract

Proteoglycans regulate important cellular pathways in essentially all metazoan organisms. While considerable effort has been devoted to study structural and functional aspects of proteoglycans in vertebrates, the knowledge of the core proteins and proteoglycan-related functions in invertebrates is relatively scarce, even for *C.elegans*. This nematode produces a large amount of non-sulfated chondroitin in addition to small amount of low-sulfated chondroitin chains (Chn and CS chains, respectively). Until recently, 9 chondroitin core proteins (CPGs) had been identified in *C.elegans*, none of which showed any homology to vertebrate counterparts or to other invertebrate core proteins. By using a glycoproteomic approach, we recently characterized the chondroitin

glycoproteome of *C.elegans*, resulting in the identification of 15 novel CPG core proteins in addition to the 9 previously established. Three of the novel core proteins displayed homology to human proteins, indicating that CPG and CSPG core proteins may be more conserved throughout evolution than previously perceived. Bioinformatic analysis of the primary amino acid sequences revealed that the core proteins contained a broad range of functional domains, indicating that specialization of proteoglycan-mediated functions may have evolved early in metazoan evolution. This review specifically discusses our recent data in relation to previous knowledge of core proteins and GAG-attachment sites in Chn and CS proteoglycans of *C.elegans* and humans, and point out both converging and diverging aspects of proteoglycan evolution.

F. Noborn
Department of Laboratory Medicine, Sahlgrenska
Academy at the University of Gothenburg, Gothenburg,
Sweden

G. Larson (✉)
Department of Laboratory Medicine, Sahlgrenska
University Hospital, Sahlgrenska Academy at the
University of Gothenburg, Gothenburg, Sweden
e-mail: goran.larson@clinchem.gu.se

Keywords

Proteoglycan · *Caenorhabditis elegans* ·
Glycosaminoglycan · Chondroitin ·
Chondroitin sulfate · Core protein · Protein
functional domains · Attachment site ·
Metazoan evolution

1 Introduction

The notion of “proteoglycans” as discrete entities first became apparent during the 1950s (Yanagishita 1993). By isolating and analyzing material from bovine cartilage, it was found that glycosaminoglycan (GAG) chains were associated with a protein component (Yanagishita 1993; Schatton and Schubert 1954). These compounds were referred to as “mucoproteins” although it was unclear at that time whether the GAG-protein association involved covalent bonds or not. In the following years, a covalent association was indeed demonstrated between chondroitin sulfate and serine residues (Muir 1958; Lindahl 2014). Furthermore, a tetrasaccharide “linkage region” [Glucuronic acid (GlcA) – Galactose (Gal) – Galactose (Gal) – Xylose (Xyl)] was identified that covalently linked the GAG chain to specific serine residues of the corresponding core proteins (Roden and Smith 1966). Since then, core proteins have gradually become recognized as distinct molecular entities, each with differences in their protein structures and cellular functions, as well as with differences in the number and types of GAG chains attached (Lindahl 2014; Murdoch and Iozzo 1993; Kjellen et al. 1989; Lindahl et al. 2015).

The identification of proteoglycans is often difficult from a methodological perspective, since proteoglycan identification requires the combined sequencing of a given core protein, together with the characterization of which type, and where along the amino acid sequence the GAG chain(s) are attached. Biochemical and immunological techniques are often hampered by the size and heterogeneity of the GAG side chains, which preclude effective core protein sequencing and characterization. Molecular cloning techniques offer a solution to these difficulties by allowing the identification and sequencing of mRNAs and protein coding genes. However, these methods do not provide information on any post-translational modifications, which creates ambiguity as to the identification of a proteoglycan (Kjellen et al. 1989; Bourdon et al. 1985). Therefore, studies on identifying

proteoglycans have earlier been focused mostly on isolation and characterization of a single core protein in specific model systems, whereas unbiased and global characterizations of all proteoglycans of a specific tissue or organism have not been systematically attempted.

The number of core proteins identified in vertebrates is limited. Less than 20 heparan sulfate proteoglycans (HSPGs) and about 60 chondroitin sulfate proteoglycans (CSPGs) have so far been identified in humans (Lindahl 2014; Zhang et al. 2018; Noborn et al. 2015; Nasir et al. 2016). This is a very limited number in relation to other types of glycoproteins, such as N- and O-glycosylated proteins, which are counted in their thousands (Nilsson et al. 2013; Joshi et al. 2018). We have recently developed a glycoproteomic approach that may assist in identifying how many and which type of proteoglycans are indeed expressed in different animal tissues and species. The aim was to characterize linkage regions, attachment sites and identities of CS core proteins (Noborn et al. 2015). In this approach, trypsin-treated proteoglycans were enriched from various sample matrices by strong-anion-exchange chromatography, and then digested with chondroitinase ABC to specifically reduce the CS chain lengths. The preparations were thereafter analyzed by nLC-MS/MS and the data from remaining linkage regions, linked to tryptic or semi-tryptic peptides, was processed by a novel glycopeptide search algorithm. Analysis of human urine and CSF resulted in the identification of 13 novel CSPGs, many of which were previously defined as peptide prohormones (Noborn et al. 2015). This suggested that many novel proteoglycans and proteoglycan-related functions are yet to be discovered, and that new methodological approaches may assist in such an endeavor.

While proteoglycans in vertebrates has been the focus of several structural and functional studies, the knowledge of proteoglycans in invertebrates is still relatively scarce, even for the otherwise well-studied nematode *C.elegans*. Until recently, 5 HSPGs and 9 CPG core proteins had been identified in the nematode (Wilson et al. 2015; Olson et al. 2006). Using our

glycoproteomic strategy, we mapped the chondroitin glycoproteome of *C.elegans*, confirming the identities of the 9 previously established core proteins, but also identifying an additional 15 chondroitin core proteins (Noborn et al. 2018). Three of the novel core proteins displayed homologies to human proteins, which was surprising since no chondroitin core proteins have previously been found to display homology to human proteins, and were therefore not considered to be well-conserved throughout evolution (Olson et al. 2006). Bioinformatic analysis of the primary amino acid sequences was performed to provide insights of the structural domain organization of each core protein. This analysis revealed a previously unknown structural complexity of CPGs in *C.elegans*, indicating that complex proteoglycan-related functions may have evolved early in metazoan evolution.

Additional glycoproteomic analyses of proteoglycans of various animals, vertebrates as well as invertebrates, are likely to expand our understanding of the structural heterogeneity of Chn and CS core proteins during metazoan evolution. However, at present it is difficult to fully appreciate the evolutionary aspects on core protein alterations in large, simply because the number of studies on core proteins in invertebrates is too limited (and that available studies typically focus only on a single core protein). Thus, this review will concentrate on the recent findings of CPGs and CSPGs in *C.elegans* and humans and points to similarities and differences between core proteins between these two evolutionary distant species. Although core proteins are the primary focus of this review, the initiating GAG-biosynthetic machinery in *C.elegans* and humans will also be discussed to highlight both converging and diverging aspects of proteoglycan evolution. References to relevant reviews relating to structural diversity of GAGs and proteoglycans in other organisms are given in their conceptual contexts in the following paragraphs. Our general and specific conclusions are summarized in Fig. 1, exemplifying our conclusions on some evolutionary principles of proteoglycan development from *C.elegans* to *Homo sapiens*.

2 Proteoglycan Diversity from *C. elegans* to Humans

Glycoconjugates constitute the structurally most diverse group of organic molecules in nature. This diversity poses a great challenge in analyzing glycan structures and also in assigning glycan-specific functions (Joshi et al. 2018; Gagneux et al. 2015; Mulloy et al. 2009). Although GAGs constitute a small subgroup of all glycans, their structural complexity is still considerable. GAGs are divided into four subclasses depending on the repeating disaccharides of the polysaccharide chains: heparan sulfate (HS)/heparin (GlcA/IdoA-GlcN), chondroitin sulfate (CS) /dermatan sulfate (DS) (GlcA/IdoA-GalNAc), keratan sulfate (Gal-GlcNAc) and hyaluronan (GlcA-GlcNAc). The molecular heterogeneity is influenced by large differences in polysaccharide chain length, domain organization and unique monosaccharide modifications, e.g. O- and N-sulfations, phosphorylations, sialylations etc. All GAGs, except for hyaluronan, are invariably attached to core proteins to form unique proteoglycans (Hascall et al. 2014; Weigel et al. 1997; Saied-Santiago and Bulow 2018). This provides additional complexity since different core proteins, have unique gene coded primary, secondary and tertiary sequences, with major consequences for where, which type and which number of GAG chains that are initiated, extended and modified.

