Deep sequencing of complex proteoglycans: a novel strategy for high coverage and site-specific identification of glycosaminoglycan-linked peptides

Joshua A. Klein\textsuperscript{1,2}, Le Meng\textsuperscript{1}, Joseph Zaia\textsuperscript{1,2}

\textsuperscript{1}Dept. of Biochemistry, Center for Biomedical Mass Spectrometry

\textsuperscript{2}Bioinformatics Program

Boston University, Boston, MA

Running title: proteoglycan deep seq

Address for correspondence:

Joseph Zaia

Boston University Medical Campus

670 Albany St., Rm. 509

Boston, MA 02118

(v) 1-617-638-6762

(f) 1-617-639-6760

(e) jzaia@bu.edu
Abstract

Proteoglycans are distributed in all animal tissues and play critical, multifaceted, physiological roles. Expressed in a spatially- and temporally-regulated manner, these molecules regulate interactions among growth factors and cell surface receptors and play key roles in basement membranes and other extracellular matrices. Due to the high degree of glycosylation by glycosaminoglycan (GAG), N-glycan and mucin-type O-glycan classes, the peptide sequence coverage of complex proteoglycans is revealed poorly by standard mass spectrometry-based proteomics methods. As a result, there is little information concerning how proteoglycan site specific glycosylation changes during normal and pathological processes. Here, we developed a workflow to improve sequence coverage and identification of glycosylated peptides in proteoglycans. We applied this workflow to the small leucine-rich proteoglycan decorin and the hyalectan proteoglycans; neurocan, brevican, and aggrecan.

We characterized glycosylation of these proteoglycans using LC-MS methods easily implemented on instruments widely used in proteomics laboratories. For decorin, we assigned the linker-glycosite and three N-glycosylation sites. For neurocan and brevican, we identified densely glycosylated mucin-like regions in the extended domains. For aggrecan, we identified 50 linker-glycosites and mucin-type O-glycosites in the extended region and N-glycosites in the globular domains, many of which are novel and have not been observed previously. Most importantly, we demonstrate an LC-MS and bioinformatics approach that will enable routine analysis of proteoglycan glycosylation from biological samples to assess their role in pathophysiology.
Introduction

At present, 43 distinct genes have been reported for proteoglycans (1) and are found ubiquitously in intracellular vesicles, on cell surfaces, in pericellular spaces and in extracellular matrices. Proteoglycan backbone sequences contain modules, conserved among family members, that bind to extracellular matrix collagens, glycoproteins, and hyaluronan to form three dimensional networks, the structures and functions of which are tissue-, development-, and disease-specific. It must be emphasized that the functions of this relatively small number of gene products are elaborated by spatially and temporally regulated glycosylation (2). These molecules are of interest for their roles in the dysregulated cell signaling mechanisms associated with a wide variety of cancers (3-9) and for understanding of the roles extracellular matrix structural cues play in permitting cancer growth, angiogenesis and metastasis (10).

Extracellular matrix proteoglycans include the small leucine rich proteoglycans (SLRPs) and the hyalectans. Mature chondroitin sulfate (CS) chains on these proteoglycans consist of a linker tetrasaccharide, and a disaccharide repeat region. Following digestion with a chondroitin lyase enzyme, a hexasaccharide composed of X-Y-GlcAβ3Galβ3Galβ4Xylβ-O- is left attached to a serine residue (Figure S 3). CS proteoglycans contain sub-stoichiometric phosphorylation of Xyl due to the activities of xylosylphosphorylase and xylosylphosphatase enzymes (11-13). They can carry modifications to the linker tetrasaccharide including sulfation of one or both Gal residues (14, 15). Fucosylation of Xyl and sialylation of Gal adjacent to Xyl have been reported for bikunin (16).

The SLRPs, including 18 gene products and many splice variants, comprise the largest family of leucine rich repeat containing proteins. They contain homologous 24 amino acid leucine rich repeats and protein cores of 36-42 kDa (1). Two of the five SLRP classes are non-canonical and are not modified with GAG chains. SLRP biological functions arise in part through
multifaceted protein binding interactions. The crystal structure of decorin, the most studied family member, shows a curved solenoid structure that typifies the SLRPS (17). Protein binding occurs primarily through the amino acid side chains on the concave side of the solenoid. The SLRPs bind and modulate the structures of collagens. They also impact cellular signaling through their interactions with receptor tyrosine kinases and growth factors.

The hyalectans comprise aggrecan, versican, brevican and neurocan, each of which contains N-terminal globular G1 domains with homology to link proteins that bind the hyaluronan (HA) polysaccharide, anchoring the molecules to the cell surface. All contain a C-terminal G3 domain with EGF-like repeats, a complement regulatory module, and a C-type lectin module (1). Aggrecan contains a G2 domain, adjacent to G1, which has link homology. The hyalectans also contain \( N \)-glycosylation sequons in the globular domains that have been studied using mass spectrometry (18).

Much of the information regarding patterns of proteoglycan expression derives from western blots, immunohistochemistry, and other antibody-based studies (19, 20). While such antibodies indicate the presence of individual proteoglycan epitopes, the epitope structures can be difficult to determine and the level of antibody specificity is often unclear. Lectins are also used for staining tissue proteoglycans but leave the underlying proteoglycan structure undefined (21-23).

Using conventional proteomics workflows, protein levels are inferred from the abundances of a few representative unmodified peptides. Since small, homogeneous PTMs including phosphorylation, acetylation, and methylation can be predicted readily from genomics databases, they have been studied widely using proteomics (24, 25). Complex glycosylation alters the dissociation behavior of peptides, and is heterogeneous as a rule. These properties multiply the number of molecular forms (proteoforms) that must be considered in a proteomics
search space. As a result, specialized analytical and bioinformatics methods are necessary (26, 27).

Enzymatic removal of GAG chains leaves a linker tetrasaccharide plus one disaccharide attached to the protein or peptide. Larson et al. identified characteristic oxonium ions in tandem mass spectra of linker-peptides of CS and HS proteoglycans (16). They also observed a set of linker substituents, including sulfate, phosphate, fucose, and sialic acid (28-30). They used this approach to identify proteoglycan core proteins based on the linker-peptide sequences. In a study of CSPGs enriched from human urine or cerebrospinal fluid, these authors identified linker-glycopeptides from proteoglycans including decorin, brevican and neurocan (30).

Our goal is to characterize proteoglycan glycosylation as completely as possible. We focus in this work on one member of the SLRP family (decorin) and three hyalectans (neurocan, brevican and aggrecan). We demonstrate analytical and bioinformatics workflows that address proteoglycan complexity. The LC-MS measurements were conducted using widely available instrumentation. We used our publicly available, open source GlycReSoft software to calculate bioinformatics search space sets for proteoglycan samples and assign glycopeptides. We assigned the proteoglycan N-, mucin-type O-, and GAG-glycosylation by calculating theoretical glycopeptide lists against which higher energy dissociation LC-MS data were searched.

