

# Expanding the chondroitin glycoproteome of *Caenorhabditis elegans*

Received for publication, August 1, 2017, and in revised form, October 23, 2017. Published, Papers in Press, November 14, 2017. DOI 10.1074/jbc.M117.807800

✉ Fredrik Noborn<sup>‡</sup>, Alejandro Gomez Toledo<sup>‡</sup>, Waqas Nasir<sup>‡</sup>, Jonas Nilsson<sup>‡</sup>, Tabea Dierker<sup>§</sup>, Lena Kjellén<sup>§</sup>, and ✉ Göran Larson<sup>‡,1</sup>

From the <sup>‡</sup>Department of Clinical Chemistry and Transfusion Medicine, Institute of Biomedicine, University of Gothenburg, SE-413 45 Gothenburg and the <sup>§</sup>Department of Medical Biochemistry and Microbiology, Uppsala University, SE-751 23 Uppsala, Sweden

Edited by Gerald W. Hart

Chondroitin sulfate proteoglycans (CSPGs) are important structural components of connective tissues in essentially all metazoan organisms. In vertebrates, CSPGs are involved also in more specialized processes such as neurogenesis and growth factor signaling. In invertebrates, however, knowledge of CSPGs core proteins and proteoglycan-related functions is relatively limited, even for *Caenorhabditis elegans*. This nematode produces large amounts of non-sulfated chondroitin in addition to low-sulfated chondroitin sulfate chains. So far, only nine core proteins (CPGs) have been identified, some of which have been shown to be involved in extracellular matrix formation. We recently introduced a protocol to characterize proteoglycan core proteins by identifying CS-glycopeptides with a combination of biochemical enrichment, enzymatic digestion, and nano-scale liquid chromatography MS/MS analysis. Here, we have used this protocol to map the chondroitin glycoproteome in *C. elegans*, resulting in the identification of 15 novel CPG proteins in addition to the nine previously established. Three of the newly identified CPGs displayed homology to vertebrate proteins. Bioinformatics analysis of the primary protein sequences revealed that the CPG proteins altogether contained 19 unique functional domains, including Kunitz and endostatin domains, suggesting direct involvement in protease inhibition and axonal migration, respectively. The analysis of the core protein domain organization revealed that all chondroitin attachment sites are located in unstructured regions. Our results suggest that CPGs display a much greater functional and structural heterogeneity than previously appreciated and indicate that specialized proteoglycan-mediated functions evolved early in metazoan evolution.

Humans express at least 50 chondroitin sulfate proteoglycans (CSPGs),<sup>2</sup> each unique by structural differences in the

This work was supported by Swedish Research Council Grant 8266 (to G. L.) and by governmental grants to the Sahlgrenska University Hospital (to G. L.). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains Tables S1–S4.

<sup>1</sup> To whom correspondence should be addressed: Institute of Biomedicine, University of Gothenburg, Bruna Stråket 16, Sahlgrenska University Hospital, SE-413 45 Gothenburg, Sweden. Tel.: 46-0-31-342-1330; Fax: 46-0-31-82-84-58; E-mail: [goran.larson@clinchem.gu.se](mailto:goran.larson@clinchem.gu.se).

<sup>2</sup> The abbreviations used are: CSPG, chondroitin sulfate proteoglycan; nLC, nano-scale liquid chromatography; HS, heparan sulfate.

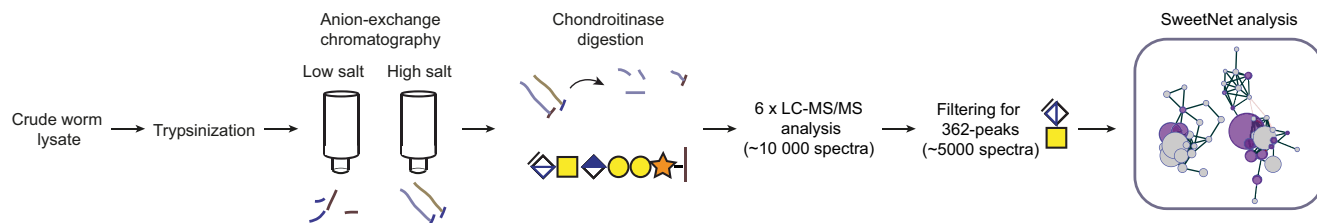
This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

amino acid sequence of the core proteins (1–3). CSPGs are important components in cartilage and other connective tissues, where they interact with fibrous proteins to provide a hydrated matrix that resists compressive forces. Apart from this role as a structural component, CSPGs also contribute to more specialized functions such as angiogenesis and neurogenesis (4, 5). Certain CSPGs, such as neurocan, are essential for neural differentiation and act as inhibitory cues for axonal outgrowth (6). Moreover, CSPGs are essential for storage and secretion of bioactive components such as proteases and pro-hormones (1, 7, 8). However, information on CSPG core protein primary structure and proteoglycan-related functions is limited and very few such studies have been performed even for the otherwise well-studied nematode *C. elegans*.

The chondroitin sulfate biosynthesis typically commences in the endoplasmic reticulum/Golgi compartments with the enzymatic transfer of a  $\beta$ -linked xylose (Xyl) to specific serine (Ser) residues of the core protein sequence (9, 10). The assembly continues with the enzymatic addition of two galactoses (Gal) and one glucuronic acid (GlcA) residue, completing the formation of a characteristic tetrasaccharide linkage region (GlcA $\beta$ 3Gal $\beta$ 3Gal $\beta$ 4Xyl $\beta$ 1Ser). The polymerization continues through the enzymatic addition of alternating units of GlcA and *N*-acetylgalactosamine (GalNAc) (11). In vertebrates, the polysaccharides undergo extensive modification by chondroitin-specific epimerases and sulfotransferases. The epimerases convert a subset of GlcA to iduronic acid (IdoA) and sulfotransferases add sulfate groups at various positions of the GlcA and GalNAc residues, generating glycosaminoglycans (GAGs) with considerable structural complexity (12). *Caenorhabditis elegans* seems to lack chondroitin-specific epimerases and so far only a single sulfotransferase has been identified (catalyzing GalNAc 4-*O*-sulfation) (13, 14). However, the presence of both 4-*O*- and 6-*O*-sulfated GalNAc residues in *C. elegans* CS indicates that at least one additional sulfotransferase is expressed by the worms (13).

The importance of chondroitin in *C. elegans* has been demonstrated by studying structural mutations, affecting genes required for chondroitin biosynthesis, which result in developmental abnormalities such as impaired vulval morphogenesis and altered neuronal migration (15–17). However, information on the core proteins involved in such processes is often scarce and so far only nine chondroitin core proteins have been iden-

## Chondroitin proteoglycans in *C. elegans*



**Figure 1. Scheme for identifying CPGs in *C. elegans*.** The work flow includes enrichment of CPGs from crude worm extract, enzymatic hydrolysis, LC-MS/MS analysis, and subsequent interpretation of mass spectra using the *SweetNET* software.

tified in *C. elegans* (CPG-1 to CPG-9) (3). Moreover, analyses of their amino acid sequence indicated that these core proteins only contain two types of functional domains, peritrophin-A chitin-binding domains in CPG-1 and CPG-2 and C-type lectin domains in CPG-5 and CPG-6 (3). Although the chitin-binding domains of CPG-1/CPG-2 are important for the assembly of the egg shell layer in the growing embryo (18), the role of C-type lectin domains on CPG-5 and CPG-6 is not yet determined. Given that vertebrate CSPGs display a wide range of functional diversity, it is likely that additional CPG core proteins may be present in *C. elegans* containing functional domains that are not only related to extracellular matrix formation, but also to accommodate more specialized functions. We argued that information on chondroitin chain attachment sites and core protein identities may be of great value to further delineate chondroitin-mediated functions in *C. elegans*.