Proteoglycans have a long evolutionary history and are expressed in all bilateral animals investigated to date (Couchman and Pataki 2012). HS appeared early in metazoan organisms and essentially all cells produce complex sulfated HS structures (Esko and Lindahl 2001). In contrast, CS/DS from lower organisms have a limited structural complexity, a complexity which increases with evolutionary higher organisms (Yamada et al. 2011). For extensive information on proteoglycan diversity and GAG-specific functions we highly recommend the following reviews (Couchman and Pataki 2012; Iozzo and Schaefer 2015; Kjellen and Lindahl 2018; Weiss et al. 2017). Here, examples are selected to focus

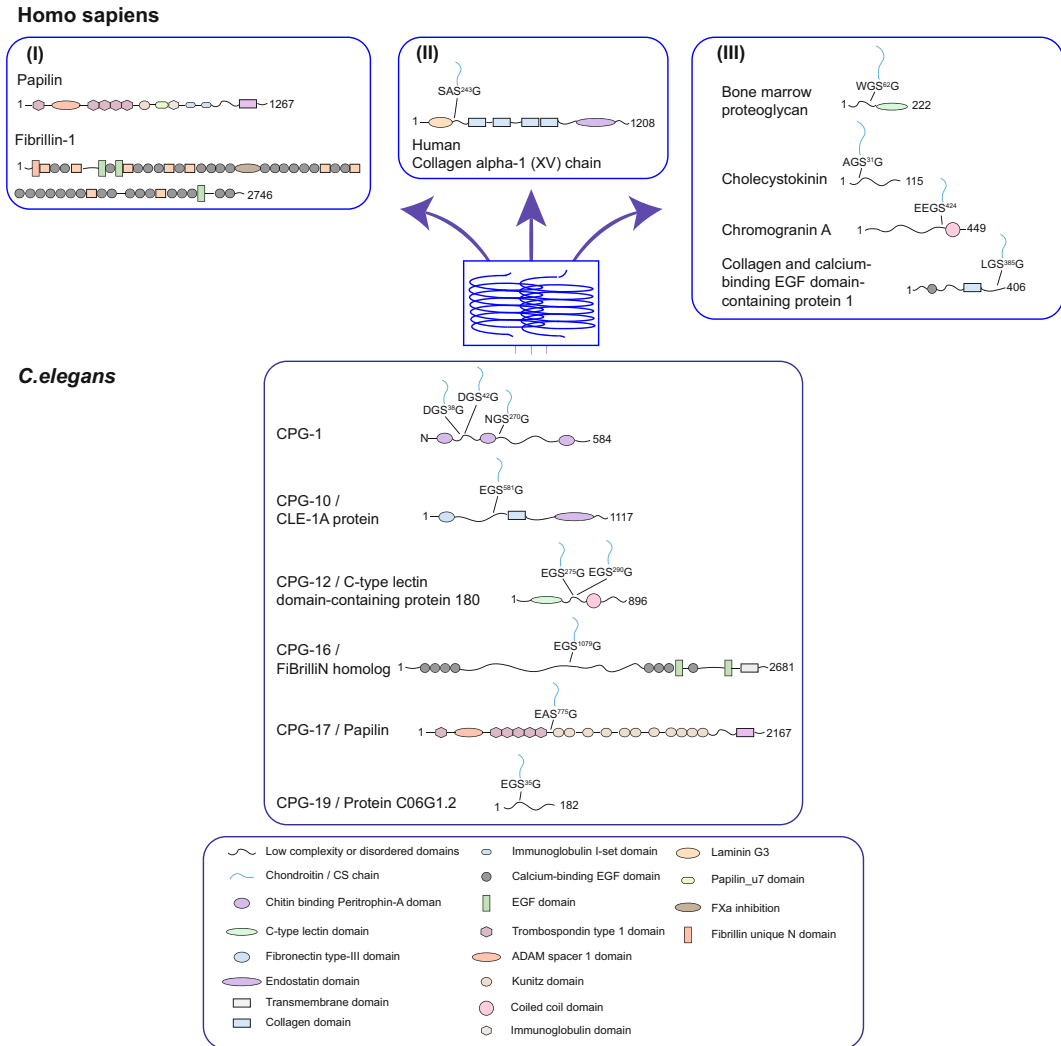


Fig. 1 Schematic illustration of evolutionary principles bridging millions of years of proteoglycan development from *C.elegans* to *Homo sapiens*. (I) Divergent evolution where the GAG chain is lost but the protein is conserved, (II) Parallel evolution conserving both the GAG chain and the protein, (III) Convergent evolution where conserved GAG chains are added to

novel core proteins. Note that some functional core protein domains are conserved throughout evolution whereas other core proteins lack such domains. The spirals/arrows are representing the evolutionary multi-interaction steps for the two primary constituents of proteoglycans, i.e. GAGs and core proteins

primarily on different aspects of CPGs in *C. elegans* and are not meant to provide a comprehensive review on proteoglycan structure and evolution in general. Hopefully, this review will provide some new aspects in proteoglycan structure and perhaps inspire to novel concepts that can be experimentally tested.

3 Structural Diversity of Chondroitin and Heparan Sulfate Proteoglycans in Invertebrates

The number of CSPGs identified in humans is around 60 (Noborn et al. 2015; Nasir et al.

2016) but in invertebrates, the number of CSPGs is even lower and the reports are restricted to only a few species. A proteomic-based study identified tryptic peptides from versican, neurocan (CSPG3) and neuroglycan (CSPG4-NG2) in the gastropoda *Achatina fulica* (Gesteira et al. 2011). Since these proteins are well-established CSPGs in vertebrates, assumptions of their CS substitutions were also made in *A. fulica*. Moreover, two populations of CSPGs with different molecular weights were isolated from squid skin (*Ilex illecebrosus*) using a combination of ion-exchange chromatography and ultracentrifugation (Karamanos et al. 1990). Biochemical analysis showed different amino acid composition of these core proteins, although the exact peptide sequences could not be resolved (Karamanos et al. 1990). Surprisingly, information on CSPGs is lacking in *Drosophila melanogaster*, regarded as one of the most studied invertebrates in glycobiology (Zhu et al. 2019), which supports our perception that there is a general gap of knowledge of these structures in invertebrates.

However, earlier studies identified various HSPGs in *D. melanogaster* and *C. elegans* which display homologies to vertebrate core proteins. There are 5 known HSPGs in *D. melanogaster* that are all homologues of mammalian counterparts: syndecan, 2 glypicans (*dally* and *dlp*), perlecan (*trol*) and testican (*cow*) (Bernfield et al. 1999; Nakato et al. 1995; Baeg et al. 2001; Park et al. 2003; Chang and Sun 2014; Nakato and Li 2016). Similar to these findings, five homologues to vertebrate genes encoding for HSPG core proteins have been identified in *C. elegans*: syndecan (*sdn-1*), 2 glypicans (*lon-2* and *gpn-1*), perlecan (*unc-52*) and agrin (*agr-1*) (Blanchette et al. 2015; Consortium CeS 1998; Rogalski et al. 1993; Hrus et al. 2007; Hutter et al. 2000; Blanchette et al. 2017). The *unc-52* gene encodes the homologue of the vertebrate gene perlecan, a major component of the extracellular matrix, which in vertebrates is substituted with both HS and CS (Rogalski et al. 1993; Yamada et al. 2002; Noborn et al. 2016). There is of course a possibility that additional HSPGs, yet unidentified and maybe not conserved, may

exist in both *Drosophila* and *C. elegans*. Nevertheless, the findings so far indicate a high degree of conservation of genes encoding for HSPG core proteins throughout evolution.

Nine CPGs have previously been identified in *C. elegans*, which were designated CPG-1 to CPG-9 (Olson et al. 2006). In contrast to the HSPGs, none of these core proteins showed homology to vertebrate proteins or to proteins in other invertebrates such as *Drosophila melanogaster* (Olson et al. 2006). Two of the CPGs (CPG-1 and CPG-2) in *C. elegans* contain chitin-binding domains and were therefore assumed to interact with the chitin layer in the cuticle (Wilson et al. 2015). Detailed functional analysis showed that CPG-1 and CPG-2 are indeed important for the hierarchical assembly of the egg shell layer during embryogenesis, resulting in an outer vitelline layer, a middle chitin layer and an inner CPG-1 and CPG-2 layer (Olson et al. 2012). This specific function confers to the classical notion of CSPGs as structural components in cartilage and other connective tissues. Since vertebrate CSPGs display a wide range of functional diversity, we argued that additional CPG core proteins are likely present in *C. elegans*, which not only relate to extracellular matrix formation, but may also accommodate more specialized functions.

