

# Chapter 4

## Characterization of Proteoglycanomes by Mass Spectrometry



Christopher D. Koch and Suneel S. Apte

**Abstract** As composites of a core protein and several chemically distinct types of glycosaminoglycan (GAG) chains, proteoglycans are diverse molecules that occupy a unique niche in biology. They have varied and essential roles as structural and regulatory molecules in numerous physiological processes and disease pathology. In regard to cellular context, some link the interior of the cell to the extracellular matrix (ECM) as transmembrane or membrane-anchored molecules with a major role in cell adhesion and signal transduction. Others reside in pericellular matrix, where they influence crucial aspects of cell behavior, and several reside in interstitial ECM as components of structural macromolecular networks. Because of their unique composition, they can be challenging to identify and characterize using conventional biochemical or antibody-based methods. In contrast, the GAG component, despite its immense chemical diversity, typically carries a strong net negative charge which can be exploited to advantage for affinity-isolation and enrichment of proteoglycans from any biological system in a core protein-, GAG-, tissue-, and species-agnostic manner by anion exchange chromatography. This method, when coupled with high resolution liquid-chromatography tandem mass spectrometry (LC-MS/MS) can be used to define the proteoglycanome of any cell type, tissue or organism. A proteoglycanomics strategy can be further refined by inclusion of additional orthogonal affinity steps or fractionation for greater specificity and to deliver proteoglycans with distinct specified characteristics. Moreover, elimination of the GAG chain chemically and/or obliteration of the core protein enables glycomics characterization of GAG structure. Enzymatic digestion of GAGs on tryptic peptides allows mapping of glycopeptides, which has been used for identification of novel proteoglycans and

---

C. D. Koch

Department of Laboratory Medicine, Yale-New Haven Health, New Haven, CT, USA

Department of Biomedical Engineering, Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA

e-mail: [c.koch@yale.edu](mailto:c.koch@yale.edu)

S. S. Apte (✉)

Department of Biomedical Engineering, Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA

e-mail: [aptes@ccf.org](mailto:aptes@ccf.org)

© Springer Nature Switzerland AG 2020

S. Ricard-Blum (ed.), *Extracellular Matrix Omics*, Biology of Extracellular Matrix 7,  
[https://doi.org/10.1007/978-3-030-58330-9\\_4](https://doi.org/10.1007/978-3-030-58330-9_4)

to precisely define sites of GAG attachment. Recent application of proteoglycanomics to human aorta and human aortic aneurysms demonstrated its potential to identify tissue and disease proteoglycanomes and the detailed method that was used is provided here for application to other tissues or biological systems.

## 4.1 Introduction

Proteoglycans (PGs) are composite molecules in whom glycosaminoglycan (GAG) chains are covalently attached to a polypeptide backbone referred to as a proteoglycan core protein (Iozzo and Schaefer 2015). The attachment typically occurs to Ser residues adjacent to a Gly residue and usually within an acidic sequence context, although keratan sulfate attachment can occur not only to Ser, but also to Thr and Asn through distinct linkages (Brinkmann et al. 1997; Funderburgh 2002). Proteoglycans are integral components of the extracellular matrix (ECM) of most tissues, and some, such as syndecans and glypicans, are specialized and important transmembrane and membrane-anchored cell-surface components, respectively (Filmus and Capurro 2014; Mitsou et al. 2017; Gondelaud and Ricard-Blum 2019). Hyaluronan-binding aggregating chondroitin sulfate proteoglycans (CSPGs) such as aggrecan are quantitatively abundant in structural tissues such as cartilage and intervertebral disc, where they provide unique biophysical properties (Hascall and Heinegard 1974; Heinegard and Saxne 2011). Specifically, the ability of the aggregates to hold large amounts of water and thus exert an internal tissue swelling pressure, as well as electrostatic charge repulsion between the aggregates makes them indispensable for compression load-bearing in these tissues (Buschmann and Grodzinsky 1995). Aggregating proteoglycans in pericellular ECM of mesenchymal cells have the potential to control focal adhesion formation, cell shape and genetic programs (Mead et al. 2018). In the brain, a diverse group of aggregating PGs (aggrecan, versican, brevican, neurocan) (Zimmermann and Dours-Zimmermann 2008) are prominent components of the limited amount of ECM that is present, and their swelling pressure may ensure mechanical buffering within the cranium; furthermore, perineuronal nets of a subset of neurons have PGs as a major component, where they are thought to insulate the soma (cell body) of the neuron against rewiring of established neuronal circuits (Fawcett et al. 2019). Aggrecan, for example, is critical in this regard (Rowlands et al. 2018). In the view of the Nobel laureate Roger Tsien, holes in perineuronal nets, which are formed on conclusion of the juvenile critical period and mark the closure of neuronal plasticity, could be the seat of long-term memory (Tsien 2013). Major roles in developmental and cancer cell signaling by heparan sulfate proteoglycans (HSPGs) are attributed to the binding of growth factors and morphogens to the highly sulfated GAG chains (Bandari et al. 2015; Ortmann et al. 2015; Sarrazin et al. 2011; van Wijk and van Kuppevelt 2014; Yu and Woessner Jr. 2000). Similar roles are reprised during inflammation with respect to cytokines (Gondelaud and Ricard-Blum 2019; Bartlett et al. 2007). In addition, some PGs or their fragments can act as damage-associated molecular

