Recent developments in proteoglycan purification and analysis

Proteoglycans are ubiquitous biomolecules in the body located in the extracellular matrix, on the cell surface and also within the cells. They contain at least one glycosaminoglycan (GAG) chain covalently attached to a core protein and may also present N- or O-linked glycans. The high structural diversity and distribution relate to the various biological functions of proteoglycans. In recent years, new members have enlarged the proteoglycan family and advances in molecular biology and glycobiology contributed to elucidate more of the biological functions of proteoglycans. In order to study the structure of a proteoglycan molecule and relate it to its function (or dysfunction), its isolation and purification from cell culture or tissue extracts is necessary. Next to the widely used anion exchange chromatographic methods, techniques based on lectin affinity chromatography have created new possibilities to increase the degree of purity. Introduction of electrospray ionization (ESI) and matrix–assisted laser desorption-ionization (MALDI) sources, together with tandem mass spectrometry (MS/MS or MSⁿ), mark a further important step towards the structural analysis of glycosaminoglycans. The aim of this review is to present the most recent advances in proteoglycan purification and analysis.
2.1 Introduction

Proteoglycans are ubiquitous molecules in eukaryotic species carrying as hallmark at least one glycosaminoglycan (GAG) side chain covalently attached to a protein core. GAGs are linear polymeric structures composed of repeating disaccharide units of one hexosamine (D-glucosamine or D-galactosamine) alternating with one uronic acid (D-glucuronic acid or L-iduronic acid) or galactose (keratan sulfate) [1]. Depending on the disaccharide composition, GAGs can be divided into four categories: chondroitin sulfate (CS) and dermatan sulfate (DS), heparin and heparan sulfate (HS), keratan sulfate (KS) and hyaluronan (HA). Hyaluronan, the largest GAG (> 2500 disaccharide units) [2], is the only non-sulfated GAG and is not covalently bound to a protein core. It is considered a PG because it associates with a wide variety of proteins. HA is synthesized by one of three different transmembrane hyaluronan syntheses (HAS1, HAS2 and HAS3) with the active sites on the inner face of the plasma membrane [3]. During synthesis, the growing HA chain is extruded into the extracellular space [4,5], allowing unconstrained polymer growth. All other GAGs are synthesized and assembled as PGs in the endoplasmic reticulum and Golgi apparatus and secreted in a way similar to other glycoproteins. CS, DS, heparin and HS are built on a common tetrasaccharide sequence (-O-Xyl-Gal-Gal-GlcA-), which is O-linked to the protein cores via a serine or threonine residue, while KS may be attached to the protein cores via both O- and N-linkages [6-8]. The type of GAG present on a given proteoglycan may vary among different cells and different GAG types may be present on the same protein core (e.g. aggrecan, syndecan) [9]. Therefore, for the classification of proteoglycans it is necessary to mention their tissue location (intracellular, on the surface of the cells or in the extracellular matrix) in addition to the GAG type.

The distribution of proteoglycans is variable. They are abundantly expressed by fibroblasts and chondrocytes, but also by cell types that do not typically take part in extracellular matrix or tissue construction, like those of the immune system or metastasizing tumor cells [10]. Articular cartilage, blood vessels, skin, and cornea are a few examples of tissues rich in proteoglycans. The structure and size of these molecules may vary enormously, particularly with respect to their glycosaminoglycan content. Perlecan is the largest characterized proteoglycan, possessing a 4400 amino acid protein core with three GAG substitution sites [11]. Aggrecan may bind 100 CS and 30 KS chains to its 2000 amino acid protein core [12]. Leucine-reach repeat proteoglycans have relatively short core proteins (300-400 amino acids), but are still larger than some cell surface proteoglycans (e. g. serglycin) possessing less than 200 amino acids [13].

The structural diversity and distribution of proteoglycans determines their various functions. Although it is difficult to discriminate between functions mediated by the GAGs themselves or by the protein cores, it seems that GAGs are responsible for targeting of PGs to the proper (extra)cellular environment [9], modulation of growth-factor activities [14], binding and organization of water molecules, repulsion of negatively charged molecules, while the protein cores are mostly mediating protein-protein interactions [15-17]. Through the combination of GAG and protein core mediated functions, PGs act as tissue organizers [18], influence tissue’s mechanical properties [19], control cell proliferation (e.g. tumor cell growth and invasion) [20], regulate collagen fibrillogenesis [21] and skin tensile strength [22], and influence corneal transparency [23]. Heparin is well known for its anticoagulant activity [24]. The role of DS as anticoagulant has also been proven [25], as well as its activity against venous thrombosis [26]. In the neural system sulfated PGs behave as axon guidance molecules during brain development [27,28] and are involved in the repair process of injured Central Nervous System (CNS) and disrupted Blood Brain Barriers (BBD) [29]. The compelling evidence for participation of proteoglycans in a multitude of biological processes.
as well as the association of hampered proteoglycan functions with disease development or progression (e.g. rheumatoid arthritis, atherosclerosis, vascular and pulmonary diseases) call for additional, more detailed structural investigations which, in turn, stimulate further methodological developments.

The purpose of this review is to highlight the developments in analytical strategies applied to proteoglycans, from their isolation and purification toward their structural analysis, revealing especially the composition of the attached glycosaminoglycans. Special attention will be accorded to mass spectrometry, which has contributed significantly to the detailed analysis of PGs, although we feel that its potential has not been fully exploited. The structural analysis of branched N- and O-linked oligosaccharides, while sharing some methodological aspects with glycosaminoglycans, is beyond the scope of this review. This topic was covered by Mechref and Novotny, who reviewed the methodological advances in the analysis of glycoproteins and their glycan structures [30].