Indeed, in our investigation of the chondroitin glycoproteome of *C. elegans*, we found 15 novel core proteins that were designated CPG-10 through -24, in accordance with previous introduced terminology (Olson et al. 2006). Six of the 15 novel core proteins were previously uncharacterized proteins, and were only annotated in UniProt based on the open reading frame (ORF) names (e.g. Protein C45E5.4/CPG-18) (Noborn et al. 2018). The other novel CPGs have previously been assigned names based on phenotypes in mutation studies (e.g. High incidence of males, isoform b/CPG-14), or based on sequence similarities to vertebrate proteins (e.g. FiBrilliN homolog/CPG-16). The identified core proteins displayed a wide range in their molecular weights, from 7.1 kDa (CPG-9) to 568 kDa (high incidence of males, isoform b/CPG-14). The number of chondroitin attachment

sites also varied depending on the core protein, from one (e.g. CPG-3, CPG-5) to four sites (CPG-4). Moreover, bioinformatics analysis of the primary amino acid sequences revealed that the core proteins contained a broad range of functional domains, assuming their involvement in a wide-range of physiological functions. In total, 19 unique domains were retrieved from the 24 core protein sequences. Apart from the expected chitin-binding domains on CPG-1 and CPG-2, additional domains were identified that indicate their involvement in extracellular matrix formation, such as fibronectin type-III domain (CLE-1A protein/CPG-10) and collagen domain (COLLAGEN/CPG-11). Other identified domains indicate a role in more specialized functions, such as thrombospondin type-1 domain (Papilin/CPG-17) and endostatin domain (CLE-1A protein/CPG-10), both of which are known to be involved in axon guidance and neuronal development (Adams and Tucker 2000; Ackley et al. 2001). Notably, 9 core protein sequences did not retrieve any hits and displayed only low complexity/disordered domains.

Bioinformatic analysis was also conducted on human CSPGs, previously identified in human samples with the same approach. The analysis retrieved 40 unique domains for 28 core proteins sequences (Noborn et al. 2015; Nasir et al. 2016). Certain domains were found in both species, such as collagen domain and the Kunitz domain. Of the 50 unique domain structures identified in the two species, 31 were uniquely found in human CSPGs, 10 uniquely found in *C. elegans* CPGs, and 9 found in both species. Moreover, sequences that only display disordered domains were also found in humans, although to a lesser degree than in *C. elegans*. Three of 28 human core proteins (10.7%) displayed this characteristic, compared to 9 out of 24 (37.5%) in *C. elegans*. This may indicate a selection process where core proteins with functional domains are conserved throughout evolution. A certain amount of research bias regarding detection of functional domains in the data base (e.g. more information of human proteins), may however also explain the higher incidence of known domains in human. Nevertheless, this analysis suggests a great structural

and also functional diversity of CPGs in *C. elegans* and indicates that some, but not all, functions overlap with those of human CSPGs. Furthermore, this indicates that also specialized CS proteoglycan-mediated functions may have evolved early in metazoan evolution.

4 Evolutionary Aspects of CS Biosynthesis in *C.elegans* and Humans

Although *C.elegans* is a well-studied model organism with regard to genomics, proteomics and certain aspects of glycosylation (Consortium CeS 1998; Antoshechkin and Sternberg 2007; Shim and Paik 2010; Schachter 2004), information on CS proteoglycans and proteoglycan-mediated functions is limited. This is unfortunate since *C.elegans* is often used to study the influence of genes and proteins in evolutionary conserved processes (Maduro 2017; Vuong-Brender et al. 2016). Such processes, e.g. morphogen distribution in embryogenesis of e.g. *D. melanogaster*, have been shown to involve HS proteoglycans which fine tunes the cellular response (Nakato and Li 2016; Bishop et al. 2007). Structural information of CS proteoglycans in *C.elegans* would therefore probably assist to our functional understanding of these processes in the worm.

The CS (and GAG) biosynthesis is always initiated by the transfer of a Xyl to a serine residue in the core protein. The xylosylation typically occurs at certain serine residues with a glycine residue at the carboxyl-terminal side (-SG-), and with a cluster of acidic residues in close proximity (Esko and Zhang 1996). This motif was initially observed for vertebrate core proteins and a similar motif has also been suggested for *C.elegans* (Olson et al. 2006). The chondroitin sulfate biosynthesis continues with the addition of two galactose (Gal) and one glucuronic acid (GlcA) residue, completing the formation of the consensus tetrasaccharide linkage region. The biosynthesis then continues with polymerization of the chain through the addition of alternating units of N-acetylgalactosamine (GalNAc) and GlcA

residues. The individual enzymes for each step in the chondroitin biosynthesis in *C.elegans* have been well established. In a mutagenesis experiment, eight mutations that perturb vulval development in the growing embryo were identified (designated *sqv* or squashed vulva). All eight mutations (*sqv-1* to 8) produced similar phenotypes, such as a defective vulval epithelial invagination and for some mutations oocyte development was also affected (Wilson et al. 2015; Herman et al. 1999). Moreover, all *sqv* genes showed homology to vertebrate enzymes and were found to be involved in different aspects of the GAG-biosynthesis. Biochemical analysis showed that *sqv-6*, *sqv-3*, *sqv-2*, and *sqv-8* encode for vertebrate homologues of glycosyltransferases required for the formation of the tetrasaccharide linkage regions, whereas *sqv-1*, *sqv-4*, and *sqv-7* encode proteins that have roles in nucleotide sugar metabolism and transport (Wilson et al. 2015; Herman and Horvitz 1999; Bulik et al. 2000; Berninsone et al. 2001; Hwang and Horvitz 2002; Hwang et al. 2003; Izumikawa et al. 2004). Taken together, all components required for the initial part of the biosynthesis is highly conserved between *C.elegans* and humans, including nucleotide sugar precursors and their transport into the Golgi, as well as enzymes required for linkage formation and chain polymerization (Olson et al. 2006).

In vertebrates, the chondroitin polysaccharide undergoes extensive modifications of sulfotransferases and chondroitin-specific epimerases (Mizumoto et al. 2013; Ly et al. 2011). This results in complex yet defined CS/DS structures that may interact with various protein ligands with different degree of specificities (Le Jan et al. 2012; Mizumoto et al. 2015; Sugiura et al. 2016). In contrast, chondroitin in *C.elegans* is considerably less complex and the general view was, until recently, that the nematode only produces non-sulfated chondroitin (Yamada et al. 1999; Toyoda et al. 2000). This was puzzling since *C.elegans*, which belongs to Ecdysozoa clade, appeared to be an exception to other animals within the same clade, such as *D. melanogaster*, which was known to produce sulfated structures (Toyoda et al. 2000).

Moreover, even animals in the evolutionary older phylum of Cnidarians, containing simple organisms such as hydrozoans, produce CS (Yamada et al. 2011). This paradox was recently settled when two separate groups demonstrated that Chn may indeed be sulfated in *C.elegans*, although to a smaller extent (Dierker et al. 2016; Izumikawa et al. 2016). So far only one single sulfotransferase, which catalyzes the GalNAc 4-*O* sulfation has been identified in *C.elegans*. However, the presence of both 4-*O* and 6-*O* sulfated GalNAc residues was shown by MS/MS analysis of CS disaccharides, and indicated that at least one additional sulfotransferase should be expressed in the nematode (Dierker et al. 2016). In contrast, *C.elegans* seems to lack the chondroitin-specific epimerases present in vertebrates and no DS structures have yet been detected in this nematode. Taken together, apart from the epimerase, all the components required for CS biosynthesis are highly conserved between *C.elegans* and humans, demonstrating an essential role for CS throughout metazoan evolution (Olson et al. 2006; Yamada et al. 2002; Berninsone et al. 2001).

5 Glycosaminoglycan Diversity from *C.elegans* to Human

However, not all aspects of GAG-evolution seems well conserved. Hyaluronan (HA), a non-sulfated GAG composed of long repeating units of GlcNAc and GlcA disaccharides, seems to have appeared quite late in evolution (Csoka and Stern 2013). The genome in *C.elegans* does not contain the necessary synthases for HA and there is no structural evidence of HA in the nematode (Yamada et al. 1999; Stern 2003). HA has various biological roles and is a prominent component of hydrated matrices in the extracellular matrix. Since Chn/CS only contains a few percent of sulfated disaccharides in *C.elegans* (Dierker et al. 2016), the large majority of the chains are therefore likely non-sulfated structures. Apart from their sizes, chondroitin and HA are relatively similar in structure with differences only in the isomeric identities of the HexNAc residues

(GalNAc vs GlcNAc). It has been suggested that chondroitin in *C.elegans* is a possible HA ancestor, carrying out functions in *C.elegans* that are assigned to HA in vertebrates (Stern 2003). While vertebrates have evolved two different GAG structures, CS and HA, to accommodate separate cellular functions, one may speculate that this structural-functional specialization also occurs in *C.elegans* to a certain extent. It is thus possible that non-sulfated chondroitin accommodate HA-like functions (e.g. provide hydrated matrices); whereas sulfated chondroitin structures accommodate more specialized functions (e.g. provide binding motifs to specific ligands).