patterns to stoke inflammation (Nastase et al. 2018). In many tissues such as tendons, skeletal muscle, lungs, cornea and blood vessels, proteoglycans are quantitatively minor components but serve crucial roles such as compression load-bearing, regulation of collagen fibril assembly and sequestration of growth factors (Ezura et al. 2000; Robinson et al. 2017; Zhang et al. 2006). One reason why proteoglycan steady state levels may be low in many cells and matrices is that they are among the most dynamic components of ECM and thus turned over quite rapidly, particularly in cell-proximate ECM.

Proteoglycan core proteins can be very large, e.g., perlecan, aggrecan and versican, or quite small, e.g., the leucine-rich proteoglycans. PG sub-groups are usually defined by a combination of size and other unique properties such as the chemical nature of their GAG chains, i.e., whether they are chemically defined chondroitin sulfate (CS), heparan sulfate (HS) or keratan sulfate (KS). Dermatan sulfate (DS) sometimes termed CS-B, is chemically similar to CS but contains iduronate (Thelin et al. 2013). This classification of PGs is imperfect, since different types of GAG can be present on the same core protein, e.g., aggrecan typically contains KS chains, yet the CS chains dominate and it is usually considered a CSPG. Moreover, some proteoglycan core proteins are not constitutively modified, and these could be regarded as “part-time” proteoglycans. However, for the purpose of this chapter, any molecule that carries a covalently attached GAG chain, if only in a small proportion of the core protein, is operationally defined as a proteoglycan.

Most GAGs are sulfated and thus share the useful property of having a net negative charge, enabling their isolation using anion exchange chromatography (AEC). This characteristic of PGs is also the basis of their detection in tissue sections using basic dyes such as alcian blue and safranin O or toluidine blue metachromasia of highly negatively charged GAGs such as heparin (Scott 1985). Their staining intensity is also a useful guide, albeit neither absolute nor specific, to GAG abundance. Tissue identification of individual GAGs and PG gene products is typically done using specific antibodies to the GAG chain or core protein. The former has the advantage over anti-core protein antibodies of being species-agnostic. Immunostaining and western blot can be used to detect the core proteins but relies on antibodies with high specificity. Additionally, the GAG-dense environment in which a core protein epitope may reside may mask antibody reactivity in tissue sections, requiring an intimate understanding of the molecules in order to properly design epitope retrieval strategies. However, the use of core protein and GAG antibodies as a targeted approach to examine the entire proteoglycan landscape of a tissue is neither practical nor efficient.

An untargeted approach, such as shotgun mass spectrometry would allow simultaneous detection of many proteoglycans within a complex cell or tissue extract while circumventing the need for quality antibodies and epitope retrieval protocols. Analysis of proteins by mass spectrometry is, however, limited by the caveat that high abundance proteins are preferentially detected during data-dependent analysis of mass spectra, and low abundance or highly modified components like proteoglycans may go undetected. A solution is to divide the sample into multiple fractions, using one or preferably, two orthogonal fractionation methods (Ly and Wasinger

2011). However, analysis of multiple fractions increases the instrument run time and requires analysis of data from multiple MS runs. An alternative to routine fractionation for reduction of sample complexity is enriching for the desired components, or excluding undesired ones (Ly and Wasinger 2011). We capitalized on the net negative charge of the GAGs to isolate and identify PG core proteins by mass spectrometry from the aorta (Cikach et al. 2018), the largest artery in the body, which contains a large repertoire of ECM molecules sandwiched in the space between elastic lamellae and arrays of smooth muscle cells.

Our approach utilized a well-characterized technique for proteoglycan enrichment from aortic tissue, i.e., isolation by AEC prior to analysis by LC-MS/MS (Cikach et al. 2018). AEC elution conditions can be adjusted to maximize proteoglycan yields and successful elution of proteoglycans can be evaluated using a variety of biochemical techniques such as fluorophore-assisted carbohydrate electrophoresis (providing precise delineation of the GAG type), safranin O staining (non-specific, but provides quantifiable staining intensity and indicates abundance of the GAGs in the eluted fractions) and western blot with a specific core protein or GAG antibody (Cikach et al. 2018). Combinations of these orthogonal methods can be used to determine exactly which PGs and GAGs were enriched, and LC-MS/MS can be subsequently used for unbiased identification of PGs extracted from essentially any tissue. Here, we describe the approach that was used to determine the proteoglycanome of the human aorta (Cikach et al. 2018). The proteoglycan isolation and quantitation methods are similar to those previously described by Carrino and colleagues (Carrino et al. 1991, 1994).