**Materials and methods:**

**Materials:** Bovine articular cartilage aggrecan, bovine articular cartilage decorin, chymotrypsin, endoproteinase Glu-C, dithiothreitol, iodoacetamide, sodium dodecyl sulfate, urea, MS-safe protease and phosphatase inhibitors, and chondroitinase ABC were purchased from Sigma-Aldrich (St. Louis, MO). Recombinant human neurocan and brevican were purchased from R&D systems. As described in the manufacturer data sheet (Figure S 1), the human neurocan was expressed with a C-terminal 6-His tag in Chinese hamster ovary cells with a predicted core
protein molecular weight of 141.4 kDa. The molecular weight observed using reducing SDS-PAGE was 140-290 kDa. Activity was determined by measuring enhancement of neuroglioma cell adhesion. As described in the manufacturer data sheet (Figure S 2), human brevican was expressed in a mouse myeloma cell line with a C-terminal 6-His tag with a predicted core protein molecular weight of 97.7 kDa. The molecular weight observed using reducing SDS-PAGE was 130-160 kDa. Activity was assessed using functional ELISA assay in which brevican bound biotinylated hyaluronan with an estimated $K_d < 3 \text{nM}$. Trypsin/Lys-C mix was purchased from Promega Biosciences (Madison, WI). The BCA protein quantification kit was purchased from Thermo-Fisher Scientific (Waltham, MA).

Proteoglycan preparation and digestion: Proteoglycans were dissolved in water and protein concentration was measured using a BCA kit. For decorin, brevican and neurocan, proteolytic digestion was carried out using 1:50 enzyme:protein ratio trypsin and Lys-C mixture (Promega Corporation, Madison, WI) and 100 µg proteoglycan in 100 µL 50 mM ammonium bicarbonate with 0.025% ProteaseMax surfactant (Promega) at 37°C overnight. To increase peptide coverage for aggrecan, Glu-C digestion was performed following trypsin/Lys-C digestion. All digestions were performed on a Thermo mixer with 30 sec shaking after 15 min. Enzymes were inactivated by heating at 98°C for 5 min.

Enrichment of GAG-linked peptides: The enrichment workflow is shown in Figure S 4. Proteolytic digests were loaded on a 0.5 mL 10 KDa molecular weight cutoff filter (Millipore Burlington, MA), and centrifuged at 14000Xg for 20 min. A 25 µL volume of 0.5 M NaCl was added to the retentate and the filter was centrifuged at 14000Xg for 10 min. A 50 µL volume of water was added to the retentate and the filter was centrifuged at 14000Xg for 10 min. GAG-linked peptides were retained by the 10 kDa molecular weight cutoff filter. Chondroitinase ABC (5 mU) was added to the retentate and incubated at 37°C overnight. The solutions were filtered through a 10 kDa molecular weight cutoff filter as above. Fractions were analyzed using
reversed phase LC-MS as described below. Plots of the extracted ion chromatograms for the ΔHexAHexNAc oxonium ion at m/z 362.1 showed that fraction retained by the 10 kDa contained the linker-peptides, (Figure S 5). As will be described in detail for each proteoglycan sample, CS-linker-peptides were enriched in the retained fraction.

Liquid chromatography mass spectrometry: Enriched digests were cleaned with Omix C-18 tips (Agilent Technologies, Santa Clara, CA) and 100 ng quantities were loaded on a Q-Exactive HF mass spectrometer (Thermo Scientific, San Jose, CA) equipped with an Advion NanoMate nanoESI source, coupled to a Waters nanoAcquity nanoLC system. A Water Xbridge reversed-phase column (150 μm × 100 mm) with 1.7 μm BEH C18 resin and a Waters trap column (180 μm × 20 mm) packed with 5 μm Symmetry™ C18 stationary phase were used for online desalting and separation of the proteomics samples as previously described (31). The mass spectrometer was programmed to acquire data-dependent tandem MS using the top 20 most abundant multiply charged precursor ions and three microscans per spectrum for both MS and tandem MS. Single LC-MS runs were acquired for each proteoglycan sample. All LC-MS data have been posted to the Pride Archive (accession number PXD008855).

Proteomics database search

LC-MS data files were searched against the Uniprot/Swissprot database using Peaks Studio ver. 8.5 (Bioinformatics Solutions, Inc., Waterloo, ON, Canada) with 15 ppm precursor and 0.02 Da product ion error tolerance, 1% false discovery rate and at least 2 unique peptides for protein identification. Carbamidomethyl cysteine was a fixed modification. A maximum of 3 variable modifications were allowed for hydroxyl proline: 15.99; HexNAc: 203.08-serine/threonine; HexHexNAc-serine/threonine: 365.13; ubiquitination-K: 114.04; hydroxyl lysine: 15.99; hydroxyl lysine-Hex: 178.05; hydroxyl lysine-Hex-Hex: 340.10; phosphorylation-serine/threonine/tyrosine: 79.97. For decorin, brevican and neurocan, trypsin was specified with
up to 3 missed cleavages and 1 non-specific cleavage. For aggrecan, trypsin plus endoproteinase Glu-C were specified with up to 3 missed cleavages and 1 non-specific cleavage. The Peaks software used a decoy fusion method to calculate false discovery rate (FDR) as published (32). The FDR was set in the Peaks software to 0.1%. The Peaks Studio 8.5 search exports have been provided as supplementary information.

**Mass Spectral Preprocessing using GlycReSoft (33, 34)**

The workflow for analysis of glycoproteomics samples using GlycReSoft is shown in Figure S 6. We de-isotoped and charge state deconvoluted MS1 and MS2 scans for each sample using an averagine model of C 10.93 H 15.75 N 1.6577 O 6.477 S 0.02 for MS1 and C 4.938 H 7.758 N 1.357 O 1.477 S 0.0417 respectively in GlycReSoft ver 0.3.1 (available at www.bumc.bu.edu/cbms). MS1 scans were averaged with the MS1 scan immediately preceding and proceeding it to boost signal and stabilize isotopic pattern consistency. We provide all raw mass spectral data for readers to verify our results using the GlycReSoft software. We also provide the GlycReSoft reports as supplemental files in interactive HTML format that show assigned and annotated tandem mass spectra.