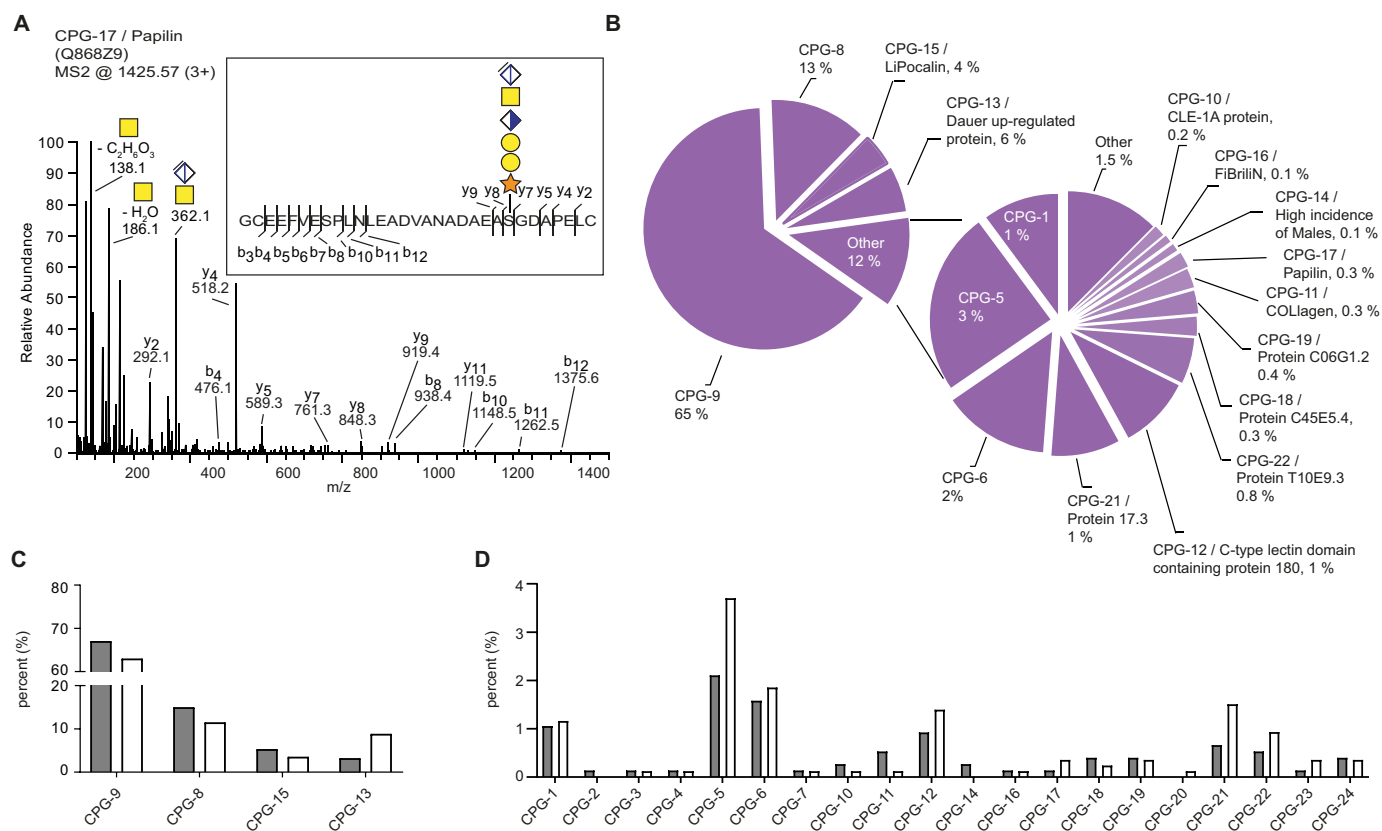
We recently introduced a method to characterize CSPGs from human tissue fluids using a combination of anion-exchange chromatography for enrichment, enzymatic digestion for reduction of the length of the CS chains, and subsequently nLC-MS/MS for structural characterization of CS glycopeptides and core proteins (1). This protocol enabled site-specific identification of both novel and established core proteins and revealed prohormones as a novel class of vertebrate proteoglycans (1). In the present work we set out to map the chondroitin glycoproteome in *C. elegans*, through analysis of the data using *SweetNET*, a recently developed bioinformatics software for glycopeptide MS/MS spectral analysis (2). Our approach resulted in the identification of the nine previously established CPG core proteins as well as 15 novel CPG core proteins, three of which displayed homology to vertebrate proteins. Bioinformatics analysis of the primary protein sequences revealed that the core proteins contained 19 unique functional domains, including calcium-binding EGF domains, immunoglobulin domains, and Kunitz domains. This suggests that CPGs display a much greater functional and structural heterogeneity than previously appreciated and indicates that specialized proteoglycan-mediated functions evolved early in metazoan evolution. Collectively, these data may assist in the efforts of finding and elucidating novel functional roles of proteoglycans both in *C. elegans* and in vertebrates.

## Results

### Identification of 15 novel chondroitin core proteins in *C. elegans*

We set out to map the chondroitin glycoproteome in *C. elegans* using our recently developed glycoproteomic approach that identifies chondroitin/CS attachment sites and provides identities of

the core proteins (1). *C. elegans* samples were collected from a population of a strain lacking two heparan sulfate sulfotransferases (*hst-6* and *hst-2*) and the material was solubilized by consecutive passages through hypodermic needles with decreasing diameters. To obtain defined chondroitin glycopeptides suitable for structural analysis, the sample was incubated with trypsin and passed over an anion exchange column that had been equilibrated with a low-salt buffer (0.2 M NaCl). This procedure enriches chondroitin glycopeptides as the positively charged matrix retains anionic polysaccharides and their attached peptides, whereas neutral or positively charged peptides flow through the column. The bound structures were eluted stepwise with three buffers of increasing sodium chloride concentrations (0.4, 0.8, and 1.5 M NaCl) and the three fractions were collected and desalted. The fractions were treated with chondroitinase ABC to reduce the complexity of the chondroitin chains that generate free disaccharides and a residual hexasaccharide structure still attached to the core protein. The residual hexasaccharide is composed of the linkage region and a  $\Delta$ GlcA-GalNAc disaccharide, dehydrated on the hexuronic acid to form  $\Delta$ hexuronic acid ( $\Delta$ HexA) (1, 19). The chondroitinase-treated fractions were analyzed with positive mode nLC-MS/MS at normalized collision energy levels set to generate abundant peptide, as well as glycosidic fragmentations, necessary for glycopeptide identification (2). The general workflow for the glycopeptide enrichment, MS/MS analysis, and the subsequent *SweetNET*-assisted analysis is illustrated in Fig. 1. In total, six data files from two different preparations were generated. As MS2-fragmentation of chondroitin glycopeptides is expected to generate prominent oxonium ions at  $m/z$  362.1, corresponding to the terminal dehydrated disaccharide structure  $[\Delta\text{HexAGalNAc}]^+$ , all spectra collected were screened for the presence of the  $m/z$  362.1 ion ( $m/z$  range 362.10–362.11). Spectra lacking this diagnostic ion were dismissed and the filtered data set was then used for molecular networking and database annotation. A single molecular network was generated and initial Mascot database searches were performed to identify chondroitin-substituted peptides (see “Experimental procedures”). The output data were integrated with the *SweetNET* platform to annotate the network with glycopeptide information. Furthermore, the distribution of precursor ion  $m/z$   $\Delta$  shifts between nodes of the network identified the presence of chemical artifacts, such as carbamidomethyl derivatization (+57 Da) and ammonium ion adducts (+17 Da). Additionally, typical  $m/z$   $\Delta$  shift of 203 Da, corresponding to HexNAc residues, were also identified. The networks were then iteratively propagated based on these observed  $m/z$   $\Delta$  shifts. All generated



**Figure 2. SweetNET-assisted identification of novel CPGs.** A, MS2 fragment spectrum of CPG-17/Papilin (O76840) ( $m/z$  1425.57; 3+). B, pie charts based on spectral counts of annotated MS/MS showing the relative abundance of each CPGs. 88% of annotated spectra were related to CPG-8 (13%), CPG-9 (65%), CPG-13/Dauer up-regulated protein (6%), or CPG-15/LiPocalin (4%) (left). The other 20 CPGs accounted together only for 12% of all the annotated spectra (right). C and D, the relative abundance of each CPG in the two sample preparations based on spectral counts of annotated MS/MS spectra (white bars, preparation 1; gray bars, preparation 2). The graphs show the relative abundance of the five most abundant CPGs (C) and the relative abundance of the other 19 low abundant CPGs (D).

hits were validated and interpreted with regard to peptide sequence, glycan structure, and precursor mass (Table S1).