## 4.2 Extraction of Proteoglycans from Tissue

1. Snap freeze tissue in liquid nitrogen immediately after collection and store at  $-80^{\circ}\text{C}$  until use.
2. Finely dice tissue with a scalpel or surgical scissors in a clean petridish on ice. Weigh diced tissue and add 1 mL ice-cold proteoglycan extraction buffer per 100 mg tissue. The extraction buffer is: 4 M guanidine hydrochloride, 2% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate [CHAPS], 50 mM sodium acetate, adjusted to pH 6.0 with HCl or NaOH. 1 tablet of complete Mini EDTA-free Protease Inhibitor (Roche) is added per 6–10 mL.
3. Homogenize tissue in the extraction buffer with a mechanical homogenizer such as an Ultra-Turrax T2 or T10 homogenizer (IKA Works Inc.), taking care to keep the sample on ice as homogenization will quickly warm the sample.
4. Rotate homogenized tissue end-over-end at  $4^{\circ}\text{C}$  for a minimum of 16 h.
5. Clarify the homogenate by centrifugation for 15 min at  $20,000\times g$ .
6. Retain the supernatant as it contains both cellular and many extracellular matrix proteins including proteoglycans. Highly crosslinked proteins, including many collagens and elastic fibers, will remain in the sample pellet unless extracted by additional steps.

The extraction process would be similar for a cell culture monolayer, with the difference that after the medium is aspirated, the cells should be washed several times with serum-free medium prior to addition of the extraction buffer. This minimizes ion suppression from abundant serum proteins such as albumin during LC-MS/MS.

### 4.3 Proteoglycan Isolation by Anion Exchange Chromatography

1. **Buffer exchange:** While guanidine hydrochloride allows for efficient extraction of proteoglycans and proteins from tissue, it is not compatible with AEC due to the presence of chloride as a counterion. Therefore, buffer exchange to another chaotropic agent such as urea must precede the AEC step. This can be accomplished by a number of methods including dialysis and centrifugal filtration; we use gel filtration, which is an efficient, rapid and economical method using columns that are inexpensive and easy to prepare.
  - a. Prepare columns by cutting the tops off 5 mL, 10 mL or 25 mL plastic serologic pipettes and lightly pack the tips with glass wool. Connect a stopcock to the bottom of the pipette for volume control, such as a length of surgical tubing with removable clamp. Swell Sephadex G-50 fine resin in G50 buffer at a ratio of 20–25 mL/g of resin for at least 24 h. G50 buffer is 8 M urea, 0.5% CHAPS, 50 mM sodium acetate, 150 mM NaCl adjusted to pH 7.0 with HCl or NaOH. Heating on a hot plate-stirrer aids urea dissolution.
  - b. Load swollen Sephadex G-50 fine resin into the column (Table 4.1) and pack by gravity flow, adding G-50 buffer as needed. Once the column is packed, allow the G50 buffer to almost completely enter the column, i.e., leaving a meniscus to prevent the resin from drying.
  - c. Taking care not to disturb the resin bed, add the appropriate sample volume (see Table 4.1) to the column and allow it to enter the resin by gravity. Close the stopcock when the last of the sample has just entered the resin.

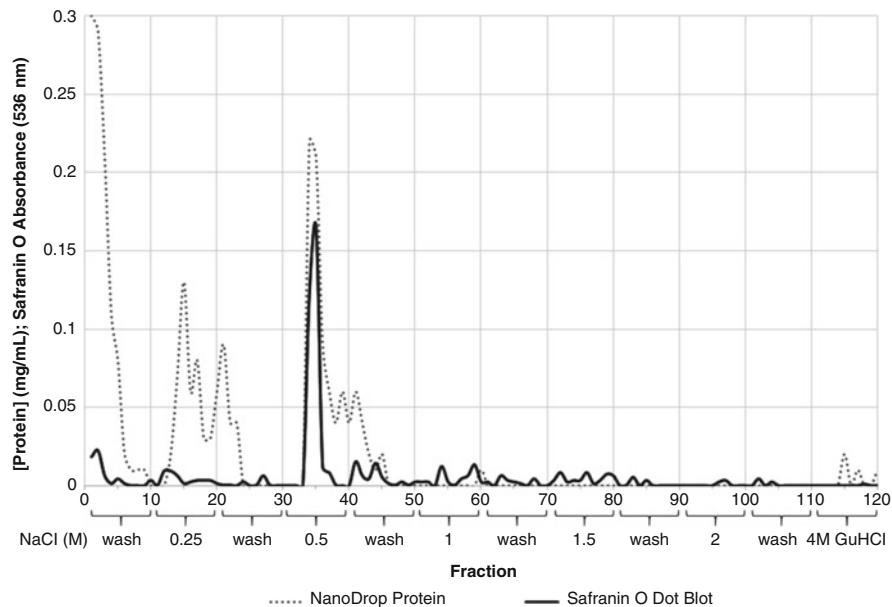
**Table 4.1** Volumes for buffer exchange using Sephadex G50 fine. The sample volume dictates the size of column(s) required. Buffer exchange can be accomplished with larger sample volumes through the use of sample aliquots and multiple columns. The pre- $V_0$  volume will not contain any proteins and is obtained immediately upon the sample fully entering the resin. Proteins will elute immediately following the pre- $V_0$  volume and are expected to be fully contained in the corresponding  $V_0$  volume, although volumes for specific applications may vary and should be determined experimentally