**Glycopeptide Database Construction**

We constructed a glycomics search space that included $N$-, mucin type $O$-, and linker-saccharides. We used a combinatorial search space for $N$-glycans ($\text{Hex}_{0-3}$, $\text{HexNAC}_{2-8}$, $\text{Fuc}_{0-4}$, $\text{NeuAc}_{0-4}$; $\text{HexNAC} > \text{Fuc}$; $\text{HexNAC} > \text{NeuAc}$; sulfate = $\text{HexNAC} - 2$). We manually curated a list of 148 human mucin-type O-glycans with NeuAc and 427 bovine with NeuAc or NeuGc. We manually curated a list of 20 saccharides that corresponded to the linker tetrasaccharides plus $\Delta$HexA-HexNAC disaccharide plus 0-3 sulfate, 0-1 Fuc, 0-1 phosphate, 0-1 NeuAc. For human proteoglycans, the search space contained only NeuAc (a total of 3025 compositions). For
bovine proteoglycans, both NeuAc and NeuGc were allowed (a total of 8791 compositions). We constructed a glycopeptide search space for each sample using the mzIdentML file containing proteomics search results for that file, using all identified peptides as well as all theoretical peptides produced by an in-silico digest of those proteins with one glycosylation site. In order to build a search space for bi-glycosylated glycopeptides, we used results from the search for singly glycosylated peptides. All search space files have been posted to the Pride archive along with the raw MS data (Pride number will be given in the final accepted manuscript).

**Glycopeptide Database Search Identification**

We carried out glycopeptide identification by searching the preprocessed mass spectra against the associated databases using a 10 parts-per-million (ppm) error tolerance for the precursor mass and a 10 ppm error tolerance for product ions. We aggregated identified glycopeptides with a spectrum-level FDR of 5%. Our final result set contains only those identified in the search of the two glycosylation site database. The GlycReSoft reports have been provided as supplementary information.

The scoring function in GlycReSoft ranges between 0 and positive infinity, being based on $-10 \times \log_{10}(p \text{ value})$ from a binomial test of the matched peak intensity against all peaks + $-10 \times \log_{10}(p \text{ value})$ from a binomial test of the number of matched fragments given the number of theoretical fragments and mass accuracy constraints, weighted by the glycopeptide sequence coverage. This score is further augmented by a small bias towards higher mass accuracy following a Gaussian distribution, and bias towards glycan compositions which contain signature ions present in the matched scan. Larger scores correspond to higher confidence assignments, and a score below 20 tends to be unreliable. The glycopeptide false discovery rate is estimated using the target-decoy strategy (35), and for each glycopeptide the reported q-value (36) is the lowest FDR at which a glycopeptide can be accepted.
Glycopeptide identification comments

In many cases O-glycopeptides did not produce adequate glycosylated product ions to localize the glycosylation site. In other instances, O-glycopeptides could be solved by using a single large O-glycan or with two smaller O-glycans which sum to the same composition. When no discriminating ions were present, we reported all possible configurations supported by the data. In some scans, we observed additional HexNAc-bearing peptide backbone product ions.

Many of the linker glycopeptides of aggrecan contain short recurring peptide subsequences, making the glycosylation at each occurrence ambiguous. As these digest products would be indistinguishable under the experimental conditions, we assigned all evidence for each recurring glycopeptide to all valid sites. Different proteases would be necessary to separate the glycosylation patterns at each recurring subsequence using the flanking non-redundant sequences.

Sulfated N-glycopeptides were identified by precursor exact mass and the presence of sulfated oxonium ions. Since the intact masses of sulfate + 2 HexNAc (504.1261 Da) and 3 Hexose (504.16903 Da) are very close, it was possible to discriminate these ions below 4291.126 Da using a mass accuracy of 10 ppm.

Results

Glycoproteomics of bovine cartilage decorin, a small leucine rich proteoglycan

Decorin is a SLRP, present in all connective tissues, that participates in collagen fibrillogenesis, angiostasis, wound repair, and autophagy (37-39). Through binding growth factors and receptor signaling molecules (38), decorin has been implicated in diseases including atherosclerosis (40), fibrosis (41) and cancers (42-46). Decorin carries one CS chain near the mature N-
terminus and three \( N \)-glycosylation codons (see Figure S 7A, C). The CS chain composition varies according to tissue and disease state and includes dermatan sulfate-like repeats (47, 48).

We digested bovine articular cartilage decorin with trypsin and filtered the resulting peptides using a 10 kDa molecular weight cutoff membrane. The retained fraction was digested using chondroitinase ABC to remove CS chain repeats, leaving a disaccharide and linker tetrasaccharide attached to the core protein. A proteomics database search determined that the sample contained decorin and significant quantities of cartilage oligomeric matrix protein, the hyalectan aggrecan, and the SLRPs biglycan and fibromodulin (Figure S 7B). We used the proteomics search export to build a glycopeptide search space that included \( N \)-glycosylation, mucin-type \( O \)-glycosylation and linker saccharides.

Peptide \(^{30}\text{DEASGIGPEEHFPEVPEIEMGPVCPR}^{58}\) corresponded to the decorin N-terminal peptide, containing acidic residues followed by an SG sequence as a consensus site for GAG modification (49). Figure 1 shows an example linker-peptide tandem mass spectrum for this peptide with a trisulfated linker saccharide. While the peptide sequence is clear, the fragment ion masses do not definitively assign the glycosylation to the SG position. The assignment was therefore made from the consensus xylosylation site (50). Figure 2 shows the set of linker-peptide glycoforms detected at position 33 (51). The two most abundant of these carried three and two sulfate groups, respectively. The abundances for those linkers with no intensity bar fell below 1e7. All of these linker compositions were attached to peptide backbones assigned with high confidence. There is insufficient precursor mass error to allow the existing linker compositions to be confused with each other, so the ones without an EIC were likely caused by low abundance or incomplete deconvolution. Others have reported presence of fucose and sialic acid substitutions on the linker saccharide of bikunin (16, 52). We identified sialylated linker compositions with NeuAc and NeuGc, respectively. We detected bi- and mono-sulfated linker compositions but no fucosylated compositions for this decorin sample. Xylose
phosphorylation regulates proteoglycan biosynthesis (13) and occurs sub-stoichiometrically in mature proteoglycans (12, 53), possibly reflecting a balance of kinase and phosphorylase activities. For decorin, the xylose-phosphate containing linker saccharides were low in abundances.

Decorin N-glycosylation helps regulate core protein transit and secretion (54). Among the three bovine decorin N-glycosylation sequons, we detected an abundant set of glycoforms at positions 211 and 262 (Figure S 7E), and a single glycoform at position 303. Note that for these proteoglycan samples, we allowed up to two sites of glycosylation in the glycoproteomics search. For this reason, GlycReSoft was not able to distinguish between a single site of N-glycosylation and two sites of O-glycosylation for some glycopeptides and displayed both in the pileup diagram (Figure S 7D). High mannose and hybrid N-glycan compositions were abundant at site 211 but absent from site 262 (Figure S 7E), consistent with the latter being processed more efficiently to complex N-glycans.