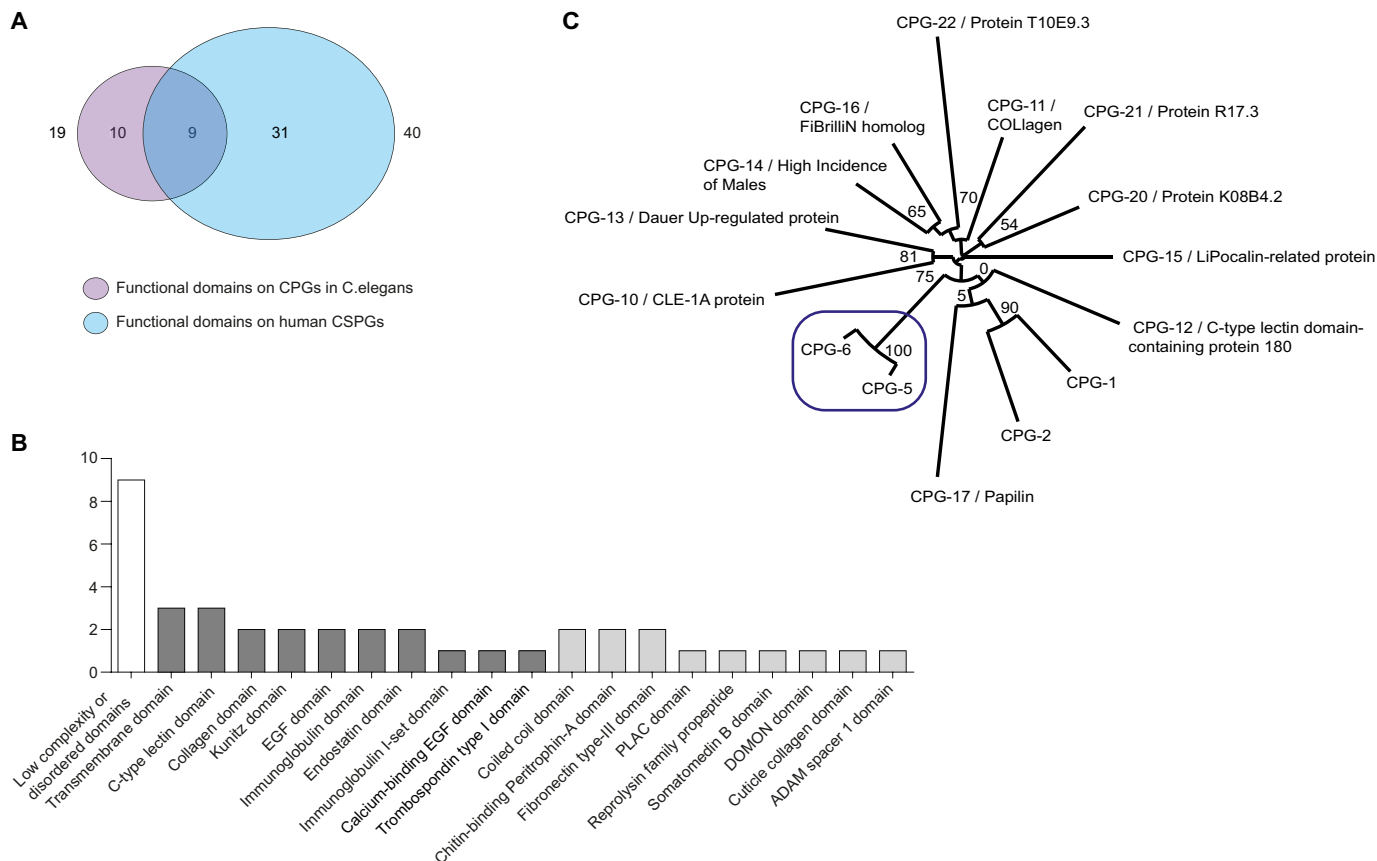
The SweetNET analysis provided annotation of several clusters in the network that corresponded to 17 different CPGs, together with seven additional CPGs that did not generate any clusters but were identified based on a single glycopeptide precursor mass for each of the core proteins (Table S1). In total, our protocol identified all the previously established core proteins (CPG-1 to CPG-9) as well as 15 novel core proteins, which were designated CPG-10 to CPG-24. Some of the novel core proteins have previously been assigned names based on sequence similarities to vertebrate proteins, such as CPG-16/Fibrillin homolog and CPG-17/Papilin, whereas others, such as CPG-22/Protein T10E9.3, are annotated in UniProt based only on the open reading frame (ORF) name. Two chondroitin glycopeptides derived from separate parts of the proteins were identified in five of the CPGs (CPG-1, CPG-8, CPG-9, CPG-12, and CPG-22). A representative MS2 spectrum of a novel CPG (CPG-17/Papilin) is shown in Fig. 2A. As expected from the filtration procedure, a prominent oxonium ion at  $m/z$  362.1 was observed corresponding to the terminal disaccharide fragment ion  $[\Delta\text{HexAGalNAc}]^+$ . Furthermore, several  $y$ - and  $b$ -ions were observed that enabled Mascot identification of the peptide sequence. Pie charts based on spectral counts of annotated MS2 scans were used to assess the abundance of each CPG (Fig. 2B). A large proportion of the annotated spectra were related to,

CPG-8 (13%), CPG-9 (65%), CPG-13/Dauer up-regulated protein (6%), and CPG-15/LiPocalin (4%) (Fig. 2B, left). The remaining CPGs accounted for only 12% of the total annotated spectra and several of the novel CPGs constituted 0.2–0.3% of the total CPG count (Fig. 2B, right). Of note, CPG-17/Papilin, which displayed distinctive glycopeptide fragmentation, as shown in Fig. 2A, constituted only about 0.3% of the total CPG count. Furthermore, to assess the variability of the relative abundance of each CPG in the two preparations, we compared the spectral counts of annotated CPGs in each of the two preparations (Fig. 2, C and D). Although some variation in relative abundance was observed for certain CPGs (e.g. CPG-5), the majority of the core proteins were found at a very similar level in the two preparations. Moreover, of the 24 identified CPGs, 21 CPGs were found in both sample preparations: only CPG-2 and CPG-14 were unique for preparation 1 (gray bars), whereas CPG-20 was unique for preparation 2 (white bars) (Fig. 2D).

#### Characteristics of the novel CPGs

Some of the CPGs now identified have previously been the focus for functional studies and are therefore relatively well-characterized. CPG-17/Papilin is a basement membrane component essential for embryonic development (20) and the CPG-16/Fibrillin homolog is required for the structural integrity of the nematode epidermal layer (21). However, other CPGs have only been predicted based on their ORF gene, and no previous

## Chondroitin proteoglycans in *C. elegans*



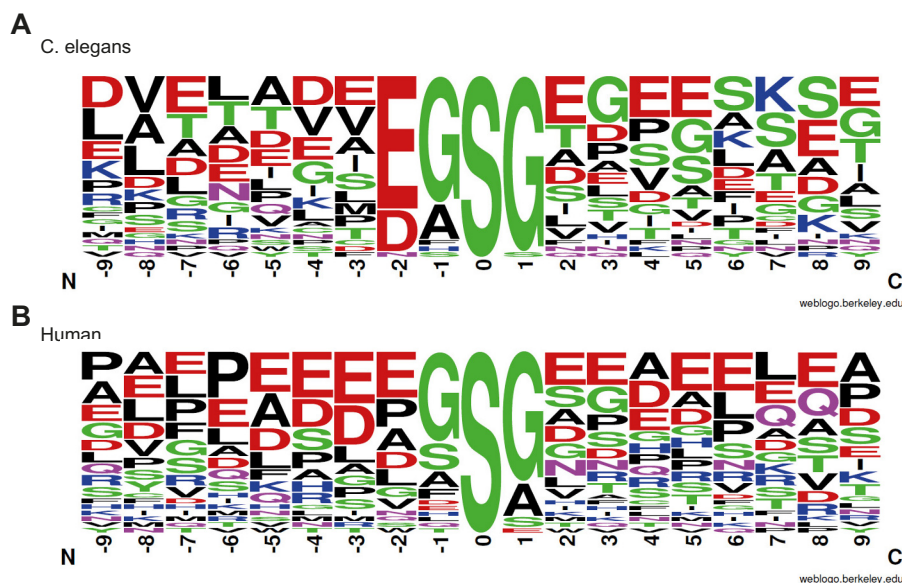
**Figure 3. Functional domains of proteoglycans in *C. elegans* and humans.** The protein sequence of each identified CPG-1 to CPG-24 was searched against the Pfam database to get information on which functional domains were present and the search identified 19 unique domains. Another search was performed for our previously identified CSPGs in humans (1, 2), providing 40 unique domains (see “Experimental procedures”). **A**, venn diagram showing the distribution of functional domains found in both species. **B**, domains found in CPGs of *C. elegans* and in CSPGs in humans are shown as *dark gray bars*, whereas functional domains unique for *C. elegans* are shown as *light gray bars*. Nine core proteins did not retrieve any hits in the Pfam database and displayed only low complexity or disordered domains (*white bar*). **C**, phylogenetic analysis based on the amino acid sequences of the core proteins that contained functional domains. The analysis was made using the Phylogeny.fr web service and *numbers* represent bootstrap support values based on 100 bootstrap replicates. The analysis shows that CPG-5 and CPG-6 are highly homologous, whereas the other core proteins do not display any apparent homologies.