Typically, a “GAG-perspective” or a “core protein perspective” is applied when studying the role of proteoglycans in various pathophysiological settings. This structural and conceptual separation is natural, given their vast structural heterogeneity and the limited number of analytical methods that provides integrated GAG-protein characterization. However, integrating structural information on the GAG chains, their attachment sites and the potential functional domains of the corresponding core protein, will likely provide new perspective when studying proteoglycan-related functions. For instance, the effect of chondroitin on neuronal migration in *C.elegans* has been studied by targeting two proteins in the chondroitin biosynthetic pathway: the chondroitin synthase (SQV-5) and a UDP-sugar transporter (SQV-7) (Pedersen et al. 2013). Worms with hypomorphic alleles in these proteins showed aberrant migration of hermaphrodite-specific neurons (HSN) (Pedersen et al. 2013). Although a functional relationship between reduced Chn synthesis and impaired neuronal migration was established, the molecular involvement of the corresponding core protein (s) remains unclear. Different scenarios are possible: the migration requires free Chn, or the migration requires Chn attached to a specific core protein, or even the active involvement of both Chn and a specific core protein. In fact, it was recently shown that neuexin, an essential component in synapse organization, was modified with HS (Zhang et al. 2018). The binding of neuexin to its post-synaptic partner, neuroligin,

involved an intrinsic mode of interaction, which required both the HS chain and the protein domain of neuexin. This underlines the importance of site-specific characterization to further delineate GAG-mediated functions in all organisms.

6 Chondroitin Sulfate and Core Proteins

The selective binding of specific protein ligands to structural variants of GAG chains regulates a diverse set of biological- and pathological processes (Kjellen and Lindahl 2018; Salanti et al. 2015; Kreuger et al. 2006; Sarrazin et al. 2011). Determining the fine-structure of binding domains or, when possible, intact GAG chains is therefore essential for understanding GAG-protein interactions and their down-stream cellular events. As *C.elegans* was recently found to have CS structures (Dierker et al. 2016; Izumikawa et al. 2016), characterization of the sulfate distribution on the polysaccharides will likely improve our understanding of CSPG- and CPG-related functions.

In our glycoproteomic approach, the Chn and CS chains are depolymerized with chondroitinase ABC, generating free disaccharides and a residual hexameric structure composed of the linkage region and a GlcA-GalNAc disaccharide, dehydrated on the terminal GlcA residue (Noborn et al. 2015; Noborn et al. 2018). This strategy reduces the complexity of the analysis significantly, but at the same time structural information towards the non-reducing end is omitted. Our analysis of CPGs in *C.elegans* did not however reveal any sulfate groups on the residual hexasaccharide structure, although the method is fully capable of detecting such modifications (Noborn et al. 2015). This may suggest that the sulfate groups are located further out on the chains, or in quantities below the present limit of detection. Regardless of their position, one may speculate whether the sulfate modifications are evenly distributed between the 24 different CPGs, or if only a subset of CPGs carries sulfated structures. Regulation of GAG-biosynthesis is

believed to be largely cell-specific as for instance GAG-structures from one mouse tissue differ from those of other mouse tissues (Kjellen and Lindahl 2018; Ledin et al. 2004). Cell-specific co-expression of GalNAc 4-*O* sulfotransferase and certain CPGs may thus result in CS chains on only a subset of core proteins, in a cell specific manner. Moreover, the modification pattern may also involve type of core proteins, although such reports are relatively scarce (Li et al. 2011). Apart from these two principles of regulation, the sulfation pattern may also be lineage specific, in that the sulfation varies in response to developmental stages and possibly disease states (Shao et al. 2013).

If sulfate groups are limited to a subset of CPGs, one may speculate which CPGs that carries sulfated structures. Three homologues to human proteins were found in *C.elegans*; CLE-1A protein/CPG-10, FiBrilliN/CPG-16 and Papilin/CPG-17. The CLE-1A protein/CPG-10 is encoded by the *cle-1* gene which produces three developmentally regulated protein isoforms (CLE-1A-C), which are expressed predominantly in neurons (Ackley et al. 2003). The CLE-1A protein is the homologue to human collagen alpha-1 XV/XVIII (Ackley et al. 2003). Interestingly, we recently found that the human collagen XV alpha-1 chain is substituted with CS in human tissue fluids (Noborn et al. 2015) and this is to our knowledge the first example of an invertebrate chondroitin core protein that shows homology to a vertebrate counterpart. Since all vertebrate core proteins carry CS chains, one may thus speculate that the three vertebrate homologues are likely candidates to be substituted with CS. Moreover, each of the CPG-vertebrate homologues contains functional domains that assume involvement in specialized proteoglycan-mediated functions. As mentioned previously, CLE-1A/CPG-10 contains an endostatin domain and deletion of this domain resulted in worms with defects in cell migration and axon guidance (Ackley et al. 2001, 2003). In vertebrates, CS inhibits nerve regeneration upon binding to the receptor protein tyrosine phosphatase sigma (RPTP σ), an interaction that requires uniform distribution of sulfate groups along the CS chain (Shen et al. 2009; Coles et al. 2011;

Katagiri et al. 2018; Sakamoto et al. 2019). Given that advanced functions, such as neurogenesis, require CS with certain sulfate distribution in vertebrates, it is plausible that this is also a requirement in *C.elegans*. Moreover, Papilin/CPG-17 is also claimed to be involved in neurogenesis in *C.elegans*, regulating and forming specific nerve tracts, although this potential role of the Chn or CS chain is unclear (Ramirez-Suarez et al. 2019). Regardless of the *in vivo* situation, future structural studies, using site-specific sequencing of longer GAG chains, will likely determine which core proteins (all or a subset) that indeed carry sulfated structures. We recently showed site-specific sequencing of longer chains in perlecan (8-mer and 10-mers), indicating that a similar approach is feasible also for CSPGs/CPGs (Noborn et al. 2016).

7 Attachment Motifs in *C.elegans* and Humans

The composition and sequence of certain amino acid in defined motifs influence whether a given serine residue is selected for GAG-biosynthesis. This attachment motif was originally observed for vertebrates core proteins and may assist in the prediction of potential GAG-sites (Esko and Zhang 1996; Zhang and Esko 1994). Large scale analysis of attachment motifs in invertebrates is still lacking and it is unknown to which degree invertebrate motifs conform to the vertebrate counterpart. We prepared a frequency plot of the neighboring amino acids in the region from -9 to +9 of the glycosylated serine residue in *C. elegans*. As a comparison we aligned 20 human CS-sites that we previously identified in human urine and CSF, identified with the same analytical procedure (Noborn et al. 2015; Noborn et al. 2018). The analysis showed that the *C.elegans*-attachment motif was similar to the vertebrate counterpart, although with certain exceptions. In both species, the glycosylated serine residue was characteristically flanked by a glycine residue in the C-terminal direction and acidic residues were present in proximity to the attachment site. However, a more stringent motif was seen in *C.*

elegans in the immediate N-terminal direction. A large portion of the sequences (80%) had “Glu” or “Asp” at the -2 position and “Gly” or “Ala” at the -1 position ([ED] – [GA] – S – G). Two vertebrate xylosyltransferases (XT-I and XT-II) have been identified, whereas only a single xylosyltransferase has been found in the nematode (Wilson 2004). One may speculate that the less stringent motif in humans reflect the activities of two different xylosyltransferases, each with slightly different specificities with regard to the amino acid motifs that are required for the enzymes to bind and initiate the first step in the GAG-linkage region. The mouse XT-I and XT-II display different tissue-specific expression pattern: XT-I is highly expressed in mouse testis, kidney, and brain, while XT-II is highly expressed in mouse liver (Ponighaus et al. 2007). Our frequency plot of the human motif was based on CS-sites found in both urine and CSF thereby representing a mixture of CSPGs from different tissues. Preparing separate plots based on which tissue the CSPGs derives from, different CS-attachment motifs may emerge, which would probably represent differences in XT-I and XT-II specificities.

The attachment motif ([ED] – [GA] – S – G) defined in *C.elegans* was further used to investigate if additional potential CPGs may be present in the nematode. A search against the Swiss-prot data base for sequences containing this motif resulted in the identification of 19 additional potential CPGs, indicating that the chondroitin glycoproteome in *C.elegans* may probably expand even further with future studies (Noborn et al. 2018). Notably, since Swiss-prot is a curated data base, additional hits may be retrieved when searches are made against a more general data base, such the NCBI protein database. Nevertheless, additional CPGs are likely to be identified and this bioinformatic strategy may be useful for identifying potential CPGs/CSPGs also in other model organisms, such as *Danio rerio* and *Drosophila melanogaster*.