Figure 3 shows the decorin unmodified and glycosylated peptides identified, totaling 80% coverage. This included mucin-type O-glycosylation of peptide 339-346 near the decorin C-terminus based on a tandem mass spectrum score of 46.054 (see supplemental files). As far as we know, this is the first report of comprehensive glycopeptide assignments for decorin and demonstrates an approach that will be useful comparative studies on site-specific glycosylation.

The proteomics database search identified aggrecan in the decorin sample with high coverage for the G1 and G2 globular domains and low coverage for the extended and G3 domains (Figure S 8A). An abundant set of N-glycopeptides was detected that corresponded to positions 332 in the G1 and 666 in the G2 domains (Figure S 8B, 3C). We identified no aggrecan linker-peptides in this sample. This is consistent with the conclusion that a population of truncated aggrecan molecules co-purified with decorin.
We also detected biglycan in the decorin sample. The biglycan sequence contains two SG attachment sites in the N-terminal region and was given an eponym to indicate two GAG chains (50). Biglycan tryptic peptide coverage from the proteomics database search of the decorin sample was 58%, indicating that this SLRP co-purified with decorin. We identified an abundant set of N-glycopeptides at the sequons at positions 270 and 311 (Figure S 9). These included sulfated compositions consistent with presence of sulfated N-acetyllactosamine units. No linker-peptides were detected for biglycan from this sample, possibly because the peptide, containing two acidic linker-saccharides, were not selected for data dependent acquisition.

Fibromodulin present in the decorin sample had a peptide coverage of 57% and N-glycopeptides were identified at positions 165 and 200, respectively (Figure S 10). We detected sulfated N-glycan compositions in low abundances, consistent with keratan sulfate units on fibromodulin (55).

**Glycoproteomics of brain hyalectan proteoglycans**

**Neurocan.** Consisting of an interacting network composed of hyaluronan, hyalectans and other glycoproteins, perineuronal nets are essential in control of plasticity in the adult brain (21). Interestingly, enzymatic digestion of CS chains, the bulk of which are bound to hyalectan core proteins, restores synaptic plasticity temporarily in adult brain (56) and memory in mice with tau pathology (57). In addition, perineuronal net structures become altered during aging, repair, learning, memory and in response to drugs of abuse (21). Much of the understanding of the roles of hyalectans in brain relies on immunohistochemical staining and/or genetic expression analyses. Therefore, in order to understand brain physiology and pathophysiology, it is necessary to integrate knowledge of extracellular matrix proteoglycan structure.

Neurocan is the predominant hyalectan in developing brain (58, 59) and is upregulated in response to brain injury (60). In adults, neurocan is associated with PNNs that surround axons
of some neurons (61). It contains N- and C-terminal globular domains homologous to other hyalectan family members and a 600 amino acid central domain rich in serine, threonine and proline and lacking in cysteine. This domain has been reported to carry about three CS chains located proximal to the globular domains. The neurocan central domain has been reported to carry approximately 40 mucin-type O-glycans (62). Proteome coverage reported in Uniprot KB is limited to the G1 and G3 globular domains with very sparse coverage of the extended domain (Figure S 11A).

We found recombinant human neurocan sample to contain significant levels of thrombospondin-1 and basement membrane-specific proteoglycan (PGBM, perlecan) (Figure S 11B). For neurocan, we identified 174 glycopeptides, corresponding to two N-glycosites and 48 mucin type O-glycosites (Figure S 11C, D). After inclusion of glycopeptides, the total coverage of neurocan was 59%. An example tandem mass spectrum of a glycopeptide modified with two mucin-type O-glycans is shown in Figure 4. The spectrum shows abundant peptide backbone fragmentation (blue and red ions) and peptide+saccharide ions (goldenrod-color ions). The peptide backbone product ions identify the glycosylation sites unambiguously but to not directly determine the glycan composition at each site. We report the most reasonable localization given the data alone. Hyalectan proteoglycans are extremely complex and it is appropriate for us to use the most reasonable localization from the data.

We detected linker peptides proximal to the G1 domain in the range of 360-415 (Figure S 11D). While the peptide sequence and glycan composition are confident, the glycosite was ambiguous. The most abundant neurocan linker saccharide composition contains xylose phosphate and no sulfation (Figure 5). The neurocan linker-glycopeptides are shown in Figure S 11E. The average neurocan linker composition is significantly less highly sulfated than that of decorin but distinct from that of aggrecan, as discussed below. In particular, linker with phosphorylated xylose was significantly more abundant in neurocan than aggrecan.
Figure 6 shows the overall coverage including the abundant mucin-type glycosylation in the extended domain. This is the first detailed map of glycosylation in the neurocan extended domain. Using this methodology, investigators will be able to determine changes to neurocan extended domain glycosylation associated with neuropsychiatric and other brain diseases.

**Brevican** is an important component of PNNs expressed in developmentally regulated manner (63). Secreted by neurons and glial cells and incorporated into the brain extracellular matrices, its expression is upregulated in response to injury (64) in brain tumor microenvironments, and has been associated with tumor aggressiveness (65, 66). Brevican contains the N-terminal G1 and C-terminal G3 globular domains characteristic of the hyalectans. The shortest hyalectan, the brevican central non-homologous domain of 290 amino acids is thought to contain 1-5 CS attachment sites and is considered to be a part-time proteoglycan with a portion of the molecular population lacking CS chains (67). The sequence contains two N-glycosylation sequons at positions 129, and 336.

The tryptic peptide proteome showed the recombinant human brevican sample was pure (Figure S 12A, C). We identified 260 total glycopeptides, including the two N-glycosylation sites and 49 O-glycosylation sites (Figure S 12D). An example N-glycopeptide tandem mass spectrum is shown in Figure 7 in which glycosylation is localized to the sequon by the presence of ions that include saccharide modifications. The extended region contains a mucin-like array of O-glycans, including two sites of hydroxylprolination (Figure S 12D). Several multiply glycosylated peptides were detected, an example of which is shown in Figure 8. The SL and TR glycosites were assigned based on the absence of saccharide modification for y1-y5, and the presence of saccharide modifications for y10-y24. Since the glycans undergo significant collisional dissociation, it was not possible to determine the glycan mass at each of the glycosites. The total brevican coverage after inclusion of glycopeptides was 79%, including 52
glycosites (Figure 9). To the best of our knowledge, this is the first report of a mucin-like domain that includes hydroxylprolination sites for brevican.