experimental or translational data exist. Thus, bioinformatics analysis was employed to gain insights into the domain architecture of each CPG. Initial searches against the Pfam database resulted in the identification of 19 unique domains, distributed over 15 different core proteins (Table S2). We then compared the nematode functional domains with the functional domains present in human CSPGs. For this purpose we used our previously identified CSPGs in human tissue fluids, which were identified using the same protocol (1, 2). Pfam searches of 28 human CSPGs provided 40 unique functional domains (Table S3). Of all 50 domain structures identified, 31 were uniquely found in human CSPGs, 10 uniquely found in *C. elegans* CPGs, and 9 found in both species (Fig. 3A). Well-known domains, such as the collagen domain and the Kunitz domain, were found in both species (Fig. 3B, *dark gray bars*), whereas other domains, such as the chitin-binding peritrophin-A domain and dopamine  $\beta$ -monooxygenase domain (DOMON), were only found in the nematode (Fig. 3B, *light gray bars*). Furthermore, nine nematode core proteins did not retrieve any hits within the Pfam database but displayed only low complexity or disordered domains (Fig. 3B, *white bars*).

A complete list of the CPGs identified, including the tryptic peptide sequences identified, the functional domains, and the

predicted molecular masses is shown in Table S2. The Pfam analysis indicated that CPG-11/COLLagen contains both a collagen domain and cuticle collagen domain, suggesting its involvement in extracellular matrix formation. Kunitz domains, identified in, e.g. CPG-17/Papilin, typically acting as protease inhibitors are present also in core proteins of human CSPGs, bikunin being one example (22). Moreover, the analysis showed that the CPG-10/CLE-1A protein contains three functional domains, including an endostatin domain in the C-terminal end. Interestingly, CPG-10/CLE-1A protein accumulates at high levels in the nematode nervous system and the endostatin domain has been demonstrated to specifically regulate cell migration and axon guidance (23). Three core proteins were predicted to contain transmembrane domains, including CPG-11/COLLagen, CPG-15/LiPocalin-related protein, and CPG-16/Fibrillin homolog. Vertebrates express multiple membrane CSPGs (e.g. CD44, CSPG5) (1), but none of the nematode membrane-spanning CPGs showed homology to any vertebrate counterparts. This discrepancy in structure-function relationships may indicate that membrane spanning CPGs/CSPGs have evolved separately in different organisms.

Generally, the phylogenetic analysis based on the amino acid sequences revealed no extensive homologies between the



**Figure 4. Definition of a chondroitin attachment motif in *C. elegans*.** A, statistical analysis of aligned sequences of chondroitin attachment sites in *C. elegans* and B, of chondroitin sulfate attachment sites in humans. Sequence logos show the frequency of each amino acid in the region from  $-9$  to  $+9$  of the glycosylated serine residue. The figures were prepared using WebLogo (35).

majorities of the core proteins, suggesting that they have evolved separately throughout evolution (Fig. 3C). However, one pair of core protein homologs was identified in the nematode, indicating that co-evolution occurs in certain cases. Thus, in agreement with a previous report, we found similarity between CPG-5 and CPG-6 (3)(Fig. 3C). Furthermore, analysis of core protein molecular mass revealed a wide range of sizes, spanning from 7.1 (CPG-9) to 568 kDa (CPG 14/high incidence of males, isoform b), where 5 of the 24 core proteins displayed a molecular mass of  $>100$  kDa.

Three of the identified core proteins have previously been shown to display sequence similarity to vertebrate proteins (20, 21, 23). CPG-17/Papilin is a homolog to human Papilin and inhibition of Papilin synthesis in the nematode results in defective cell arrangement and embryonic death (20). The CPG-16/FiBrilliN homolog shows sequence similarity with vertebrate fibrillins, which are essential for formation of elastic fibers in connective tissue. Mutations of fibrillin genes in humans are linked to connective tissue diseases, including Marfan syndrome, characterized by abnormal fibrous connective tissue, which affects the ocular, cardiovascular, and skeletal systems (21). Moreover, the CPG-10/CLE-1A protein is the homolog to human collagen  $\alpha$ -1 XV/XVIII and is involved in cell migration and axon guidance (23). Taken together, these data suggest that the structural and functional diversity of CPGs in *C. elegans* are much greater than previously appreciated.

#### Definition of the chondroitin attachment motif in *C. elegans*

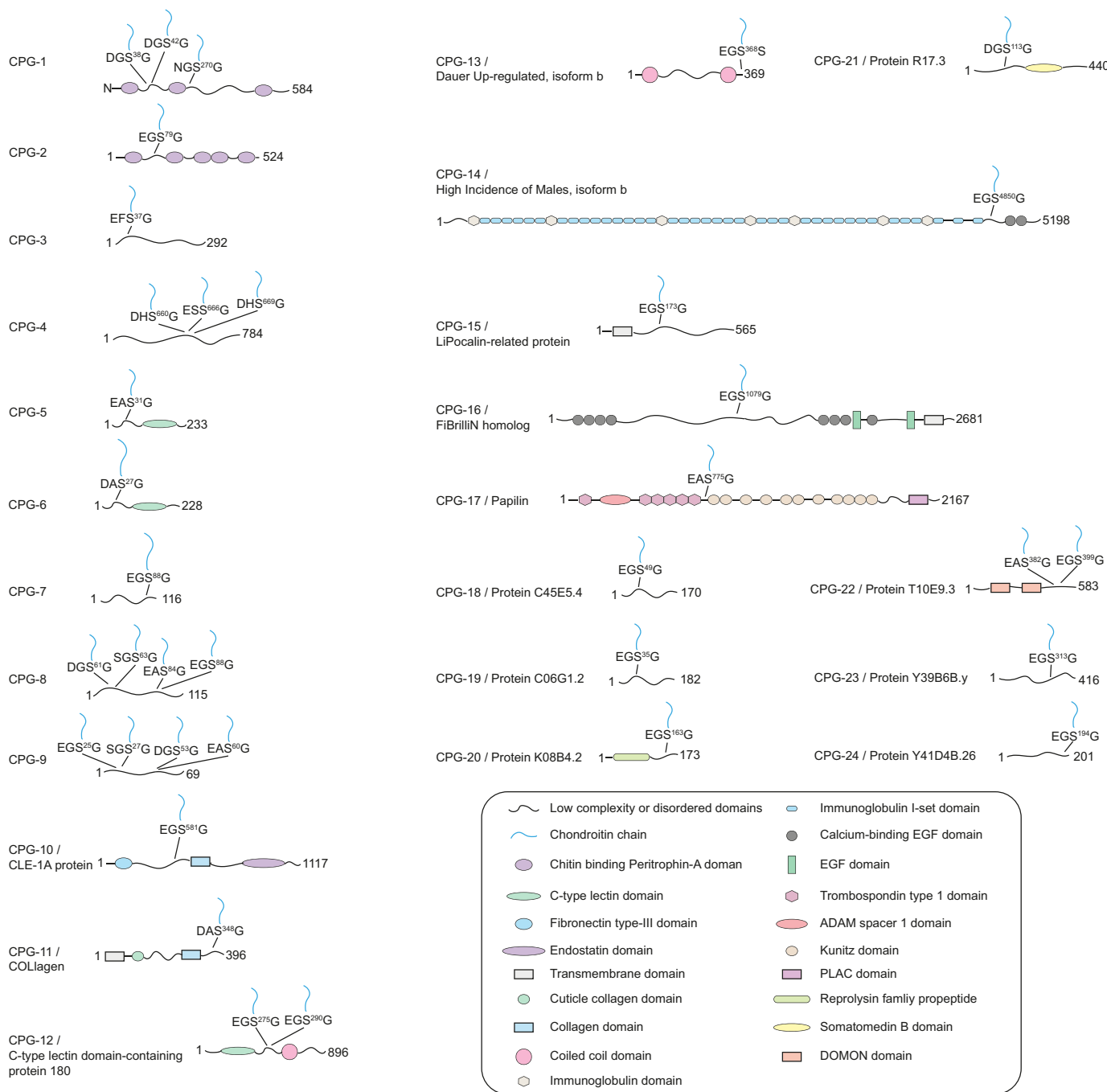
In vertebrates, certain features of the core protein seem to influence whether a certain serine residue is modified with a xylose to initiate GAG biosynthesis. The glycosylated serine residue is typically flanked by a glycine residue in the C-terminal direction and is also located close to a number of acidic residues in close proximity (10). To investigate whether the chondroitin glycosylation motifs in *C. elegans* conform to these criteria, we prepared a frequency plot of the neighboring amino