Pipette/column size (mL)	Sephadex G50 fine (mL)	Sample (mL)	Pre- $V_0$ (mL)	$V_0$ (mL)
5	4	1	0.25	1.5
10	8	2	0.5	3
25	24	6	1.5	9

- d. Carefully add G50 buffer to the column and resume gravity flow. Immediately start collecting the eluate as pre- $V_0$ . This eluate will be G50 buffer without proteins and can be discarded. The volume is specifically adjusted for the column sizes and sample volumes referenced in Table 4.1.
- e. Starting immediately after the pre- $V_0$  volume, collect the appropriate  $V_0$  volume. This fraction will deliver the proteoglycans and proteins in the AEC-compatible G50 buffer.
- f. Columns can be reused for subsequent samples after washing with 2–4 column volumes of G50 buffer; however single use may be more economical given the volume of G50 buffer that may be required for this.

## 2. Anion Exchange Chromatography

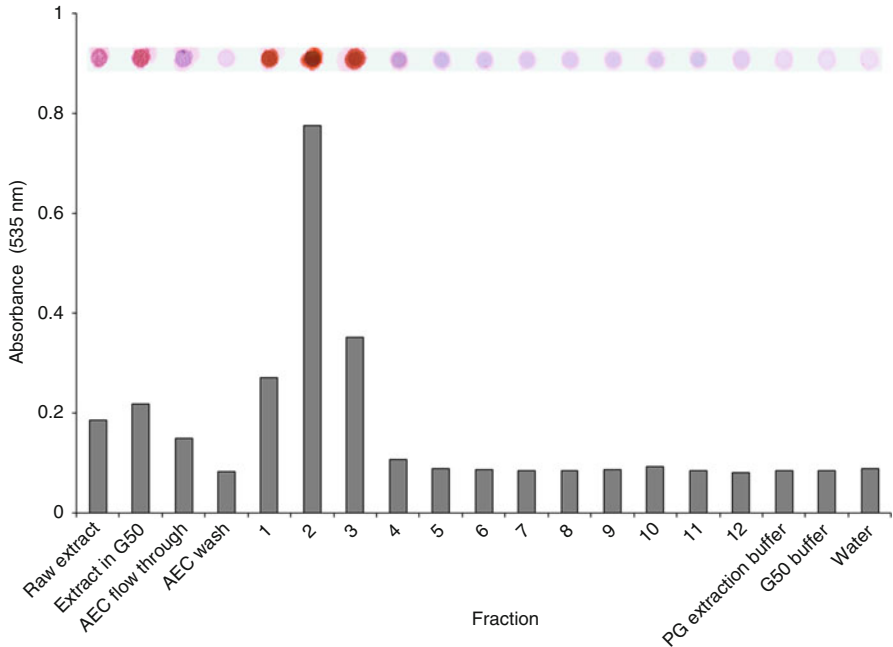
- a. Swell diethylaminoethyl (DEAE)-Sephacel with G50 buffer. Pack 4 mL swollen DEAE resin by gravity flow of G50 buffer into a glass column fitted with a stopcock as described above. Once the column is packed, allow the G50 buffer to almost completely enter the column taking care not to let the resin dry. Add the sample to the column without disturbing the resin bed and allow it to enter the resin by gravity flow. Collect the flow-through and retain for future analysis or discard. Wash the column with five column volumes (20 mL) of wash buffer (8 M urea, 0.5% CHAPS, 50 mM sodium acetate, 250 mM NaCl, pH 7.0). This fraction will contain weakly anionic proteins. It can be collected and retained for future analysis but is usually discarded.
  - b. Add elution buffer (same as wash buffer above, but with 0.5–1 M NaCl) to the column and collect 2 mL fractions. Most proteoglycans will elute with 0.5 M NaCl (Fig. 4.1), however proteoglycans containing high anion charge densities such as versican and aggrecan may elute best at higher concentrations (up to 1 M) of NaCl. In our experience, 1 M NaCl is sufficient to elute all proteoglycans with the highest concentrations eluting in fractions 1–3 (6 mL total eluate volume).
- ## 3. Quantitation of isolated proteoglycans.
- The isolated fractions first undergo buffer exchange by dialysis to 20 mM HEPES, 150 mM NaCl, pH 7.2, following which a safranin O staining and spectrophotometry method (Carrino et al. 1991) is applied to quantify proteoglycan content of the fractions. For total protein, sample absorbance at 280 nM is used. A representative safranin O-based quantitation profile of fractions is shown in Fig. 4.2.
- a. Prepare 0.45  $\mu\text{m}$  nitrocellulose by soaking in water for at least 1 min and mount it into a dot blot apparatus.
  - b. Pipette 25  $\mu\text{L}$  (1 part) of each DEAE eluate fraction (isolated proteoglycans) into 250  $\mu\text{L}$  (10 parts) 0.02% safranin O into each well. Mix briefly by pipetting and allow the mixture to stand for 1 min. Apply a vacuum to the dot blot apparatus to collect the precipitate onto the nitrocellulose.
  - c. Wash each well three times with 200  $\mu\text{L}$  water, drawing water through the nitrocellulose under vacuum. Wells with significant precipitation (high proteoglycan/glycosaminoglycan content) may require higher vacuum pressure