**Aggrecan** is a large and complex hyalectan, heavily modified by CS chains in the extended domains. The average aggrecan molecule has been estimated to carry 90 CS chains based on the total masses of core protein, CS, and the average mass of a CS chain (68). The bovine aggrecan sequence has 141 SG and 35 SA sub-sequences that may be considered candidate CS attachment sites (69). It also contains 9 *N*-glycosylation sequons and many potential mucin-type *O*-linked glycosylation sites. The protein coverage for aggrecan reported in UniprotKB is confined primarily to the globular domains (Figure S 13A). To the best of our knowledge, site-specific identification of peptides with CS chains in aggrecan has not been reported previously.

In initial experiments we found that use of trypsin as the only protease produced extended region glycopeptides that were too large for analysis using HCD LC-tandem MS. We therefore digested aggrecan with trypsin and Glu-C and then enriched linker-peptides (Figure S 4). Since individual CS chains have a molecular weight of over 10 kDa (70), we utilized a 10 kDa cut-off spin filter to enrich the linker-peptides. After enrichment, the linker-peptides were exhaustively digested with chondroitinase ABC to depolymerize the CS chains and leave linker-hexasaccharides attached to peptides. We analyzed the enriched fraction using LC-MS and performed a database search that showed aggrecan to be the most abundant protein (Figure S 13B). As expected, diagnostic oxonium ions from linkers were readily observed in the upper retained portion of the 10 kDa filter but were barely detected in the lower flow-through portion (Figure S 5D).

Linker glycosites were clustered in the extended region from 994-1453 (Figure S 13D). Several of the trypsin/Glu-C peptides from the extended domain had identical sequences and could not be distinguished. These included ISASGVEDL/ISR (1046-1057, 1317-1328, 1356-1367),
DISGLPSGGEVHLE (1011-1025, 1032-1046), DLSVLPSGEGHLE(1226-1239, 1246-1259, 1343-1356), L/IPSSEGPE(1057-1065, 1135-1143, 1173-1181, 1192-1200, 1269-1277, 1328-1336), and ISASGVE(1025-1032, 1046-1053, 1065-1072, 1086-1093, 1297-1394, 1317-1324, 1356-1363).

Bi-glycosylated linker-peptides were abundant in the trypsin/Glu-C digest dataset. Figure 10 shows a representative tandem mass spectrum of an aggrecan peptide modified with two linker saccharides. Although peptide backbone fragment ions were sufficiently abundant to assign both sites of xylosylation unambiguously, the glycan composition at each site could not be assigned. For many of the bi-glycosylated peptides, only one of the glycosites could be assigned confidently based on peptide backbone product ions. As a result, it is difficult to compare differences in linker-saccharide composition at different aggrecan glycosites. Looking at the average linker saccharide composition for aggrecan (Figure 11), the most abundant glycoforms have 0 sulfate groups and a very low degree of xylose phosphorylation. Neurocan linker peptides showed similar extent of sulfation but much higher xylose phosphorylation (Figure 5). By contrast, the most abundant decorin linker-peptides had 3 sulfate groups. In addition, decorin linker-saccharide showed sialic acid substitution (Figure 2).

Bovine aggrecan contains 144 SG and 35 SA sequences. We assigned a total of 50 linker glycosites, 41 to SG and 9 to SA. The glycosites corresponded to 9 different tri-amino acid sequences (Figure 12A). The most abundant tetra-amino acid glycosites were PSGE (12 instances), ASGV (8) and ISAS (7). Among XSG glycosites, P was the most common side chain in the first position (40%) followed by A (24%) and the fourth position was occupied most commonly by V (31%) and E (29%) (Figure 12B, C). We compared the expression of XSG sequence motifs in aggrecan among 9 mammalian species (Figure S 14A) and observed the order of abundance of PSG>ASG>LSG, VSG. The expression of SGX sequence motifs
showed an order of abundance of SGE>SGL>SGV (Figure S 14B). It therefore appears that the types of motifs occupied by linker mirror those present in the aggrecan protein sequences.

Among SA glycosites, 70% had I in the first position and all had S in the fourth position (Figure 12D, E). All SA glycosites occurred in SASG sequences. In many cases such sequences were modified with two linker-saccharides, making assignment of the SA glycosites unambiguous.

The two aggrecan \( N \)-glycosites occurred at NQT sequons in homologous peptide sequences in the G1 (position 332) and G2 (position 666) domains, respectively. The G1 site was occupied by a set of complex \( N \)-glycoforms ranging from bi-lateral-antennary, in agreement with the glycoforms distribution previously reported (18). The G2 \( N \)-glycosite was occupied by a set of bi- and tri-antennary complex \( N \)-glycans qualitatively similar in relative abundances to those at the G1 site (Figure 13). Both sites included sulfated glycoforms. Mucin-type \( O \)-glycosites occurred in the regions 987-90, 1490-92, 1780-86, 1871-75, and 1897-99. Overall, the density of mucin glycosites was much lower than observed in the extended domains of brevican and neurocan, respectively.

**Conclusions**

Until the present, much of the knowledge of tissue proteoglycan glycosylation has derived from classical biochemical and antibody-binding studies. To date, glycoproteomics methods have been used to characterize bikunin (16), CSPGs from urine and spinal fluid (28, 30), *C. elegans* (71), and the CS/HS PG perlecan (29). We report glycoproteomics methods to assess site-specific glycosylation of the SLRP decorin and the hyalectans aggrecan, brevican and neurocan. Our methods assign peptides glycosylated by linker saccharides, \( N \)-glycans or mucin-type \( O \)-glycans. We used LC-MS instruments and collisional dissociation methods that are widely available in proteomics laboratories. Our software used to interpret the data (GlycReSoft) is open source and publicly available.
The advantage to the use of mass spectrometry based omics methods is that the proteoglycan samples need not be pure; it is possible to define glycoproteomics search space information based on the measured proteome of the proteoglycan samples. The method assigns N-, mucin-type O- and Linker-glycopeptides from the same LC-MS data. Decorin is the archetypical SLRP and we demonstrate that nearly complete proteomics and glycoproteomics coverage can be obtained from a single data acquisition. We demonstrate that the extended domains of brevican and neurocan contain dense regions of mucin glycosylation. Linker glycosylation was not detected in brevican and was much lower in abundance than mucin glycosylation for neurocan. Since it is likely that glycosylation of the extended domains varies according to physiological context, it seems reasonable to conclude that brevican and neurocan mucin-regions impact the functions of these molecules. Aggrecan is vastly complex and we demonstrate the first assignments of site-specific linker-glycosylation in the extended domain.