acids in the region from  $-9$  to  $+9$  of the glycosylated serine residue (Fig. 4A). As a comparison, a frequency plot of the CS attachment motifs in humans was prepared from data of our previously verified CS sites in human tissue fluids (Fig. 4B) (1, 2) (Table S3). In both species, the glycosylated serine residue is characteristically flanked by a glycine residue in the C-terminal direction and both chondroitin and chondroitin sulfate-modified sequences display a high abundance of acidic residues in both C- and N-terminal directions. However, the *C. elegans* sequence is relatively more conserved in the immediate N-terminal direction, where a large proportion of the sequences (80%) conform to “Glu” or “Asp” at the  $-2$  position and “Gly” or “Ala” at the  $-1$  position. To investigate whether additional CPGs may be present in the nematode proteome, the ScanProsite bioinformatic tool was used to search the Swiss-Prot database using the  $([ED] - [GA] - S - G)$  motif. The retrieved hits were filtered for the presence of a signal peptide and sequences without a signal peptide were excluded. These criteria retrieved the 11 identified CPGs that were annotated in the Swiss-Prot database (CPG-1 to CPG-9, CPG-12, and CPG-17) but also identified additionally 19 potential CPGs in this database (Table S4). In conclusion, although our MS analysis enabled the identification of a significant number of novel CPGs, this suggested that additional CPGs are yet to be found in *C. elegans*.

#### Mapping the chondroitin glycoproteome in *C. elegans*

Our specific and general findings are summarized in Fig. 5, showing a map of the chondroitin glycoproteome in *C. elegans*. The presence of functional domain(s), the chondroitin attachment site(s), and the attachment motif(s) for each core protein is schematically illustrated. The CPGs identified in the present study are CPGs 10 to 24. Interestingly, inspection of the domain organization for each core protein reveals that all chondroitin attachment sites are located in low complexity or in disordered domains. The chondroitin glycosylation sites were found throughout the

## Chondroitin proteoglycans in *C. elegans*



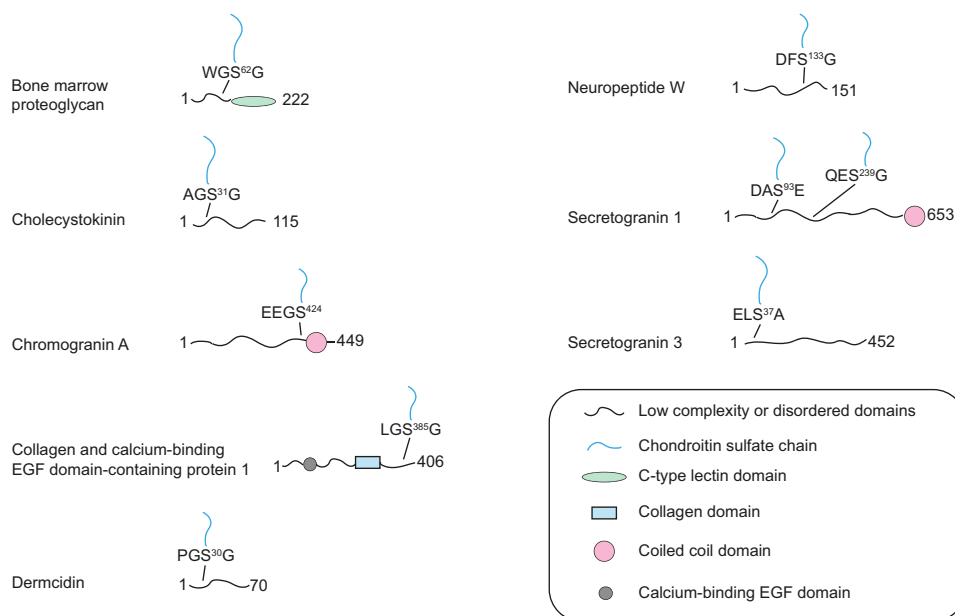
**Figure 5. Map of the chondroitin glycoproteome in *C. elegans*.** The scheme illustrates all the CPGs (CPG-1 to CPG-24) identified in the present study, where the novel CPGs identified are named CPG-10 to -24. The presence of functional domain(s), the chondroitin attachment site(s) for each core protein are shown. The key for various functional domains is provided in the box.

entire proteins, with no apparent enrichment toward the N- or C-terminal end. Furthermore, some attachment sites were found distant from a functional domain (e.g. CPG-16), whereas others were found in close proximity to a functional domain (e.g. CPG-13). Moreover, while reviewing a set of human CSPGs it was evident that the CS attachment sites in human core proteins are also located in low complexity or in disordered domains (Fig. 6). The CSPGs reviewed belong to the family of human prohormones that, similar to the *C. elegans* CPGs, contain functional domains such as C-type lectin domain, collagen domain, as well as core proteins without any functional domains.

Taken together, these data showed that nematode chondroitin and human chondroitin sulfate glycosylations are restricted to low complexity or disordered domains, suggesting that this glycosylation pattern may be a general feature of chondroitin glycosylation across species.

### Discussion

*C. elegans* has a compact genome that is well-suited for genetic manipulations and for determining the influence of each gene product in development and physiology (24). Although there has been significant progress to support our



**Figure 6. Map of chondroitin sulfate attachment sites in human prohormones.** The scheme illustrates human CSPGs that belong to the subclass of prohormones, which were identified in a previous study (1). The presence of functional domain(s), the chondroitin attachment site(s) for each core protein are shown. Inspection of the domain organization for each core protein reveals that all chondroitin sulfate attachment sites are located in low complexity or disordered domains. The key for various functional domains is provided in the box.

understanding of how different genes and proteins influence various biological processes, information on protein glycosylation in the nematode remains relatively scarce (25).

We report here the identification of 15 novel CPGs in the nematode *C. elegans*. The bioinformatics workflow for glycopeptide spectral analysis provided direct annotation of ~5,000 chondroitin-modified glycopeptide spectra and enabled the identification of these novel CPGs. One earlier proteomic based strategy for studies of GAG-attachment sites in *C. elegans* resulted in the identification of 9 core proteins (CPG-1 to CPG-9) (3). In that study, enriched GAG-peptides were treated with sodium hydroxide, causing  $\beta$ -elimination of the sugar chains and resulting in reactive serine residues that were subsequently tagged with dithiothreitol. The tagged serine residues allowed site-specific characterization with MS/MS. However, as  $\beta$ -elimination releases both GAGs and mucin-type O-glycans this strategy only provides tentative GAG assignments and thus requires additional experiments to confirm the GAG-nature of the glycosylation. In contrast to the complete release of the glycan by  $\beta$ -elimination, enzymatic treatment with, e.g. chondroitinase ABC, as used here, reduces the length of the chondroitin chain and generates free disaccharides and a residual hexasaccharide structure still attached to the peptide (1). Higher energy collision dissociation fragmentation of the resulting glycopeptides generates abundant peptide fragmentation that gives the identity of the peptide sequences. The fragmentation also generates glycan-specific fragment ions, including diagnostic oxonium ions that give the identity of the glycan, as well as the isomeric glycan identity (GalNAc versus GlcNAc) (2, 26–28). Here, all 24 core proteins were confirmed to be of chondroitin-type based on identification of specific glycopeptide ions (i.e. peptide + xylose) and specific GalNAc oxonium ion patterns in the low  $m/z$  range. The abundance of each CPG was assessed using spectral counts and indicated that several of

the novel core proteins constitute less than 0.2% of the total CPGs. By excluding dominating CPGs from the sample preparation, e.g. by immune depletion with antibodies targeting abundant CPGs, it may be possible to identify additional low abundance CPGs. Nevertheless, the present protocol enabled the identification of several novel CPGs and a similar strategy may thus be successful in identifying proteoglycans also in other important model organisms, such as *Drosophila melanogaster* and *Danio rerio*.