**Fig. 4.1** AEC elution profile of aortic GAGs. Aortic proteoglycans were eluted from a 4 mL DEAE-Sephacel column with G50 buffer containing increasing concentrations of NaCl. Ten 1 mL fractions were collected for each elution condition. Each elution was preceded by a 10 mL wash with G50 buffer containing 150 mM NaCl. A final elution was performed with 4 M GuHCl to ensure no proteoglycans remained on the column after the 2 M NaCl elution. Total protein and relative GAG concentration were determined for each fraction using a NanoDrop (absorbance at 280 nm) and safranin O dot blot, respectively. Most proteoglycans eluted with 500 mM NaCl, however the large proteoglycans versican and aggrecan may require 1 M NaCl for complete elution

and a longer time to empty fully. To prevent bleeding of precipitate into adjacent wells, dry the nitrocellulose in the dot blot apparatus for a minimum of 1 h.

- d. Remove the nitrocellulose from the dot blot apparatus and allow it to dry completely at room temperature. Cut or punch out the precipitate spots from the membrane and place each in a 1.5 mL microcentrifuge tube. Add 1 mL 10% cetylpyridinium chloride (CPC) to each tube and vortex vigorously. Incubate the tubes at 37 °C for 10 min, vortex vigorously, and incubate at 37 °C for an additional 10 min. Vortex vigorously after incubation and transfer a portion of the CPC solution from each tube into a 96 well plate or cuvette for absorbance measurement at 536 nm.
  - i. Fluorophore-assisted carbohydrate analysis, if available, is a specialized technique that can be used to identify and quantify hyaluronan and GAGs (Calabro et al. 2001; Midura et al. 2018).
  - ii. Western blot/ELISA using antibodies against specific GAGs and core proteins, or the GAG stubs left behind after enzymatic release can be



**Fig. 4.2** Quantitation of aortic glycosaminoglycans isolated using AEC. GAG concentration, as a surrogate marker for proteoglycans, was determined for each sample using the safranin O dot blot assay. 2 mL fractions were collected from a column containing 4 mL DEAE-Sephacel. Representative data from a single sample is shown. “Raw extract” refers to the initial sample in proteoglycan extraction buffer, and “Extract in G50” refers to the same sample after buffer exchange

used for identifying specific components. Enzymatic removal of GAGs from proteoglycans with the appropriate lyase is typically required to ensure full mobility in polyacrylamide gels and may be essential for proper epitope recognition by some antibodies.

## 4.4 Analysis of Isolated Proteoglycans by Mass Spectrometry

### 1. Sample preparation for mass spectrometry

- Lyophilize 20  $\mu\text{g}$  total protein in a SpeedVac evaporator.
- Reconstitute the dried protein in 50  $\mu\text{L}$  6 M urea, 100 mM Tris, pH 7.0.
- Reduce cysteine bonds by adding 2.5  $\mu\text{L}$  200 mM dithiothreitol (prepared fresh) and incubate at room temperature for 15 min.
- Alkylate the proteins by adding 10  $\mu\text{L}$  200 mM iodoacetamide (prepared fresh) and incubate at room temperature for 20 min, protecting from light.



- e. Quench the excess iodoacetamide by adding 10  $\mu\text{L}$  200 mM dithiothreitol and incubate at room temperature for 15 min.
- f. Reduce the urea concentration to approximately 1.2 M by diluting the sample with 160  $\mu\text{L}$  water.
- g. Adjust the pH to  $>8.0$  with 100 mM ammonium bicarbonate. This will likely take 20–30  $\mu\text{L}$  of ammonium bicarbonate; verify the correct pH by blotting a 1–2  $\mu\text{L}$  sample on litmus paper.
- h. Add trypsin at an enzyme:protein ratio of 1:20. Incubate at room temperature for 24 h or at 37  $^{\circ}\text{C}$  for 8–16 h.
- i. Desalt the trypsinized peptides using a C18 column such as a Pierce C18 spin column (ThermoFisher Scientific) following the manufacturer's instructions.
- j. Fully lyophilize the C18 eluate in a SpeedVac evaporator.
- k. Reconstitute the peptides in 30  $\mu\text{L}$  1% acetic acid.

Note: This protocol leads to injection of approximately 1  $\mu\text{g}$  of total peptides in 5  $\mu\text{L}$  on the liquid chromatography column. The starting amount of total protein (step a) and the final reconstitution volume (step k) can be adjusted to increase or decrease the final injection concentration.