The collisional dissociation methods employed here, while effective for singly glycosylated peptides, were able to distinguish sites in a subset of multiply glycosylated peptides. In order to resolve such glycosylation more effectively, electron activated dissociation (ExD) methods may be useful. The key point is to produce peptide backbone dissociation abundant enough for analysis of multiply glycosylated peptides; however, the linker-peptides are particularly acidic due to uronic acid, sialic acid, sulfate, and/or phosphate substitution. As a result, additional work is needed to define the effectiveness of positive mode ExD methods since charge states will be relatively low reducing the dissociation efficiency. It may be that negative ion ExD will prove more effective based on the negative charge of the glycopeptides, and this will require investigation.

**Acknowledgements**
This work was funded by NIH grant numbers P41GM104603, R21CA177479 and U01CA221234.

Data availability

The data discussed in the manuscript are available through the Pride Archive (Project PXD008855).
References

1. Iozzo, R. V., and Schaefer, L. (2015) Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biology* 42, 11-55.
2. Bishop, J. R., Schuksz, M., and Esko, J. D. (2007) Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* 446, 1030-1037.
3. Suhovskih, A. V., Kashuba, V. I., Klein, G., and Grigorieva, E. V. (2017) Prostate cancer cells specifically reorganize epithelial cell-fibroblast communication through proteoglycan and junction pathways. *Cell Adh Migr* 11, 39-53.
4. Schaefer, L., Tredup, C., Gubbiotti, M. A., and Iozzo, R. V. (2017) Proteoglycan neofunctions: regulation of inflammation and autophagy in cancer biology. *FEBS J* 284, 10-26.
5. Farnedi, A., Rossi, S., Bertani, N., Gulli, M., Silini, E. M., Mucignat, M. T., Poli, T., Sesenna, E., Lanfranco, D., Montebuognoli, L., Leonard, E., Marchetti, C., Cocchi, R., Ambrosini-Spaltro, A., Foschini, M. P., and Perris, R. (2015) Proteoglycan-based diversification of disease outcome in head and neck cancer patients identifies NG2/CSPG4 and syndecan-2 as unique relapse and overall survival predicting factors. *BMC Cancer* 15, 352.
6. Wade, A., Robinson, A. E., Engler, J. R., Petritsch, C., James, C. D., and Phillips, J. J. (2013) Proteoglycans and their roles in brain cancer. *FEBS J.* 280, 2399-2417.
7. Niedworok, C., Rock, K., Kretschmer, I., Freudenberger, T., Nagy, N., Sarvars, T., Vom Dorp, F., Reis, H., Rubben, H., and Fischer, J. W. (2013) Inhibitory role of the small leucine-rich proteoglycan biglycan in bladder cancer. *PLoS ONE* 8, e80084.
8. Yoneda, A., Lendorf, M. E., Couchman, J. R., and Multhaupt, H. A. (2012) Breast and ovarian cancers: a survey and possible roles for the cell surface heparan sulfate proteoglycans. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 60, 9-21.
9. Sanderson, R. D., Yang, Y., Kelly, T., MacLeod, V., Dai, Y., and Theus, A. (2005) Enzymatic remodeling of heparan sulfate proteoglycans within the tumor microenvironment: growth regulation and the prospect of new cancer therapies. *Journal of cellular biochemistry* 96, 897-905.
10. Theocharis, A. D., Gialeli, C., Bouris, P., Giannopoulou, E., Skandalis, S. S., Aletras, A. J., Iozzo, R. V., and Karamanos, N. K. (2014) Cell-matrix interactions: focus on proteoglycan-proteinase interplay and pharmacological targeting in cancer. *FEBS J* 281, 5023-5042.
11. Shibata, S., Midura, R. J., and Hascall, V. C. (1992) Structural characterization of proteoglycan biglycan in bladder cancer. *PLoS ONE* 8, e80084.
12. Ueno, M., Yamada, S., Zako, M., Bernfield, M., and Sugahara, K. (2001) Structural characterization of heparan sulfate and chondroitin sulfate of syndecan-1 purified from normal murine mammary gland epithelial cells. *J. Biol. Chem.* 276, 29134-29140.
13. Cheng, F., Heinegard, D., Fransson, L., Bayliss, M., Bielicki, J., Hopwood, J., and Yoshida, K. (1996) Variations in the chondroitin sulfate-protein linkage region of aggrecans from bovine nasal and human articular cartilages. *J Biol Chem* 271, 28572-28580.
16. Gomez Toledo, A., Nilsson, J., Noborn, F., Sihlbom, C., and Larson, G. (2015) Positive Mode LC-MS/MS Analysis of Chondroitin Sulfate Modified Glycopeptides Derived from Light and Heavy Chains of The Human Inter-alpha-Trypsin Inhibitor Complex. Mol Cell Proteomics 14, 3118-3131.

17. Scott, P. G., McEwan, P. A., Dodd, C. M., Bergmann, E. M., Bishop, P. N., and Bella, J. (2004) Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. P Natl Acad Sci USA 101, 15633-15638.

18. Zaia, J., Liu, B., Boynton, R., and Barry, F. (2000) Structural Analysis of Cartilage Proteoglycans and Glycoproteins Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. Anal Biochem 277, 94-103.

19. van Kuppevelt, T. H., Dennissen, M. A., van Venrooij, W. J., Hoet, R. M., and Veerkamp, J. H. (1998) Generation and application of type-specific anti-heparan sulfate antibodies using phage display technology. Further evidence for heparan sulfate heterogeneity in the kidney. J Biol Chem 273, 12960-12966.

20. Thompson, S. M., Fernig, D. G., Jesudason, E. C., Losty, P. D., van de Westerlo, E. M., van Kuppevelt, T. H., and Turnbull, J. E. (2009) Heparan sulfate phage display antibodies identify distinct epitopes with complex binding characteristics: insights into protein binding specificities. J Biol Chem 284, 35621-35631.

21. Sorg, B. A., Berretta, S., Blacktop, J. M., Fawcett, J. W., Kitagawa, H., Kwok, J. C., and Miquel, M. (2016) Casting a Wide Net: Role of Perineuronal Nets in Neural Plasticity. J Neurosci 36, 11459-11468.

22. Choudhary, C., and Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. Nat Rev Mol Cell Biol 11, 427-439.

23. Hu, H., Khatri, K., and Zaia, J. (2017) Algorithms and design strategies towards automated glycoproteomics analysis. Mass Spectrom Rev 36, 475-498.

24. Hu, H., Khatri, K., Klein, J., Leymarie, N., and Zaia, J. (2016) A review of methods for interpretation of glycopeptide tandem mass spectral data. Glycoconj J 33, 285-296.