We decided to designate the novel core proteins CPG-10 to -24. Regardless of whether our assigned CPG-name will serve as a synonym to already established names or as the future submitted name, this nomenclature reflects the characteristics of chondroitin carrying core proteins and provides a consistent framework for investigating CPGs in *C. elegans*. It has long been viewed that nematodes, such as *C. elegans*, only produce chondroitin without any sulfation. It was very recently shown, however, that chondroitin is indeed sulfated in *C. elegans* although to a small extent (13, 14). However, the positions of the sulfates on the polysaccharide chains as well as which core protein(s) that are modified are still unknown. As earlier studies in zebrafish and mammalian cells indicate that reduced HS sulfation result in increased CS sulfation, this study was conducted with a *C. elegans* mutant strain lacking two HS sulfotransferases (hst-6 and hst-2 double mutant) to increase our chances to detect potential sulfate modifications near the linkage region (13). We did not, however, in this study, detect any sulfate modifications on the residual hexasaccharide structure, although the method is apt to detect such modifications (1). This suggests that the sulfate groups are either localized toward the non-reducing end of the polysaccharide or that the CSPGs are below our present level of detection. Future studies with site-specific analysis of longer saccharides and with improved detection levels, may determine which CPGs are indeed modi-

## Chondroitin proteoglycans in *C. elegans*

fied with sulfates. Furthermore, one may speculate on whether the reduced HS sulfation of the mutant strain used in our experiments may provoke the synthesis of novel chondroitin proteoglycans. If so, some of the core proteins reported here may be the result of such an effect and may not be identified in wild-type strains. Although such an effect cannot be excluded, we believe this is unlikely as HS comprise only a small amount (<0.4%) of the total GAGs in *C. elegans* (13), indicating that any alteration in HS sulfation will probably have little or no effect on chondroitin core protein synthesis. Moreover, to our knowledge no alteration in chondroitin/CS core protein biosynthesis has been reported in relationship to decreased HS sulfation in any model system investigated so far.

CPG-1 and CPG-2 are two essential components of the inner eggshell layer in *C. elegans*. The proteoglycans contain chitin-binding domains and interact with chitin during mitosis, forming a rigid matrix that provides protection to the growing embryo (18). Vertebrates produce several CSPGs that are important structural components of cartilage and other connective tissues, but also contribute to the regulation of more specialized processes, such as neurogenesis, growth factor signaling, and angiogenesis (29, 30). We show here that the nematode proteoglycans contain several functional domains, including Kunitz domains and endostatin domains, associated with protease inhibition and axonal migration, respectively (22, 23). Taken together, this suggests that CPGs in *C. elegans* not only serve as structural components in extracellular matrices but are also involved in more specialized functions. These findings indicate that specialized proteoglycan-mediated functions evolved early in metazoan evolution.

With the introduction of site-specific glycoproteomic analysis for CSPGs in human tissue fluids, the number of human CSPGs is increasing and now comprises more than 50 core proteins (1, 2). Using our defined *C. elegans* sequence motif for chondroitin attachment ([ED] – [GA] – S – G) in a Swiss-Prot database search, we identified 19 additional proteins suggesting that the number of identified core proteins in *C. elegans* will probably increase further. Targeted glycoproteomics with the use of immunoprecipitation or selective enrichment of certain tissues may be used to provide experimental evidence for additional CPGs.

Bioinformatics analysis of the core protein domain organization revealed that all chondroitin attachment sites were located in low complexity or disordered domains. Similar analysis demonstrated the same characteristics for CS-attachment sites on human CSPG prohormones, suggesting that this may be a general feature of chondroitin and chondroitin sulfate glycosylation in metazoan organisms. This is of interest because other Golgi-synthesized glycans, such as the O-GalNAc type glycosylation, also often reside in disordered domains (31). Interestingly, site-specific GalNAc-type O-glycosylation in disordered domains has been shown to be an important co-regulator of proprotein and metalloproteases processing (32, 33). Although the function of GAG chains are typically associated with the selective interaction to various ligands, one may speculate that the GAG function also depend on the position of the polysaccharide on the polypeptide backbone. We recently suggested that proteolytic processing of perlecan, an extracellular matrix

proteoglycan, is influenced by a GAG site in a disordered domain, which is located in close proximity to a metalloproteinase cleavage position (26). It is thus possible that our observed chondroitin sites, also in disordered domains, influence the core protein processing in *C. elegans* in a similar manner. The functional aspects of such potential processing will be the objects of future studies.

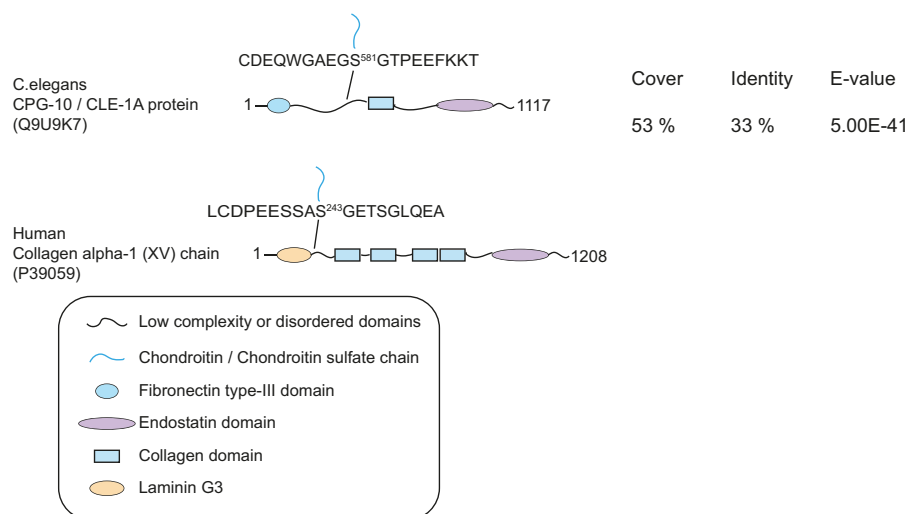
Our finding that the chondroitin attachment motif in *C. elegans* showed distinctive differences compared with the human motif suggests differences in xylosyltransferase specificities between the two species. Two vertebrate xylosyltransferases (I and II) have been identified, compared with only one in *C. elegans* (34). The chondroitin attachment site motif in *C. elegans* proteoglycans is more conserved in close proximity to the serine residue compared with the human CSPG motif. One may speculate that the less stringent motif in humans reflects the activity of the two different xylosyltransferases, each with slightly different substrate specificity. Most of the enzymes required for chondroitin/CS biosynthesis are relatively conserved between *C. elegans* and vertebrates (13). However, the majority of the identified core proteins in *C. elegans* do not show any resemblance to vertebrate proteins, apart from three of the core proteins (CPG-10/CLE-1A protein, CPG-16/FiBrillIN homolog, CPG-17/Papilin), which have sequence similarities to vertebrate counterparts (20–23). Interestingly, CPG-10/CLE-1A protein shows similarity to the human collagen  $\alpha$ -1 (XV) chain and we recently found that this protein is substituted with CS in human tissue fluids (1). The CPG-10/CLE-1A protein domain is to our knowledge the first example of an invertebrate chondroitin core protein that shows homology to a vertebrate counterpart. Comparison of the human and nematode core proteins reveal a high degree of similarity both regarding functional domains and their order of domain organization (Fig. 7). Taken together, these 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 polymerization, and in some cases also the core protein. However, because the majority of core proteins are not conserved between the species, our findings point to both converging and diverging selective forces during the proteoglycan evolution.

## Experimental procedures

### *C. elegans* maintenance and growth for CPG analysis

The strains OH4128 (*juIs76; evIs82b*), OH1421 (*hst-6(ok273)*), and OH1876 (*hst-2(ok595)*) as well as *Escherichia coli* strains OP50 and HB101 were obtained from the Caenorhabditis Genetics Center (CGC). The strain AHS50 (*evIs82b; hst-6(ok273)hst-2(ok595)*) (referred to as *hst-6 hst-2*) was generated from these by standard genetic methods and maintained on nematode growth medium agar at 20 °C on *E. coli* OP50 until the start of the experiment. To get larger amounts of material for CPG analysis, the animals were transferred to 10–20 rich nematode growth medium agar plates seeded with *E. coli* HB101. The worms were collected when almost all *E. coli* had been consumed and the material was washed in M9 buffer (22





**Figure 7. The CPG-10/CLE-1A protein (Q9U9K7) in *C. elegans* shows homology to the human protein collagen  $\alpha$ -1 (XV) chain (P39059).** The proteins display a high degree of sequence similarity regarding functional domains and their order of organization. However, the chondroitin/CS attachment site is different on the nematode compared with the human protein, as well as the amino acids surrounding the attachment site ( $-10$  to  $+10$ ).

mM  $\text{KH}_2\text{PO}_4$ , 44 mM  $\text{Na}_2\text{HPO}_4$ , 86 mM NaCl, pH 7.2). The worms were thereafter pelleted by centrifugation and then washed again in M9 buffer until no traces of bacteria were visible. The pellets were thereafter washed repeatedly with water, and then stored at  $-20^\circ\text{C}$ .