## 2. Mass spectrometry

Tryptic peptides can be identified using a number of modern mass spectrometry instruments. We used an LTQ-Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific) for its high resolution and sensitivity.

- a. Separate peptides with an in-line liquid chromatography system, e.g., Dionex Ultimate 3000 nanoflow ultrahigh pressure liquid chromatography (UHPLC) system using a 75  $\mu\text{m} \times 15$  cm, 3  $\mu\text{m}$  particle size, Acclaim PepMap 100 C18 column (Thermo Fisher Scientific) at a flow rate of 0.3  $\mu\text{L}/\text{min}$ .
- b. Elute peptides over 2 h using buffers A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) with the following LC conditions:

Time (min)	Buffer B (%)
0–5	2
5–110	Linear gradient: 2–40
110–115	Linear gradient: 40–80
115–120	80

- c. The UHPLC column is coupled to a nanospray source through a PicoTip emitter (FS360-20-15-N-20-C15, New Objective).
- d. Collect spectra using a full-ion scan at a resolution of 60,000 over the mass/charge range 300–2000.  $\text{MS}^2$  scans using collision-induced dissociation (CID) can be performed on the 20 most abundant precursor ions from  $\text{MS}^1$  scans using the data-dependent mode with dynamic exclusion.

3. Bioinformatics—The precise bioinformatics approach used will vary according to the facility or user preference. The following approach was used with data from the LTQ-Orbitrap Elite hybrid mass spectrometer
  - a. MS<sup>2</sup> spectra were matched to the UniProtKB/Swiss-Prot human database using ProteomeDiscoverer (Thermo Scientific).
  - b. The percolator function was utilized to select only matches with a Q value < 0.01 (<1% false discovery rate [FDR]).
  - c. The mass tolerance was 10 ppm for precursor ions and 0.8 Da for MS<sup>2</sup>.
  - d. Only fully tryptic peptides were considered with a maximum of three missed tryptic cleavage sites.
  - e. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine was set as a variable modification.

Notes:

- i. Additional modifications (i.e. phosphorylation) can be included, which may increase peptide identification and improve core protein coverage.
- ii. Tryptic peptides containing GAG attachment sites will not be recognized unless the GAGs are enzymatically removed prior to analysis and the appropriate modifications are sought during spectral analysis (Noborn et al. 2015, 2016).

This detailed method provides a workhorse approach for isolating proteoglycans and can serve as the foundation for a proteomics study of core proteins, glycopeptide analysis or glycomics analysis. In all these applications, it ensures reduction of sample complexity and brings the PGs to the forefront. Using this approach to define the proteoglycanome of the ascending thoracic aorta, we identified 20 distinct proteoglycan core proteins (Cikach et al. 2018).

Pursuing a similar rationale, Talusan and colleagues used ion exchange chromatography isolation of proteoglycans coupled to proteomics analysis by gel-LC-MS/MS, in which the fractions obtained by urea extraction and High Q support strong AEC were electrophoresed on gels prior to tryptic digestion of the gel slices and mass spectrometry (Talusán et al. 2005). They found a strong correlation between proteoglycan abundance and species in an atherosclerosis-prone artery (internal carotid) and an atherosclerosis-resistant artery (internal thoracic) (Talusán et al. 2005). Vijayagopal et al. (1996) isolated LDL binding proteoglycans from atherosclerotic human arteries using a combination of orthogonal chromatographies. First, dissociative extraction and ion-exchange chromatography, similar to that described above, were used to isolate proteoglycans. Next proteoglycans were sub-fractionated on an LDL affinity column and proteoglycan fractions with high-affinity binding to LDL were analyzed for GAG species. The extracts from atheromatous plaques contained a high proportion of chondroitin and heparan sulfate proteoglycans, whereas normal aorta contained more diverse GAG species. Although core proteins were not evaluated, the data suggested enrichment of a small subset of proteoglycan species in atheromas, among which the CSPGs versican and biglycan have been shown to both bind LDL and be present in plaques (Didangelos et al. 2012; Wight

and Merrilees 2004). The CSPG aggrecan, typically a component of cartilage, was also reported in atheromatous plaques and stented coronary arteries (Talusán et al. 2005; Suna et al. 2018).

Mass spectrometry (MS) is used extensively for analysis of glycans, including GAGs. These applications digest the core protein to completion to preserve the GAGs and rely heavily on sophisticated bioinformatics approaches. This is because MS identifies molecules by mass and cannot readily distinguish between isomeric structures, e.g., glucose, mannose or galactose, although fragmentation of the glycan by different methods, coupled with other analytic techniques can help resolve its structure (Rojas-Macias et al. 2019). Profiling of glycans by MS requires dedicated isolation, derivatization and characterization techniques as well as high-resolution MS instruments (Rojas-Macias et al. 2019).