Inspection of the attachment motifs in relation to the functional domains, demonstrated that all motifs were present in disordered regions of the core proteins. A similar observation was made for

mucin-type O-glycans (King et al. 2017), suggesting that glycosylation in disordered regions is a general phenomenon in metazoan organisms. Further, some of the attachment motifs in our study of *C.elegans* were found in close proximity to a functional domain (e.g. on Papilin/ CPG-17), while others were found in disordered regions distant, in the primary sequence, from any functional domains (e.g. on FiBrilliN homologue/CPG-16). It is unclear how the distance to a functional domain affects the specificity of the xylosyltransferase at a given attachment motif. In vertebrates, several proteoglycans have been identified that have a time-dependent presence of GAGs, so-called part-time proteoglycans which vary their degree of occupancy at specific sites (Iozzo and Schaefer 2015; Nadanaka et al. 1998; Aono et al. 2004; Oohira et al. 2004). Sometimes, this is regulated by the synthesis of splice variants lacking or presenting a GAG-attachment motif (Wight 2002; Pangalos et al. 1995). However, one may speculate that positioning of the attachment motif and its distance to a functional domain, may influence the efficiency of the biosynthesis and thereby contribute to the glycosylation heterogeneity seen in proteoglycans.

8 Evolutionary Aspects when Comparing Chondroitin Sulfate Proteoglycans (CSPGs) and Heparan Sulfate Proteoglycans (HSPGs)

Several HSPGs in *C.elegans* display homology to vertebrate core proteins. In line with these findings, neurexin, which was recently defined as a HSPG in mouse brain tissue, also displays a high degree of similarity between distant species (Zhang et al. 2018). The HS site is conserved in all vertebrate neurexin genes from zebrafish to human. *C.elegans* also contains a homologue to the vertebrate neurexin gene, corroborating the notion that HSPGs are highly conserved throughout evolution. Although the primary sequence of neurexin is more divergent in *C.elegans*, the nematode has a consensus HS site approximately in

the same region as that of the mouse protein (Zhang et al. 2018). Other HSPG core proteins in *C. elegans* also display this degree of similarity to vertebrate HSPGs. For instance, mouse perlecan has three SG repeats in close proximity to the N-terminal domain (62-DDASGDGLGSGDVGSGDFQMVFYFR-85), all of which are modified with HS (Noborn et al. 2016). The nematode-homologue (*unc-52*) has also several potential HS attachment sites in the primary sequence, but none of these is located in the N-terminal domain. As mentioned previously, the large majority of CPG core protein in *C. elegans* do not display homology to vertebrate counterparts. However, we found a chondroitin modification on CLE-1A protein/ CPG-10 (Q9U9K7) which display homology to the human CSPG collagen α -1 (XV) chain (P39059). These proteins display a high degree of sequence similarity regarding functional domains and their order of organization. However, the Chn or CS attachment site is different in the nematode compared to the human protein, as well as the sequence and composition of amino acids surrounding the attachment site, thus principally displaying a similar degree of conservation as found for HSPGs. Furthermore, we recently identified several novel human CSPGs in tissue samples, that had previously been defined as prohormones (Noborn et al. 2015). Cholecystokinin, a peptide hormone of the gut and central nervous system, was found to be modified with CS in its propeptide region. Alignment of mammalian cholecystokinins shows a relatively low degree of sequence homology for the CS-site. For instance, the sequences of mouse and cat contain a proline instead of a serine residue at the attachment site, thereby excluding the possibility of CS-modification. Taken together, this indicates that certain CPGs/CSPGs are conserved throughout evolution to the same extent as HSPGs, whereas others display a very short evolutionary history.

One might question why HSPGs are generally more conserved throughout evolution compared with CSPGs. The difference in conservation may reflect differences in physiological functions, as HSPGs and CSPGs often induce opposite effects

on similar cellular events. In neurogenesis, CS and HS have a dual mode of action for regulating neuronal outgrowth, where both GAGs compete for the same binding sites on RPTP σ -receptors. CS chains inhibit nerve regeneration upon binding to RPTP σ -receptors, whereas HS promotes nerve regeneration upon binding to the same receptors (Shen et al. 2009; Coles et al. 2011; Katagiri et al. 2018; Sakamoto et al. 2019). Given this proteoglycan-switch, it is conceivable that HSPGs work in strict regulation with other promoting factors to navigate the growing axon along a precisely defined path. In contrast, CSPGs, which have a negative regulatory role, may be less specific in its action, providing foremost an outer perimeter for the process. A clinical example is the potential use of chondroitinase ABC in the treatment of spinal cord injury. At the injured site, axons fail to regenerate due to the formation of a glial scar, which is composed of extracellular matrix components including CSPGs (Bradbury et al. 2002). Intrathecal administration of chondroitinase ABC degrades the CS chains and thereby increases neuronal plasticity (Hu et al. 2018). Therefore, CSPG-mediated functions may display less stringent spatiotemporal requirements compared to HSPGs. To exert a particular CSPG-mediated function a CS chain is likely necessary, but its exact attachment site along the complete amino acid sequence, or even the exact identity of the core protein may have less importance, as long as the CS chain is presented in its functional context. This would impose an evolutionary selection pressure to conserve the mechanisms for CS biosynthesis and attachment motifs, but not to the same extent to a particular core protein.

A wide range of microbial pathogens uses GAG-specific interactions for their adhesion to host tissues and invasion of target cells (Bartlett and Park 2010). In nature, *C. elegans* is found in microbe-rich environments, such as rotting plant matter, containing a multitude of microbial antagonists to the nematode (Schulenburg and Felix 2017). As parasites and pathogens reduce host fitness they often impose high selective pressure on their hosts. The nematode's natural biotic environment has therefore been suggested to have

strong impact on *C.elegans* evolution and of great importance for understanding its biology (Schulenburg and Felix 2017). Given that GAGs serve as an entry point for different pathogens, changes in the underlying genomic characteristics to introduce additional chondroitin-attachment motifs on different core proteins, may have served as a strategy to evade infections for *C.elegans* throughout evolution. A more divergent chondroitin glycoproteome may present more ‘decoy sites’ for chondroitin-binding pathogens, thereby reducing pathogen attachment and entry to specific target cells. Indeed, the complexity of glycans has been suggested to be driven by an evolutionary arms race due to the exploitation of host glycans by parasites and pathogens (Gagneux et al. 2015). One may speculate that other invertebrates, whose natural habitats present lower microbe-induced selective pressure, would have less CSPGs. Regardless of the *in vivo* situation, full appreciation of the functional roles and evolutionary perspectives of CPGs/CSPGs warrants further studies in *C.elegans* and in other invertebrates. Taken together, our findings suggest that several aspects regarding chondroitin and chondroitin sulfate proteoglycan biosynthesis are conserved throughout evolution. This includes the glycosylation motif, the mechanisms for saccharide initiation and polymerization and in some cases also the splicing and the presentation of core protein domains. However, since the majority of core proteins seems not to be conserved between the species, our findings point to both converging and diverging selective forces during the proteoglycan evolution.

9 Conclusions

Our use of a novel glycoproteomic method for identifying CS-glycopeptides enabled the identification of several novel core proteins in *C.elegans* and in humans. Bioinformatic analysis of the primary amino acid sequence revealed great structural and also functional diversity of CPGs in the nematode and indicates that some, but not all, functions overlap with those of human CSPGs. Moreover, three of the novel core

proteins display homology to vertebrate counterparts, indicating that CPG / CSPGs may be more conserved throughout evolution than previously perceived. The future use of similar glycoproteomic strategies may thus be helpful in identifying CPG / CSPGs also in other important model organisms, such as *Drosophila melanogaster* and *Danio rerio*. This will likely expand the number of identified core proteins and may also provide new perspectives on proteoglycan-mediated functions and how these have persisted or developed throughout evolution. Further, obtaining global information on attachment sites and core protein identities will likely assist in assigning CPG/CSPG specific functions, both in vertebrates and in invertebrates. In addition, novel methods to site-specifically analyze the structures of extended CS chains may also be important to better understand the structure-function relationship of CPG/CSPG-mediated functions.

Acknowledgements The work was supported by grants from the Swedish Research Council (K2014-08266, 2017-00955) and the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement (ALFGBG_721971). We would also like to thank Lena Kjellén, Hiroshi Nakato, Tabea Dierker, Alejandro Gomez Toledo, Waqas Nasir, Andrea Persson and Jonas Nilsson for fruitful discussions.