25. Noborn, F., Gomez Toledo, A., Nasir, W., Sihlbom, C., and Larson, G. (2017) Characterization of Glycan Structures of Chondroitin Sulfate-Glycopeptides Facilitated by Sodium Ion-Pairing and Positive Mode LC-MS/MS. J Am Soc Mass Spectrom 28, 229-241.

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.

27. Khatri, K., Klein, J. A., White, M. R., Grant, O. C., Leymarie, N., Woods, R. J., Hartshorn, K. L., and Zaia, J. (2016) Integrated Omics and Computational Glycobiology Reveal Structural Basis for Influenza A Virus Glycan Microheterogeneity and Host Interactions. Mol Cell Proteomics 15, 1895-1912.

28. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 11, M111 010587.
33. Khatri, K., Klein, J. A., and Zaia, J. (2017) Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Anal Bioanal Chem* 409, 607-618.

34. Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., Conley, S., Staples, G. O., Slys, G. W., Smith, R. D., and Zaia, J. (2012) GlycReSoft: A Software Package for Automated Recognition of Glycans from LC/MS Data. *PLoS ONE* 7, e45474.

35. Elias, J. E., and Gygi, S. P. (2010) Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in molecular biology* 604, 55-71.

36. Kall, L., Storey, J. D., and Noble, W. S. (2008) Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics* 24, i42-48.

37. Jarvelainen, H., Sainio, A., and Wight, T. N. (2015) Pivotal role for decorin in angiogenesis. *Matrix Biol* 43, 15-26.

38. Gubbiotti, M. A., Vallet, S. D., Ricard-Blum, S., and Iozzo, R. V. (2016) Decorin interacting network: A comprehensive analysis of decorin-binding partners and their versatile functions. *Matrix Biol* 55, 7-21.

39. Buraschi, S., Neill, T., and Iozzo, R. V. (2017) Decorin is a devouring proteoglycan: Remodeling of intracellular catabolism via autophagy and mitophagy. *Matrix Biol*.

40. Singla, S., Hu, C., Mizeraicki, A., and Mehta, J. L. (2011) Decorin in atherosclerosis. *Ther Adv Cardiovasc Dis* 5, 305-314.

41. Baghy, K., Iozzo, R. V., and Kovalszky, I. (2012) Decorin-TGFbeta axis in hepatic fibrosis and cirrhosis. *J Histochem Cytochem* 60, 262-268.

42. Bi, X. L., and Yang, W. (2013) Biological functions of decorin in cancer. *Chin J Cancer* 32, 266-269.

43. Jarvinen, T. A., and Prince, S. (2015) Decorin: A Growth Factor Antagonist for Tumor Growth Inhibition. *BioMed research international* 2015, 654765.

44. Li, S. J., Chen, D. L., Zhang, W. B., Shen, C., and Che, G. W. (2015) Prognostic value of stromal decorin expression in patients with breast cancer: a meta-analysis. *J Thorac Dis* 7, 1939-1950.

45. Neill, T., Schaefer, L., and Iozzo, R. V. (2015) Oncosuppressive functions of decorin. *Mol Cell Oncol* 2, e975645.

46. Neill, T., Schaefer, L., and Iozzo, R. V. (2016) Decorin as a multivalent therapeutic agent against cancer. *Adv Drug Deliv Rev* 97, 174-185.

47. Thelin, M., Svensson, K. J., Shi, X., Bagher, M., Axelsson, J., Isinger-Ekstrand, A., van Kuppevelt, T. H., Johansson, J., Nilbert, M., Zaia, J., Belting, M., Maccarana, M., and Malmstrom, A. (2012) Dermatan sulfate is involved in the tumorigenic properties of Esophageus Squamous Cell Carcinoma. *Cancer Res.* 72, 1943-1952.

48. Miller, M. J. C., Costello, C. E., Malmström, A., and Zaia, J. (2006) A tandem mass spectrometric approach to determination of chondroitin/dermatan sulfate oligosaccharide glycoforms. *Glycobiology* 16, 502-513.

49. Mann, D. M., Yamaguchi, Y., Bourdon, M. A., and Ruoslahti, E. (1990) Analysis of glycosaminoglycan substitution in decorin by site-directed mutagenesis. *J Biol Chem* 265, 5317-5323.