Whereas glycomics methods obliterate the core protein structure, glycoproteomics methods preserve the glycan-peptide linkage and can identify peptides to which glycans are attached, but these methods have limited ability to provide structural detail of the GAG. As an example of a glycoproteomics strategy, Noborn and colleagues have developed approaches to characterize HS and CS linkage regions, attachment sites, and identify novel proteoglycans (Noborn et al. 2015, 2016, 2018; Gomez Toledo et al. 2015). As in the method we describe here, their approach first requires isolation and enrichment of glycopeptides after trypsin digestion, using strong AEC. For identification of CS-attachment sites, for example, CS chains were depolymerized with chondroitinase ABC, leaving a residual hexameric structure composed of the linkage region and a GlcA-GalNAc disaccharide dehydrated on the terminal GlcA residue attached to the tryptic peptide. LC-MS/MS was then used to define the mass of the peptide with the residual hexasaccharide (Noborn et al. 2015).

## 4.5 Conclusions and Future Perspectives

In summary, we have briefly reviewed the immense biological significance of proteoglycans that justifies the need for specialized -omics approaches to obtain complete characterization, but it is obvious that this is a complex endeavor requiring multidisciplinary combinations of specialized techniques, instruments and skills. Regardless of whether the final goal is a proteoglycanome, glycoproteome or glycome, the strategy will employ some measure of enrichment of proteoglycans as an initial step. One can thus imagine several potential applications of the basic approaches we have outlined here. An important application is to identify the complete repertoire of “part-time” proteoglycans, since GAG-attachment contributes substantially to the functional diversity of the genome. Akin to our delineation of the aortic proteoglycanome, it is possible to use the approach to identify the proteoglycanome of any tissue or organ in development, health and disease. For example, taken together with isolation of specific organ structures such as renal glomeruli, or neural ganglia, proteoglycanomes of specific structures in organs can

also be achieved. The emerging explosion of single cell RNA sequence data for many tissues and organs, both adult and at many stages of embryonic development, will further allow a very refined determination of the proteoglycan repertoire. Another application that has been explored to only a limited extent is to define how proteoglycanomes differ in various diseases. This would be most informative if a quantitative analysis could be applied. One such approach would be to digest extracts of normal and disease tissue with trypsin, label the tryptic peptides with isobaric tags, and then combine the samples for glycopeptide enrichment steps prior to LC-MS/MS. In this way, variables introduced by preparation of each sample separately are eliminated, and moreover, the combined enriched glycopeptide pool can be analyzed in the same mass spectrometry run, avoiding run-to-run variation as well. The improved sensitivity of modern instruments and informatics tools that can combine glycopeptide analysis in multi-omics strategies, e.g., with single cell RNA sequencing, will likely make this approach important in understanding many diseases in the future.

**Acknowledgments** The authors express their gratitude to David A. Carrino, PhD for providing instruction in the proteoglycan isolation and quantitation methods used here and Belinda Willard PhD and Ling Li PhD of the Lerner Research Institute Proteomics and Metabolomics core for guidance in mass spectrometry. This work was supported by the Allen Distinguished Investigator Program, through support made by The Paul G. Allen Frontiers Group and the American Heart Association (to S.S.A).

## References

- Bandari S et al (2015) Sweet on Hedgehogs: regulatory roles of heparan sulfate proteoglycans in Hedgehog-dependent cell proliferation and differentiation. *Curr Protein Pept Sci* 16(1):66–76
- Bartlett AH, Hayashida K, Park PW (2007) Molecular and cellular mechanisms of syndecans in tissue injury and inflammation. *Mol Cells* 24(2):153–166
- Brinkmann T, Weilke C, Kleesiek K (1997) Recognition of acceptor proteins by UDP-D-xylose proteoglycan core protein beta-D-xylosyltransferase. *J Biol Chem* 272(17):11171–11175
- Buschmann MD, Grodzinsky AJ (1995) A molecular model of proteoglycan-associated electrostatic forces in cartilage mechanics. *J Biomech Eng* 117(2):179–192
- Calabro A et al (2001) Fluorophore-assisted carbohydrate electrophoresis (FACE) of glycosaminoglycans. *Osteoarthritis Cartilage* 9(Suppl A):S16–S22
- Carrino DA, Arias JL, Caplan AI (1991) A spectrophotometric modification of a sensitive densitometric Safranin O assay for glycosaminoglycans. *Biochem Int* 24(3):485–495
- Carrino DA et al (1994) Identity of the core proteins of the large chondroitin sulphate proteoglycans synthesized by skeletal muscle and prechondrogenic mesenchyme. *Biochem J* 298(Pt 1):51–60
- Cikach FS et al (2018) Massive aggrecan and versican accumulation in thoracic aortic aneurysm and dissection. *JCI Insight* 3(5)
- Didangelos A et al (2012) Novel role of ADAMTS-5 protein in proteoglycan turnover and lipoprotein retention in atherosclerosis. *J Biol Chem* 287(23):19341–19345
- Ezura Y et al (2000) Differential expression of lumican and fibromodulin regulate collagen fibrillogenesis in developing mouse tendons. *J Cell Biol* 151(4):779–788
- Fawcett JW, Oohashi T, Pizzorusso T (2019) The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. *Nat Rev Neurosci* 20(8):451–465