References

- Ackley BD, Crew JR, Elamaa H, Pihlajaniemi T, Kuo CJ, Kramer JM (2001) The NC1/endostatin domain of Caenorhabditis elegans type XVIII collagen affects cell migration and axon guidance. *J Cell Biol* 152 (6):1219–1232
- Ackley BD, Kang SH, Crew JR, Suh C, Jin Y, Kramer JM (2003) The basement membrane components nidogen and type XVIII collagen regulate organization of neuromuscular junctions in *Caenorhabditis elegans*. *J Neurosci* 23(9):3577–3587
- Adams JC, Tucker RP (2000) The thrombospondin type 1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development. *Dev Dyn* 218 (2):280–299. [https://doi.org/10.1002/\(SICI\)1097-0177\(200006\)218:2<280::AID-DVDY4>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0177(200006)218:2<280::AID-DVDY4>3.0.CO;2-0)
- Antoshechkin I, Sternberg PW (2007) The versatile worm: genetic and genomic resources for *Caenorhabditis elegans* research. *Nat Rev Genet* 8(7):518–532. <https://doi.org/10.1038/nrg2105>

- Aono S, Tokita Y, Shuo T, Yamauchi S, Matsui F, Nakanishi K, Hirano K, Sano M, Oohira A (2004) Glycosylation site for chondroitin sulfate on the neural part-time proteoglycan, neuroglycan C. *J Biol Chem* 279(45):46536–46541. <https://doi.org/10.1074/jbc.M403263200>
- Baeg GH, Lin X, Khare N, Baumgartner S, Perrimon N (2001) Heparan sulfate proteoglycans are critical for the organization of the extracellular distribution of Wingless. *Development* 128(1):87–94
- Bartlett AH, Park PW (2010) Proteoglycans in host-pathogen interactions: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med* 12:e5. <https://doi.org/10.1017/S1462399409001367>
- Bernfield M, Gotte M, Park PW, Reizes O, Fitzgerald ML, Lincecum J, Zako M (1999) Functions of cell surface heparan sulfate proteoglycans. *Annu Rev Biochem* 68:729–777. <https://doi.org/10.1146/annurev.biochem.68.1.729>
- Beminsone P, Hwang HY, Zemtseva I, Horvitz HR, Hirschberg CB (2001) SQV-7, a protein involved in *Caenorhabditis elegans* epithelial invagination and early embryogenesis, transports UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose. *Proc Natl Acad Sci U S A* 98(7):3738–3743. <https://doi.org/10.1073/pnas.061593098>
- Bishop JR, Schuksz M, Esko JD (2007) Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* 446(7139):1030–1037. <https://doi.org/10.1038/nature05817>
- Blanchette CR, Perrat PN, Thackeray A, Benard CY (2015) Glypican is a modulator of Netrin-mediated axon guidance. *PLoS Biol* 13(7):e1002183. <https://doi.org/10.1371/journal.pbio.1002183>
- Blanchette CR, Thackeray A, Perrat PN, Hekimi S, Benard CY (2017) Functional requirements for Heparan sulfate biosynthesis in morphogenesis and nervous system development in *C. elegans*. *PLoS Genet* 13(1):e1006525. <https://doi.org/10.1371/journal.pgen.1006525>
- Bourdon MA, Oldberg A, Pierschbacher M, Ruoslahti E (1985) Molecular cloning and sequence analysis of a chondroitin sulfate proteoglycan cDNA. *Proc Natl Acad Sci U S A* 82(5):1321–1325. <https://doi.org/10.1073/pnas.82.5.1321>
- Bradbury EJ, Moon LD, Popat RJ, King VR, Bennett GS, Patel PN, Fawcett JW, SB MM (2002) Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature* 416(6881):636–640. <https://doi.org/10.1038/416636a>
- Bulik DA, Wei G, Toyoda H, Kinoshita-Toyoda A, Waldrip WR, Esko JD, Robbins PW, Selleck SB (2000) sqv-3, -7, and -8, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc Natl Acad Sci U S A* 97(20):10838–10843. <https://doi.org/10.1073/pnas.97.20.10838>
- Chang YH, Sun YH (2014) Carrier of Wingless (Cow), a secreted heparan sulfate proteoglycan, promotes extracellular transport of Wingless. *PLoS One* 9(10):e111573. <https://doi.org/10.1371/journal.pone.0111573>
- Coles CH, Shen Y, Tenney AP, Siebold C, Sutton GC, Lu W, Gallagher JT, Jones EY, Flanagan JG, Aricescu AR (2011) Proteoglycan-specific molecular switch for RPTPsigma clustering and neuronal extension. *Science* 332(6028):484–488. <https://doi.org/10.1126/science.1200840>
- Consortium CeS (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282(5396):2012–2018
- Couchman JR, Pataki CA (2012) An introduction to proteoglycans and their localization. *J Histochem Cytochem* 60(12):885–897. <https://doi.org/10.1369/0022155412464638>
- Csoka AB, Stern R (2013) Hypotheses on the evolution of hyaluronan: a highly ironic acid. *Glycobiology* 23(4):398–411. <https://doi.org/10.1093/glycob/cws218>
- Dierker T, Shao C, Haitina T, Zaia J, Hinas A, Kjellen L (2016) Nematodes join the family of chondroitin sulfate-synthesizing organisms: identification of an active chondroitin sulfotransferase in *Caenorhabditis elegans*. *Sci Rep* 6:34662. <https://doi.org/10.1038/srep34662>
- Esko JD, Lindahl U (2001) Molecular diversity of heparan sulfate. *J Clin Invest* 108(2):169–173. <https://doi.org/10.1172/JCI13530>
- Esko JD, Zhang L (1996) Influence of core protein sequence on glycosaminoglycan assembly. *Curr Opin Struct Biol* 6(5):663–670
- Gagneux P, Aebi M, Varki A (2015) Evolution of glycan diversity. In: Varki A, Cummings RD et al (eds) *Essentials of Glycobiology*, Cold Spring Harbor, pp 253–264. <https://doi.org/10.1101/glycobiology.3e.020>
- Gesteira TF, Coulson-Thomas VJ, Ogata FT, Farias EH, Cavalheiro RP, de Lima MA, Cunha GL, Nakayasu ES, Almeida IC, Toma L, Nader HB (2011) A novel approach for the characterisation of proteoglycans and biosynthetic enzymes in a snail model. *Biochim Biophys Acta* 1814(12):1862–1869. <https://doi.org/10.1016/j.bbapap.2011.07.024>
- Hascall VC, Wang A, Tammi M, Oikari S, Tammi R, Passi A, Vigetti D, Hanson RW, Hart GW (2014) The dynamic metabolism of hyaluronan regulates the cytosolic concentration of UDP-GlcNAc. *Matrix Biol* 35:14–17. <https://doi.org/10.1016/j.matbio.2014.01.014>
- Herman T, Horvitz HR (1999) Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. *Proc Natl Acad Sci U S A* 96(3):974–979. <https://doi.org/10.1073/pnas.96.3.974>
- Herman T, Hartweg E, Horvitz HR (1999) sqv mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc Natl Acad Sci U S A* 96(3):968–973. <https://doi.org/10.1073/pnas.96.3.968>
- Hrus A, Lau G, Hutter H, Schenk S, Ferralli J, Brown-Luedi M, Chiquet-Ehrismann R, Canevascini S (2007) *C. elegans* agrin is expressed in pharynx, IL1 neurons and distal tip cells and does not genetically interact