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

51. Chopra, R. K., Pearson, C. H., Pringle, G. A., Fackre, D. S., and Scott, P. G. (1985) Dermatan sulphate is located on serine-4 of bovine skin proteodermatan sulphate. Demonstration that most molecules possess only one glycosaminoglycan chain and comparison of amino acid sequences around glycosylation sites in different proteoglycans. *Biochim J* 232, 277-279.
52. Lu, H., McDowell, L. M., Studelska, D. R., and Zhang, L. (2010) Glycosaminoglycans in Human and Bovine Serum: Detection of Twenty-Four Heparan Sulfate and Chondroitin Sulfate Motifs Including a Novel Sialic Acid-modified Chondroitin Sulfate Linkage Hexasaccharide. *Glycobiol Insights* 2010, 13-28.
53. Yamada, S., Okada, Y., Ueno, M., Iwata, S., Deepa, S. S., Nishimura, S., Fujita, M., Van Die, I., Hirabayashi, Y., and Sugahara, K. (2002) Determination of the glycosaminoglycan-protein linkage region oligosaccharide structures of proteoglycans from Drosophila melanogaster and Caenorhabditis elegans. *J Biol Chem* 277, 31877-31886.
54. Seo, N.-S., Hocking, A. M., Höök, M., and McQuillan, D. J. (2005) Decorin Core Protein Secretion Is Regulated by N-Linked Oligosaccharide and Glycosaminoglycan Additions. *J. Biol. Chem.* 280, 42774-42784.
55. Plaas, A. H., Neame, P. J., Nivens, C. M., and Reiss, L. (1990) Identification of the keratan sulfate attachment sites on bovine fibromodulin. *J Biol Chem* 265, 20634-20640.
56. Galtrey, C. M., Asher, R. A., Nothias, F., and Fawcett, J. W. (1997) Promoting plasticity in the spinal cord with chondroitinase improves functional recovery after peripheral nerve repair. *Brain* 130, 926-939.
57. Yang, S., Cacquevel, M., Saksida, L. M., Bussey, T. J., Schneider, B. L., Aebischer, P., Melani, R., Pizzorusso, T., Fawcett, J. W., and Spillantini, M. G. (2015) Perineuronal net digestion with chondroitinase restores memory in mice with tau pathology. *Exp Neurol* 265, 48-58.
58. Rauch, U., Feng, K., and Zhou, X. H. (2001) Neurocan: a brain chondroitin sulfate proteoglycan. *Cell Mol Life Sci* 58, 1842-1856.
59. Rauch, U., Gao, P., Janetzko, A., Flaccus, A., Hilgenberg, L., Tekotte, H., Margolis, R. K., and Margolis, R. U. (1991) Isolation and characterization of developmentally regulated chondroitin sulfate and chondroitin/keratan sulfate proteoglycans of brain identified with monoclonal antibodies. *J Biol Chem* 266, 14785-14801.
60. Asher, R. A., Morgenstern, D. A., Fidler, P. S., Adcock, K. H., Oohira, A., Braistead, J. E., Levine, J. M., Margolis, R. U., Rogers, J. H., and Fawcett, J. W. (2000) Neurocan is upregulated in injured brain and in cytokine-treated astrocytes. *J Neurosci* 20, 2427-2438.
61. Matsui, F., Nishizuka, M., Yasuda, Y., Aono, S., Watanabe, E., and Oohira, A. (1998) Occurrence of a N-terminal proteolytic fragment of neurocan, not a C-terminal half, in a perineuronal net in the adult rat cerebrum. *Brain research* 790, 45-51.
62. Retzler, C., Wiedemann, H., Kulbe, G., and Rauch, U. (1996) Structural and Electron Microscopic Analysis of Neurocan and Recombinant Neurocan Fragments. *J. Biol. Chem.* 271, 17107-17113.
63. Frischknecht, R., and Seidenbecher, C. I. (2012) Brevican: a key proteoglycan in the perisynaptic extracellular matrix of the brain. *Int J Biochem Cell Biol* 44, 1051-1054.
64. Hamel, M. G., Mayer, J., and Gottschall, P. E. (2005) Altered production and proteolytic processing of brevican by transforming growth factor beta in cultured astrocytes. *J Neurochem* 93, 1533-1541.
65. Varga, I., Hutoczki, G., Szemcsak, C. D., Zahuczky, G., Toth, J., Adamecz, Z., Kenyeres, A., Bognar, L., Hanzely, Z., and Klekner, A. (2012) Brevican, neurocan, tenascin-C and versican are mainly responsible for the invasiveness of low-grade astrocytoma. *Pathol Oncol Res* 18, 413-420.
66. Dwyer, C. A., Bi, W. L., Viapiano, M. S., and Matthews, R. T. (2014) Brevican knockdown reduces late-stage glioma tumor aggressiveness. *Journal of neuro-oncology* 120, 63-72.
67. Yamada, H., Watanabe, K., Shimonaka, M., and Yamaguchi, Y. (1994) Molecular cloning of brevican, a novel brain proteoglycan of the aggrecan/versican family. *J Biol Chem* 269, 10119-10126.
68. Hascall, V. C., and Sajdera, S. W. (1970) Physical properties and polydispersity of proteoglycan from bovine nasal cartilage. *J Biol Chem* 245, 4920-4930.
69. Esko, J. D., and Zhang, L. (1996) Influence of core protein sequence on glycosaminoglycan assembly. *Curr Opin Struct Biol* 6, 663-670.
70. Luscombe, M., and Phelps, C. F. (1967) The composition and physicochemical properties of bovine nasal-septa protein-polysaccharide complex. *Biochem J* 102, 110-119.
71. Noborn, F., Gomez Toledo, A., Nasir, W., Nilsson, J., Dierker, T., Kjellen, L., and Larson, G. (2018) Expanding the chondroitin glycoproteome of Caenorhabditis elegans. *J Biol Chem* 293, 379-389.
Figures

Figure 1. Tandem mass spectrum of a decorin linker glycopeptide 30-58 (4354.5765 Da). Color code: green, oxonium ions; red, peptide y-ions; blue, peptide b-ions; goldenrod, peptide+saccharide fragments; grey, unassigned ions.
Figure 2. Decorin disaccharide-linker-glycopeptide compositions. A total of 9 compositions were detected.
Figure 3. Diagram showing bovine decorin peptides and glycopeptides observed from the tryptic digest fraction retained by the molecular weight cutoff membrane. Each colored bar represents an assigned peptide or glycopeptide. Additional details on glycopeptide compositions are given in Figure S 7D and GlycReSoft HTML report in the Supplemental Information.
Figure 4. Tandem mass spectrum of neurocan peptide 559-581 (5069.1165 Da) with two mucin-type O-glycosites. Color code: green, oxonium ions; red, peptide y-ions; blue, peptide b-ions; goldenrod, peptide+saccharide fragments; grey, unassigned ions.
Figure 5. Neurocan average Linker saccharide composition
Figure 6. Diagram showing human recombinant neurocan peptides and glycopeptides identified. Each colored bar represents an assigned glycopeptide. Additional details are given in Figure S 11C and the GlycReSoft HTML report in the Supplemental Information.
Figure 7. Tandem mass spectrum of brevican N-glycopeptide 105-136 (4607.2538 Da). Color code: green, oxonium ions; red, peptide y-ions; blue, peptide b-ions; goldenrod, peptide+saccharide fragments; grey, unassigned ions.
Figure 8. Tandem mass spectrum of brevican O-glycopeptide 579-605 (4496.8923 Da). Color code: green, oxonium ions; red, peptide y-ions; blue, peptide b-ions; goldenrod, peptide+saccharide fragments; grey, unassigned ions. Glycosites are indicated in red.
Figure 9. Diagram showing human brevican peptides and glycopeptides identified from the tryptic digest fraction retained by the molecular weight cutoff membrane. Each colored bar represents an assigned glycopeptide. Additional details are given in Figure S 12C and the GlycReSoft HTML report in the Supplemental Information.
Figure 10. Tandem mass spectrum of aggrecan CS-linker-peptide 1032-1046 (3395.2511 Da). Color code: green, oxonium ions; red, peptide y-ions; blue, peptide b-ions; goldenrod, peptide+saccharide fragments; grey, unassigned ions. Glycosites are indicated in red.
Figure 11. Aggrecan average Linker saccharide composition
Figure 12. Comparison of linker glycosite sequence contexts in bovine aggrecan. XSGZ glycosites: A USXZ glycosites with total for each glycosite indicated with an asterisk *; B: (XSG) C (SGZ) D (XSA), E (SAZ), where X and Z are variable amino acid residues.
Figure 13. Comparison of glycosylation at two aggrecan $N$-glycosites from the trypsin plus Glu-C digest fraction retained by the molecular weight cutoff membrane.
Figure 14. Diagram showing all peptides and glycopeptides detected for bovine aggrecan from the trypsin plus Glu-C digest fraction regained by the molecular weight cutoff membrane. Each colored bar represents an assigned glycopeptide. Additional details are given in Figure S 13C and the GlycReSoft HTML report in the Supplemental Information.