#### Enrichment of chondroitin glycopeptides

Chondroitin glycopeptides were purified from the worm extract using a combination of trypsin digestion and anion exchange chromatography, modified from a previously described protocol (1). In brief, worm pellets were dissolved in 1% CHAPS buffer, and boiled for 10 min at  $96^\circ\text{C}$ . The material was solubilized by consecutive passages through hypodermic needles with decreasing diameters (18 to 26 gauge). Samples were adjusted to 2 mM  $\text{MgCl}_2$  and incubated with 38  $\mu\text{l}$  of Benzonase (Novagen) at  $37^\circ\text{C}$  for 3 h. Benzonase was inactivated for 5 min at  $96^\circ\text{C}$  and the samples were then centrifuged for 10 min at  $13,000 \times g$ . One milligram of protein was reduced and alkylated in 1 ml of 50 mM  $\text{NH}_4\text{HCO}_3$ , and thereafter trypsinized overnight ( $37^\circ\text{C}$ ) with 20  $\mu\text{g}$  of trypsin (Promega). The digested samples were applied onto DEAE (GE Healthcare) columns (600  $\mu\text{l}$  in 10 ml of Poly-Prep<sup>®</sup> Chromatography columns (Bio-Rad) and incubated for 1 h at  $4^\circ\text{C}$ . The columns were washed with three different low-salt washing solutions at  $4^\circ\text{C}$  to remove loosely bound material: 15 min with 4 ml of 50 mM Tris-HCl, 100 mM NaCl, pH 8.0; 15 min with 4 ml of 50 mM NaAc, 100 mM NaCl, pH 4.0; 30 min with 100 mM NaCl. The GAG-peptides were then eluted stepwise with three buffers of increasing NaCl concentrations at  $4^\circ\text{C}$ : 1) 20 min with 6 ml of 400 mM NaCl; 2) 20 min with 6 ml of 800 mM NaCl; and 3) 25 min with 5 ml of 1500 mM NaCl. The three fractions collected were concentrated in a SpeedVac and desalted using PD10 columns (GE Healthcare). All fractions were lyophilized and the salt-free samples were then individually treated with 1 milliunits of chondroitinase ABC (C3667, Sigma) for 3 h at  $37^\circ\text{C}$ . Prior to MS analysis, the samples were desalted using a C18 spin column (8 mg resin) according to the manufacturer's protocol (Thermo Scientific, Inc., Waltham, MA).

#### LC-MS/MS analysis and spectral filtering

The samples were analyzed on a Q Exactive mass spectrometer coupled to an Easy-nLC 1000 system (Thermo Fisher Scientific, Inc., Waltham, MA), as previously described (1). Briefly, glycopeptides (10- $\mu\text{l}$  injection volume) were separated using an analytical column with Reprosil-Pur C18-AQ particles (Dr. Maisch GmbH, Ammerbuch, Germany). The following gradient was run at 150 nl/min; 7–37% B-solvent (acetonitrile in 0.2% formic acid) over 60 min, 37–80% B-solvent over 5 min, with a final hold at 80% B-solvent for 10 min. The A-solvent was 0.2% formic acid. Spectra were recorded in positive ion mode and MS scans were performed at 70,000 resolution with a mass range of  $m/z$  600–2000. The MS/MS analysis was performed in a data-dependent mode, with the top six most abundant charged precursor ions in each MS scan selected for fragmentation (MS2) by stepped higher energy collision dissociation with normalized collision energy values of 20, 25, and 30. The MS2 scans were performed at a resolution of 35,000 (at  $m/z$  200). A total of six MS2 data files were used for the bioinformatics analysis. The files were derived from the three different fractions (400 mM NaCl, 800 mM NaCl, and 1500 mM NaCl) of two independent sample preparations. The files were comprised of  $\sim 10,000$  MS/MS scans and the raw data files were filtered by *SweetNET* for the presence of the diagnostic oxonium ion at  $m/z$  362.10 ( $\Delta\text{HexAGalNAc}$ ). The GScore (Glycan score) and GGRatio (GlcNAc/GalNAc ratio) parameters were calculated based on the oxonium ion intensities of all selected MS/MS fragmentation scans, as described (2). The parameters enable identification of false-positive glycopeptide scans (through GScore) and to determine the isomeric identity of the HexNAc (GlcNAc/GalNAc) oxonium ions (through GGRatio).

#### Molecular networking

Molecular networks were generated using the workflow at the Global Natural Products Social Molecular Networking

## Chondroitin proteoglycans in *C. elegans*

Server (GNPS) found at <http://gnps.ucsd.edu/> (38).<sup>3</sup> The data were filtered by removing all MS2 peaks within  $\pm 17$  Da of the precursor *m/z*. MS2 spectra were window filtered by choosing only the top six peaks in the  $\pm 50$  Da window throughout the spectrum. The data were then clustered with MS-Cluster with a parent mass tolerance of 2.0 Da and a MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra. A network was then created where edges were filtered to have a cosine score above 0.7 and more than six matched peaks. Edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The networks were then iteratively propagated and annotated by *SweetNET* based on known mass shifts and Mascot searches, respectively.

### Mascot database search and SweetNET data analysis

Initial database searches were performed against *C. elegans* in the UniProtKB/Swiss-Prot database (3,871 sequences; 01/03/16) and NCBI (31, 266 sequences; 07/05/16) through Mascot Distiller (version 2.3.2.0, Matrix Science, London, U.K) using an in-house Mascot Server (version 2.3.02), as previously reported for CS glycopeptides (1). The Mascot database search results were used for the initial annotation of the spectral networks regarding peptide sequence and glycan composition. The *SweetNET* data validation module is a ranking system that outputs the "reliability score" for the suggested Mascot annotated and network propagated hits. The validation score (Vscore) is comprised of two components, *i.e.* specific oxonium ions (Supplemental Table 1) and glycopeptide fragment ions. This Vscore was calculated for each MS2 scan and for each network node as described previously (2).

### WebLogo, Pfam database, and phylogenetic analysis

Statistical analysis of aligned chondroitin amino acid attachment sites of *C. elegans* and CS attachment sites in humans were performed using Weblogo (35). The human CSPG core proteins were taken from data previously reported (1, 2) and are summarized in Table S3. For the Pfam database analysis, each core protein sequence was search against the Pfam library (Pfam 30.0; Jun 2016, 16306 families) ([www.pfam.xfam.org](http://www.pfam.xfam.org))<sup>3</sup> (36). The phylogenetic analysis was made based on the core proteins containing functional domains and was made using the web service Phylogeny.fr ([www.phylogeny.fr](http://www.phylogeny.fr))<sup>3</sup> (37).

**Author contributions**—F. N., T. D., L. K., and G. L. designed the study. T. D. and L. K. prepared the samples and F. N. performed the experiments. F. N., A. G. T., W. N., J. N., and G. L. analyzed the experiments. F. N. and G. L. wrote the paper. All authors reviewed the results and approved the final version of the manuscript.

**Acknowledgment**—The Proteomics Core Facility at the Sahlgrenska Academy, University of Gothenburg, is acknowledged for running all the MS analyses.