- with genes involved in synaptogenesis or muscle function. *PLoS One* 2(8):e731. <https://doi.org/10.1371/journal.pone.0000731>
- Hu HZ, Granger N, Pai SB, Bellamkonda RV, Jeffery ND (2018) Therapeutic efficacy of microtubule-embedded chondroitinase ABC in a canine clinical model of spinal cord injury. *Brain* 141(4):1017–1027. <https://doi.org/10.1093/brain/awy007>
- Hutter H, Vogel BE, Plenefisch JD, Norris CR, Proenca RB, Spieth J, Guo C, Mastwal S, Zhu X, Scheel J, Hedgecock EM (2000) Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science* 287(5455):989–994. <https://doi.org/10.1126/science.287.5455.989>
- Hwang HY, Horvitz HR (2002) The *Caenorhabditis elegans* vulval morphogenesis gene *sqv-4* encodes a UDP-glucose dehydrogenase that is temporally and spatially regulated. *Proc Natl Acad Sci U S A* 99(22):14224–14229. <https://doi.org/10.1073/pnas.172522499>
- Hwang HY, 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(14):11735–11738. <https://doi.org/10.1074/jbc.C200518200>
- Iozzo RV, Schaefer L (2015) Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. *Matrix Biol* 42:11–55. <https://doi.org/10.1016/j.matbio.2015.02.003>
- Izumikawa T, Kitagawa H, Mizuguchi S, Nomura KH, Nomura K, Tamura J, Gengyo-Ando K, Mitani S, Sugahara K (2004) Nematode chondroitin polymerizing factor showing cell-/organ-specific expression is indispensable for chondroitin synthesis and embryonic cell division. *J Biol Chem* 279(51):53755–53761. <https://doi.org/10.1074/jbc.M409615200>
- Izumikawa T, Dejima K, Watamoto Y, Nomura KH, Kanaki N, Rikitake M, Tou M, Murata D, Yanagita E, Kano A, Mitani S, Nomura K, Kitagawa H (2016) Chondroitin 4-O-Sulfotransferase is indispensable for Sulfation of chondroitin and plays an important role in maintaining Normal life span and oxidative stress responses in nematodes. *J Biol Chem* 291(44):23294–23304. <https://doi.org/10.1074/jbc.M116.757328>
- Joshi HJ, Narimatsu Y, Schjoldager KT, Tytgat HLP, Aebi M, Clausen H, Halim A (2018) SnapShot: O-Glycosylation pathways across kingdoms. *Cell* 172(3):632–632 e632. <https://doi.org/10.1016/j.cell.2018.01.016>
- Karamanos KK, Aletras AJ, Antonopoulos CA, Hjerpe A, Tsiganos CP (1990) Chondroitin proteoglycans from squid skin. Isolation, characterization and immunological studies. *Eur J Biochem* 192(1):33–38. <https://doi.org/10.1111/j.1432-1033.1990.tb19191.x>
- Katagiri Y, Morgan AA, Yu P, Bangayan NJ, Junka R, Geller HM (2018) Identification of novel binding sites for heparin in receptor protein-tyrosine phosphatase (RPTPsigma): implications for proteoglycan signaling. *J Biol Chem* 293(29):11639–11647. <https://doi.org/10.1074/jbc.RA118.003081>
- King SLJH, Schjoldager KT, Halim A, Madsen TD, Dzielgiel MH, Woetmann A, Vakhrushev SY, Wandall H (2017) Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Adv* 1(7):429–442
- Kjellen L, Lindahl U (2018) Specificity of glycosaminoglycan-protein interactions. *Curr Opin Struct Biol* 50:101–108. <https://doi.org/10.1016/j.sbi.2017.12.011>
- Kjellen L, Pettersson I, Lillhager P, Steen ML, Pettersson U, Lehtonen P, Karlsson T, Ruoslahti E, Hellman L (1989) Primary structure of a mouse mastocytoma proteoglycan core protein. *Biochem J* 263(1):105–113. <https://doi.org/10.1042/bj2630105>
- Kreuger J, Spillmann D, Li JP, Lindahl U (2006) Interactions between heparan sulfate and proteins: the concept of specificity. *J Cell Biol* 174(3):323–327. <https://doi.org/10.1083/jcb.200604035>
- Le Jan S, Hayashi M, Kasza Z, Eriksson I, Bishop JR, Weibrecht I, Heldin J, Holmborn K, Jakobsson L, Soderberg O, Spillmann D, Esko JD, Claesson-Welsh L, Kjellen L, Kreuger J (2012) Functional overlap between chondroitin and heparan sulfate proteoglycans during VEGF-induced sprouting angiogenesis. *Arterioscler Thromb Vasc Biol* 32(5):1255–1263. <https://doi.org/10.1161/ATVBAHA.111.240622>
- Ledin J, Staatz W, Li JP, Gotte M, Selleck S, Kjellen L, Spillmann D (2004) Heparan sulfate structure in mice with genetically modified heparan sulfate production. *J Biol Chem* 279(41):42732–42741. <https://doi.org/10.1074/jbc.M405382200>
- Li F, Shi W, Capurro M, Filmus J (2011) Glypican-5 stimulates rhabdomyosarcoma cell proliferation by activating Hedgehog signaling. *J Cell Biol* 192(4):691–704. <https://doi.org/10.1083/jcb.201008087>
- Lindahl U (2014) A personal voyage through the proteoglycan field. *Matrix Biol* 35:3–7. <https://doi.org/10.1016/j.matbio.2014.01.001>
- Lindahl U, Couchman J, Kimata K, Esko JD (2015) Proteoglycans and Sulfated Glycosaminoglycans. In: Varki A, Cummings RD et al (eds) *Essentials of glycobiology*, Cold Spring Harbor, pp 207–221. <https://doi.org/10.1101/glycobiology.3e.017>
- Ly M, Leach FE 3rd, Laremore TN, Toida T, Amster IJ, Linhardt RJ (2011) The proteoglycan bikunin has a defined sequence. *Nat Chem Biol* 7(11):827–833. <https://doi.org/10.1038/nchembio.673>
- Maduro MF (2017) Gut development in *C. elegans*. *Semin Cell Dev Biol* 66:3–11. <https://doi.org/10.1016/j.semcdb.2017.01.001>
- Mizumoto S, Ikegawa S, Sugahara K (2013) Human genetic disorders caused by mutations in genes encoding biosynthetic enzymes for sulfated glycosaminoglycans. *J Biol Chem* 288(16):10953–10961. <https://doi.org/10.1074/jbc.R112.437038>
- Mizumoto S, Yamada S, Sugahara K (2015) Molecular interactions between chondroitin-dermatan sulfate and growth factors/receptors/matrix proteins. *Curr Opin*

- Struct Biol 34:35–42. <https://doi.org/10.1016/j.sbi.2015.06.004>
- Muir H (1958) The nature of the link between protein and carbohydrate of a chondroitin sulphate complex from hyaline cartilage. *Biochem J* 69(2):195–204. <https://doi.org/10.1042/bj0690195>
- Mulloy B, Hart GW, Stanley P (2009) Structural analysis of Glycans. In: Varki A, Cummings RD et al (eds) *Essentials of Glycobiology*, Cold Spring Harbor
- Murdoch AD, Iozzo RV (1993) Perlecan: the multidomain heparan sulphate proteoglycan of basement membrane and extracellular matrix. *Virchows Arch A Pathol Anat Histopathol* 423(4):237–242. <https://doi.org/10.1007/bf01606885>
- Nadanaka S, Kitagawa H, Sugahara K (1998) Demonstration of the immature glycosaminoglycan tetrasaccharide sequence GlcAbeta1-3Galbeta1-3Galbeta1-4Xyl on recombinant soluble human alpha-thrombomodulin. An oligosaccharide structure on a “part-time” proteoglycan. *J Biol Chem* 273(50):33728–33734. <https://doi.org/10.1074/jbc.273.50.33728>
- Nakato H, Li JP (2016) Functions of Heparan sulfate proteoglycans in development: insights from *Drosophila* models. *Int Rev Cell Mol Biol* 325:275–293. <https://doi.org/10.1016/bs.ircmb.2016.02.008>
- Nakato H, Futch TA, Selleck SB (1995) The division abnormally delayed (dally) gene: a putative integral membrane proteoglycan required for cell division patterning during postembryonic development of the nervous system in *Drosophila*. *Development* 121(11):3687–3702
- Nasir W, Toledo AG, Noborn F, Nilsson J, Wang M, Bandeira N, Larson G (2016) SweetNET: a bioinformatics workflow for Glycopeptide MS/MS spectral analysis. *J Proteome Res* 15(8):2826–2840. <https://doi.org/10.1021/acs.jproteome.6b00417>
- Nilsson J, Halim A, Grahn A, Larson G (2013) Targeting the glycoproteome. *Glycoconj J* 30(2):119–136. <https://doi.org/10.1007/s10719-012-9438-6>
- Noborn F, Gomez Toledo A, Sihlbom C, Lengqvist J, Fries E, Kjellen L, Nilsson J, Larson G (2015) Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. *Mol Cell Proteomics* 14(1):41–49. <https://doi.org/10.1074/mcp.M114.043703>
- Noborn F, Gomez Toledo A, Green A, Nasir W, Sihlbom C, Nilsson J, Larson G (2016) Site-specific identification of heparan and chondroitin sulfate glycosaminoglycans in hybrid proteoglycans. *Sci Rep* 6:34537. <https://doi.org/10.1038/srep34537>
- Noborn F, Gomez Toledo A, Nasir W, Nilsson J, Dierker T, Kjellen L, Larson G (2018) Expanding the chondroitin glycoproteome of *Caenorhabditis elegans*. *J Biol Chem* 293(1):379–389. <https://doi.org/10.1074/jbc.M117.807800>
- 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(6):985–994. <https://doi.org/10.1083/jcb.200603003>
- Olson SK, Greenan G, Desai A, Muller-Reichert T, Oegema K (2012) Hierarchical assembly of the egg-shell and permeability barrier in *C. elegans*. *J Cell Biol* 198(4):731–748. <https://doi.org/10.1083/jcb.201206008>
- Oohira A, Shuo T, Tokita Y, Nakanishi K, Aono S (2004) Neuroglycan C, a brain-specific part-time proteoglycan, with a particular multidomain structure. *Glycoconj J* 21(1–2):53–57. <https://doi.org/10.1023/B:GLYC.0000043748.90896.83>
- Pangalos MN, Efthimiopoulos S, Shioi J, Robakis NK (1995) The chondroitin sulfate attachment site of appican is formed by splicing out exon 15 of the amyloid precursor gene. *J Biol Chem* 270(18):10388–10391. <https://doi.org/10.1074/jbc.270.18.10388>
- Park Y, Rangel C, Reynolds MM, Caldwell MC, Johns M, Nayak M, Welsh CJ, McDermott S, Datta S (2003) *Drosophila* perlecan modulates FGF and hedgehog signals to activate neural stem cell division. *Dev Biol* 253(2):247–257. [https://doi.org/10.1016/s0012-1606\(02\)00019-2](https://doi.org/10.1016/s0012-1606(02)00019-2)
- Pedersen ME, Snieckute G, Kagias K, Nehammer C, Multhaupt HA, Couchman JR, Pocock R (2013) An epidermal microRNA regulates neuronal migration through control of the cellular glycosylation state. *Science* 341(6152):1404–1408. <https://doi.org/10.1126/science.1242528>
- Ponighaus C, Ambrosius M, Casanova JC, Prante C, Kuhn J, Esko JD, Kleesiek K, Gotting C (2007) Human xylosyltransferase II is involved in the biosynthesis of the uniform tetrasaccharide linkage region in chondroitin sulfate and heparan sulfate proteoglycans. *J Biol Chem* 282(8):5201–5206. <https://doi.org/10.1074/jbc.M611665200>
- Ramirez-Suarez NJ, Belalcazar HM, Salazar CJ, Beyaz B, Raja B, Nguyen KCQ, Celestrin K, Fredens J, Faergeman NJ, Hall DH, Bulow HE (2019) Axon-independent patterning and maintenance of somatosensory dendritic Arbors. *Dev Cell* 48(2):229–244 e224. <https://doi.org/10.1016/j.devcel.2018.12.015>
- Roden L, Smith R (1966) Structure of the neutral trisaccharide of the chondroitin 4-sulfate-protein linkage region. *J Biol Chem* 241(24):5949–5954
- Rogalski TM, Williams BD, Mullen GP, Moerman DG (1993) Products of the unc-52 gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan. *Genes Dev* 7(8):1471–1484. <https://doi.org/10.1101/gad.7.8.1471>
- Saied-Santiago K, Bulow HE (2018) Diverse roles for glycosaminoglycans in neural patterning. *Dev Dyn* 247(1):54–74. <https://doi.org/10.1002/dvdy.24555>
- Sakamoto K, Ozaki T, Ko YC, Tsai CF, Gong Y, Morozumi M, Ishikawa Y, Uchimura K, Nadanaka S, Kitagawa H, Zulueta MML, Bandaru A, Tamura JI, Hung SC, Kadomatsu K (2019) Glycan sulfation patterns define autophagy flux at axon tip via