- Filmus J, Capurro M (2014) The role of glypicans in Hedgehog signaling. *Matrix Biol* 35:248–252
- Funderburgh JL (2002) Keratan sulfate biosynthesis. *IUBMB Life* 54(4):187–194
- Gomez Toledo A et al (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(12):3118–3131
- Gondelaud F, Ricard-Blum S (2019) Structures and interactions of syndecans. *FEBS J* 286(15):2994–3007
- Hascall VC, Heinegard D (1974) Aggregation of cartilage proteoglycans. I. The role of hyaluronic acid. *J Biol Chem* 249(13):4232–4241
- Heinegard D, Saxne T (2011) The role of the cartilage matrix in osteoarthritis. *Nat Rev Rheumatol* 7(1):50–56
- Iozzo RV, Schaefer L (2015) Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. *Matrix Biol* 42:11–55
- Ly L, Wasinger VC (2011) Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. *Proteomics* 11(4):513–534
- Mead TJ et al (2018) ADAMTS9-regulated pericellular matrix dynamics governs focal adhesion-dependent smooth muscle differentiation. *Cell Rep* 23(2):485–498
- Midura RJ et al (2018) Quantification of hyaluronan (HA) using a simplified fluorophore-assisted carbohydrate electrophoresis (FACE) procedure. *Methods Cell Biol* 143:297–316
- Mitsou I, Multhaupt HAB, Couchman JR (2017) Proteoglycans, ion channels and cell-matrix adhesion. *Biochem J* 474(12):1965–1979
- Nastase MV et al (2018) Small leucine-rich proteoglycans in renal inflammation: two sides of the coin. *J Histochem Cytochem* 66(4):261–272
- Noborn F et al (2015) Identification of chondroitin sulfate linkage region glycopeptides reveals prohormones as a novel class of proteoglycans. *Mol Cell Proteomics* 14(1):41–49
- Noborn F et al (2016) Site-specific identification of heparan and chondroitin sulfate glycosaminoglycans in hybrid proteoglycans. *Sci Rep* 6:34537
- Noborn F et al (2018) Expanding the chondroitin glycoproteome of *Caenorhabditis elegans*. *J Biol Chem* 293(1):379–389
- Ortmann C et al (2015) Sonic hedgehog processing and release are regulated by glypican heparan sulfate proteoglycans. *J Cell Sci* 128(12):2374–2385
- Robinson KA et al (2017) Decorin and biglycan are necessary for maintaining collagen fibril structure, fiber realignment, and mechanical properties of mature tendons. *Matrix Biol* 64:81–93
- Rojas-Macias MA et al (2019) Towards a standardized bioinformatics infrastructure for N- and O-glycomics. *Nat Commun* 10(1):3275
- Rowlands D et al (2018) Aggrecan directs extracellular matrix-mediated neuronal plasticity. *J Neurosci* 38(47):10102–10113
- Sarrazin S, Lamanna WC, Esko JD (2011) Heparan sulfate proteoglycans. *Cold Spring Harb Perspect Biol* 3(7)
- Scott JE (1985) Proteoglycan histochemistry—a valuable tool for connective tissue biochemists. *Coll Relat Res* 5(6):541–575
- Suna G et al (2018) Extracellular matrix proteomics reveals interplay of aggrecan and aggrecanases in vascular remodeling of stented coronary arteries. *Circulation* 137(2):166–183
- Talusan P et al (2005) Analysis of intimal proteoglycans in atherosclerosis-prone and atherosclerosis-resistant human arteries by mass spectrometry. *Mol Cell Proteomics* 4(9):1350–1357
- Thelin MA et al (2013) Biological functions of iduronic acid in chondroitin/dermatan sulfate. *FEBS J* 280(10):2431–2446
- Tsien RY (2013) Very long-term memories may be stored in the pattern of holes in the perineuronal net. *Proc Natl Acad Sci U S A* 110(30):12456–12461
- van Wijk XM, van Kuppevelt TH (2014) Heparan sulfate in angiogenesis: a target for therapy. *Angiogenesis* 17(3):443–462

- Vijayagopal P et al (1996) Isolation and characterization of a proteoglycan variant from human aorta exhibiting a marked affinity for low density lipoprotein and demonstration of its enhanced expression in atherosclerotic plaques. *Atherosclerosis* 127(2):195–203
- Wight TN, Merrilees MJ (2004) Proteoglycans in atherosclerosis and restenosis: key roles for versican. *Circ Res* 94(9):1158–1167
- Yu WH, Woessner JF Jr (2000) Heparan sulfate proteoglycans as extracellular docking molecules for matrilysin (matrix metalloproteinase 7). *J Biol Chem* 275(6):4183–4191
- Zhang G et al (2006) Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. *J Cell Biochem* 98(6):1436–1449
- Zimmermann DR, Dours-Zimmermann MT (2008) Extracellular matrix of the central nervous system: from neglect to challenge. *Histochem Cell Biol* 130(4):635–653