<sup>3</sup> Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

## References

1. Noborn, F., Gomez Toledo, A., Sihlbom, C., Lenggqvist, J., Fries, E., Kjellén, L., Nilsson, J., and Larson, G. (2015) Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. *Mol. Cell Proteomics* **14**, 41–49 [CrossRef Medline](#)
2. Nasir, W., Toledo, A. G., Noborn, F., Nilsson, J., Wang, M., Bandeira, N., and Larson, G. (2016) SweetNET: a bioinformatics workflow for glycopeptide MS/MS spectral analysis. *J. Proteome Res.* **15**, 2826–2840 [CrossRef Medline](#)
3. 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 [CrossRef Medline](#)
4. Lau, L. W., Cua, R., Keough, M. B., Haylock-Jacobs, S., and Yong, V. W. (2013) Pathophysiology of the brain extracellular matrix: a new target for remyelination. *Nat. Rev. Neurosci.* **14**, 722–729 [CrossRef Medline](#)
5. Coles, C. H., Shen, Y., Tenney, A. P., Siebold, C., Sutton, G. C., Lu, W., Gallagher, J. T., Jones, E. Y., Flanagan, J. G., and Aricescu, A. R. (2011) Proteoglycan-specific molecular switch for RPTPr clustering and neuronal extension. *Science* **332**, 484–488 [CrossRef Medline](#)
6. Siebert, J. R., and Osterhout, D. J. (2011) The inhibitory effects of chondroitin sulfate proteoglycans on oligodendrocytes. *J. Neurochem.* **119**, 176–188 [CrossRef Medline](#)
7. Kjellén, L., Pettersson, I., Lillhager, P., Steen, M. L., Pettersson, U., Lehtonen, P., Karlsson, T., Ruoslahti, E., and Hellman, L. (1989) Primary structure of a mouse mastocytoma proteoglycan core protein. *Biochem. J.* **263**, 105–113 [CrossRef Medline](#)
8. Bartolomucci, A., Possenti, R., Mahata, S. K., Fischer-Colbrie, R., Loh, Y. P., and Salton, S. R. (2011) The extended granin family: structure, function, and biomedical implications. *Endocr. Rev.* **32**, 755–797 [CrossRef Medline](#)
9. Mikami, T., and Kitagawa, H. (2013) Biosynthesis and function of chondroitin sulfate. *Biochim. Biophys. Acta* **1830**, 4719–4733 [CrossRef Medline](#)
10. Esko, J. D., and Zhang, L. (1996) Influence of core protein sequence on glycosaminoglycan assembly. *Curr. Opin. Struct. Biol.* **6**, 663–670 [CrossRef Medline](#)
11. Mizumoto, S., Ikegawa, S., and Sugahara, K. (2013) Human genetic disorders caused by mutations in genes encoding biosynthetic enzymes for sulfated glycosaminoglycans. *J. Biol. Chem.* **288**, 10953–10961 [CrossRef Medline](#)
12. Kreuger, J., and Kjellén, L. (2012) Heparan sulfate biosynthesis: regulation and variability. *J. Histochem. Cytochem.* **60**, 898–907 [CrossRef Medline](#)
13. Dierker, T., Shao, C., Haitina, T., Zaia, J., Hinas, A., and Kjellén, L. (2016) Nematodes join the family of chondroitin sulfate-synthesizing organisms: identification of an active chondroitin sulfotransferase in *Caenorhabditis elegans*. *Sci. Rep.* **6**, 34662 [CrossRef Medline](#)
14. Izumikawa, T., Dejima, K., Watamoto, Y., Nomura, K. H., Kanaki, N., Rikitake, M., Tou, M., Murata, D., Yanagita, E., Kano, A., Mitani, S., Nomura, K., and 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**, 23294–23304 [CrossRef Medline](#)
15. Bulik, D. A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W. R., Esko, J. D., Robbins, P. W., and Selleck, S. B. (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**, 10838–10843 [CrossRef Medline](#)
16. Hwang, H. Y., Olson, S. K., Esko, J. D., and Horvitz, H. R. (2003) *Caenorhabditis elegans* early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**, 439–443 [CrossRef Medline](#)
17. Pedersen, M. E., Snieckute, G., Kagias, K., Nehammer, C., Multhaupt, H. A., Couchman, J. R., and Pocock, R. (2013) An epidermal microRNA regulates neuronal migration through control of the cellular glycosylation state. *Science* **341**, 1404–1408 [CrossRef Medline](#)

18. Olson, S. K., Greenan, G., Desai, A., Müller-Reichert, T., and Oegema, K. (2012) Hierarchical assembly of the eggshell and permeability barrier in *C. elegans*. *J. Cell Biol.* **198**, 731–748 [CrossRef Medline](#)
19. Lawrence, R., Brown, J. R., Al-Mafraji, K., Lamanna, W. C., Beitel, J. R., Boons, G. J., Esko, J. D., and Crawford, B. E. (2012) Disease-specific non-reducing end carbohydrate biomarkers for mucopolysaccharidoses. *Nat. Chem. Biol.* **8**, 197–204 [CrossRef Medline](#)
20. Kramerova, I. A., Kawaguchi, N., Fessler, L. I., Nelson, R. E., Chen, Y., Kramerov, A. A., Kusche-Gullberg, M., Kramer, J. M., Ackley, B. D., Siron, A. L., Prockop, D. J., and Fessler, J. H. (2000) Papilin in development: a pericellular protein with a homology to the ADAMTS metalloproteinases. *Development* **127**, 5475–5485 [Medline](#)
21. Kelley, M., Yochem, J., Krieg, M., Calixto, A., Heiman, M. G., Kuzmanov, A., Meli, V., Chalfie, M., Goodman, M. B., Shaham, S., Frand, A., and Fay, D. S. (2015) FBN-1, a fibrillin-related protein, is required for resistance of the epidermis to mechanical deformation during *C. elegans* embryogenesis. *Elife* **4**, eLife3.06565
22. Fries, E., and Blom, A. M. (2000) Bikunin—not just a plasma proteinase inhibitor. *Int. J. Biochem. Cell Biol.* **32**, 125–137 [CrossRef Medline](#)
23. Ackley, B. D., Crew, J. R., Elamaa, H., Pihlajaniemi, T., Kuo, C. J., and Kramer, J. M. (2001) The NC1/endostatin domain of *Caenorhabditis elegans* type XVIII collagen affects cell migration and axon guidance. *J. Cell Biol.* **152**, 1219–1232 [CrossRef Medline](#)
24. *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 [CrossRef Medline](#)
25. Haltiwanger, R. S., and Lowe, J. B. (2004) Role of glycosylation in development. *Annu. Rev. Biochem.* **73**, 491–537 [CrossRef Medline](#)
26. Noborn, F., Gomez Toledo, A., Green, A., Nasir, W., Sihlbom, C., Nilsson, J., and Larson, G. (2016) Site-specific identification of heparan and chondroitin sulfate glycosaminoglycans in hybrid proteoglycans. *Sci. Rep.* **6**, 34537 [CrossRef Medline](#)
27. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* **13**, 6024–6032 [CrossRef Medline](#)
28. Yu, J., Schorlemer, M., Gomez Toledo, A., Pett, C., Sihlbom, C., Larson, G., Westerlind, U., and Nilsson, J. (2016) Distinctive MS/MS fragmentation pathways of glycopeptide-generated oxonium ions provide evidence of the glycan structure. *Chemistry* **22**, 1114–1124 [CrossRef Medline](#)
29. Dityatev, A., Schachner, M., and Sonderegger, P. (2010) The dual role of the extracellular matrix in synaptic plasticity and homeostasis. *Nat. Rev. Neurosci.* **11**, 735–746 [CrossRef Medline](#)
30. Le Jan, S., Hayashi, M., Kasza, Z., Eriksson, I., Bishop, J. R., Weibrecht, I., Heldin, J., Holmborn, K., Jakobsson, L., Söderberg, O., Spillmann, D., Esko, J. D., Claesson-Welsh, L., Kjellén, L., and Kreuger, J. (2012) Functional overlap between chondroitin and heparan sulfate proteoglycans during VEGF-induced sprouting angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* **32**, 1255–1263 [CrossRef Medline](#)
31. King, S. L., J. H., Schjoldager, K. T., Halim, A., Madsen, T. D., Dzielgiel, M. H., Woetmann, A., Vakhrushev, S. Y., and Wandall, H. (2017) Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Adv.* **1**, 429–442 [CrossRef](#)
32. Goth, C. K., Halim, A., Khetarpal, S. A., Rader, D. J., Clausen, H., and Schjoldager, K. T. (2015) A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14623–14628 [CrossRef Medline](#)
33. Schjoldager, K. T., Vester-Christensen, M. B., Goth, C. K., Petersen, T. N., Brunak, S., Bennett, E. P., Levery, S. B., and Clausen, H. (2011) A systematic study of site-specific GalNAc-type O-glycosylation modulating pro-protein convertase processing. *J. Biol. Chem.* **286**, 40122–40132 [CrossRef Medline](#)
34. Wilson, I. B. (2004) The never-ending story of peptide O-xylosyltransferase. *Cell Mol. Life Sci.* **61**, 794–809 [CrossRef Medline](#)
35. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 [CrossRef Medline](#)
36. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 [CrossRef Medline](#)
37. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 [CrossRef Medline](#)
38. Wang, M., et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 PMID: 27504778