- PTPRsigma-cortactin axis. *Nat Chem Biol* 15 (7):699–709. <https://doi.org/10.1038/s41589-019-0274-x>
- Salanti A, Clausen TM, Agerbaek MO, Al Nakouzi N, Dahlback M, Oo HZ, Lee S, Gustavsson T, Rich JR, Hedberg BJ, Mao Y, Barington L, Pereira MA, LoBello J, Endo M, Fazli L, Soden J, Wang CK, Sander AF, Dagil R, Thrane S, Holst PJ, Meng L, Favero F, Weiss GJ, Nielsen MA, Freeth J, Nielsen TO, Zaia J, Tran NL, Trent J, Babcook JS, Theander TG, Sorensen PH, Daugaard M (2015) Targeting human Cancer by a glycosaminoglycan binding malaria protein. *Cancer Cell* 28(4):500–514. <https://doi.org/10.1016/j.ccell.2015.09.003>
- Sarrazin S, Lamanna WC, Esko JD (2011) Heparan sulfate proteoglycans. *Cold Spring Harb Perspect Biol* 3(7). <https://doi.org/10.1101/cshperspect.a004952>
- Schachter H (2004) Protein glycosylation lessons from *Caenorhabditis elegans*. *Curr Opin Struct Biol* 14 (5):607–616. <https://doi.org/10.1016/j.sbi.2004.09.005>
- Schatton J, Schubert M (1954) Isolation of a mucoprotein from cartilage. *J Biol Chem* 211:565–573
- Schulenburg H, Felix MA (2017) The natural biotic environment of *Caenorhabditis elegans*. *Genetics* 206 (1):55–86. <https://doi.org/10.1534/genetics.116.195511>
- Shao C, Shi X, White M, Huang Y, Hartshorn K, Zaia J (2013) Comparative glycomics of leukocyte glycosaminoglycans. *FEBS J* 280(10):2447–2461. <https://doi.org/10.1111/febs.12231>
- Shen Y, Tenney AP, Busch SA, Horn KP, Cuascut FX, Liu K, He Z, Silver J, Flanagan JG (2009) PTPsigma is a receptor for chondroitin sulfate proteoglycan, an inhibitor of neural regeneration. *Science* 326 (5952):592–596. <https://doi.org/10.1126/science.1178310>
- Shim YH, Paik YK (2010) *Caenorhabditis elegans* proteomics comes of age. *Proteomics* 10(4):846–857. <https://doi.org/10.1002/pmic.200900542>
- Stern R (2003) Devising a pathway for hyaluronan catabolism: are we there yet? *Glycobiology* 13(12):105R–115R. <https://doi.org/10.1093/glycob/cwg112>
- Sugiura N, Clausen TM, Shioiri T, Gustavsson T, Watanabe H, Salanti A (2016) Molecular dissection of placental malaria protein VAR2CSA interaction with a chemo-enzymatically synthesized chondroitin sulfate library. *Glycoconj J* 33(6):985–994. <https://doi.org/10.1007/s10719-016-9685-z>
- 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(4):2269–2275
- Vuong-Breder TT, Yang X, Labouesse M (2016) *C. elegans* embryonic morphogenesis. *Curr Top Dev Biol* 116:597–616. <https://doi.org/10.1016/bs.ctdb.2015.11.012>
- Weigel PH, Hascall VC, Tammi M (1997) Hyaluronan synthases. *J Biol Chem* 272(22):13997–14000. <https://doi.org/10.1074/jbc.272.22.13997>
- Weiss RJ, Esko JD, Tor Y (2017) Targeting heparin and heparan sulfate protein interactions. *Org Biomol Chem* 15(27):5656–5668. <https://doi.org/10.1039/c7ob01058c>
- Wight TN (2002) Versican: a versatile extracellular matrix proteoglycan in cell biology. *Curr Opin Cell Biol* 14 (5):617–623
- Wilson IB (2004) The never-ending story of peptide O-xylosyltransferase. *Cell Mol Life Sci* 61 (7–8):794–809. <https://doi.org/10.1007/s00018-003-3278-2>
- Wilson IBH, Cummings RD, Aebi M (2015) Nematoda. In: Varki A, Cummings RD et al (eds) *Essentials of Glycobiology*, Cold Spring Harbor, pp 321–333. <https://doi.org/10.1101/glycobiology.3e.025>
- Yamada S, Van Die I, Van den Eijnden DH, Yokota A, Kitagawa H, Sugahara K (1999) Demonstration of glycosaminoglycans in *Caenorhabditis elegans*. *FEBS Lett* 459(3):327–331. [https://doi.org/10.1016/s0014-5793\(99\)01286-7](https://doi.org/10.1016/s0014-5793(99)01286-7)
- Yamada S, Okada Y, Ueno M, Iwata S, Deepa SS, Nishimura S, Fujita M, Van Die I, Hirabayashi Y, Sugahara K (2002) Determination of the glycosaminoglycan-protein linkage region oligosaccharide structures of proteoglycans from *Drosophila melanogaster* and *Caenorhabditis elegans*. *J Biol Chem* 277(35):31877–31886. <https://doi.org/10.1074/jbc.M205078200>
- Yamada S, Sugahara K, Ozbek S (2011) Evolution of glycosaminoglycans: comparative biochemical study. *Commun Integr Biol* 4(2):150–158. <https://doi.org/10.4161/cib.4.2.14547>
- Yanagishita M (1993) A brief history of proteoglycans. *Experientia* 49(5):366–368. <https://doi.org/10.1007/bf01923581>
- Zhang L, Esko JD (1994) Amino acid determinants that drive heparan sulfate assembly in a proteoglycan. *J Biol Chem* 269(30):19295–19299
- Zhang P, Lu H, Peixoto RT, Pines MK, Ge Y, Oku S, Siddiqui TJ, Xie Y, Wu W, Archer-Hartmann S, Yoshida K, Tanaka KF, Aricescu AR, Azadi P, Gordon MD, Sabatini BL, Wong ROL, Craig AM (2018) Heparan sulfate organizes neuronal synapses through Neurexin partnerships. *Cell* 174(6):1450–1464 e1423. <https://doi.org/10.1016/j.cell.2018.07.002>
- Zhu F, Li D, Chen K (2019) Structures and functions of invertebrate glycosylation. *Open Biol* 9(1):180232. <https://doi.org/10.1098/rsob.180232>