2.2 Proteoglycan purification

Proteoglycans can be purified from tissues, cells or cell culture supernatants using similar approaches. In this review we will focus on the purification of PGs from tissue, as this is the most complex starting material for proteoglycan purification. The purification of proteoglycans from tissue is often complicated by limited availability, their large molecular size and the necessity of chaotropes for efficient extraction.

2.2.1 Extraction from tissue

Slicing, grinding, or pulverization is required in order to increase the area of tissue available for extraction and to reduce the extraction time. Tissue homogenates are further incubated with suitable extraction buffers for 12 to 48 h followed by re-extraction of the pellet resulting after centrifugation. The extraction yields are tissue and time dependent. Inactivation of proteases is accomplished by including protease inhibitors in the extraction buffers, performing the extraction at low temperatures (4°C) and under pH conditions unfavorable for protease activity. Occasionally wide-spectrum proteases (proteinase K or papain) can be used to solubilize the tissue when the interest lies mainly in GAG characterization.

Different concentrations of several salts (LiCl, CaCl₂, MgCl₂) [31] have been employed for proteoglycan extraction, promoting solubilization by dissociating intermolecular interactions between proteoglycans and other molecules. In 1969, Sajdera and Hascall [32] reported the extraction of cartilage proteoglycans with 3-4M guanidine hydrochloride (GuHCl) followed by further purification by cesium chloride (CsCl) density gradient centrifugation. 4M GuHCl has proven to offer the best extraction yield and has become the most widely used extraction agent. Brand et al [33] extracted bovine articular cartilage with increasing concentrations of GuHCl and concluded that the ionic strength employed for the efficient solubilization of PGs is directly proportional to their size. However, for tissues less abundant in proteoglycans, GuHCl extraction results also in a high amount of other proteins. Chaotropes and detergents are especially necessary for extraction of proteoglycans localized intracellular or in cell membranes preventing formation of large aggregates. If for example cell cultures are the target, addition of 0.1% Triton X-100 may solubilize cell membranes. Inositol hexaphosphate [34,35] or phosphatidylinositol specific phospholipase C [34,36] are needed for releasing membrane associated PGs. EDTA dissolves hydroxyapatite mineral and therefore inclusion of EDTA in the extraction buffer permits to distinguish between matrix associated PGs (4M GuHCl soluble) and mineral associated PGs (4M GuHCl insoluble/4M GuHCl + 0.5M EDTA soluble) in fetal porcine calvarial bone [37].
HSPGs associated with murine erytholeukemia cells were separated into two urea-soluble classes (with the molecular weights of 280 kDa and 80 kDa) versus one high molecular weight (360 kDa) urea-insoluble [38] class. Extensive denaturation of the protein components of proteoglycans occurs in concentrated solutions of these chaotropic reagents. Removal of denaturant by dialysis does not always restore the protein conformation to its native state. When it is intended to use the protein component for further activity studies it is necessary to preserve its native structure to avoid compromising the biological activity. Chaotrope-free extraction buffers have been described for soluble PGs [39], preserving their native state, but with relatively lower extraction yields.

### 2.2.2 Purification

Following their extraction, proteoglycans are isolated and further separated into different classes based on their physico-chemical properties. The most important feature of PGs is their polyanionic nature conferred by sulfate substituents on carbohydrate residues and by carboxyl groups of uronic acid moieties. In addition, some protein cores may contribute also specific properties. Membrane PGs (e.g. perlecan, agrin, bamacan) possess a phosphatidylinositol moiety, which increases their hydrophobicity [40], freely soluble secreted PGs (biglycan, decorin and fibromodulin) show an amphipathic nature [41-43], while aggrecan or versican have peptide sequences that confer lectin-like properties [44]. Therefore, using different types of chromatography, PGs may be fractionated based on charge density (anion exchange chromatography), size and shape (size exclusion chromatography), surface hydrophobicity (hydrophobic interaction chromatography), or interaction with affinity ligands recognizing the distribution of specific amino acids or carbohydrate residues at their surface (affinity chromatography). An example of the work flow used to purify decorin core protein from human lung tissue is presented in Fig. 1.

![Diagram](image.png)

**Fig. 1.** Schematic representation of the work flow used to purify decorin (DCN) core protein from human lung tissue (Reprinted from ref. [65] Copyright © 2006, with permission from Elsevier.)
Proteoglycans can be separated as intact molecules or subjected to chemical/enzymatic treatment depending on whether the interest lies in the structural characterization of the GAG chain(s) or the protein cores. Separation of intact PGs is usually achieved by gradient density centrifugation or by chromatographic methods like size exclusion or anion exchange chromatography.

Centrifugation in density gradients of cesium chloride (CsCl) or cesium bromide (CsBr) with addition of protein denaturants separates proteoglycans from non-covalently associated proteins and constitutes the first step in most preparative methods. One run typically takes 2-3 days. By using either non-denaturing (0.5M GuHCl) or denaturing (4M GuHCl) buffers, PGs can be isolated as aggregates (with hyaluronan and link protein) or as monomers [45].

Anion-exchange chromatography (AEC) is the method of choice for the isolation/purification of PGs extracted from tissues or cells with a relatively low ratio of proteoglycan to protein. Separation of PGs by AEC involves binding of PGs to a cationic stationary phase via their anionic GAG chains followed by elution with elevated salt concentration. The most common resins employed in AEC of PGs contain the weak anion exchanger diethylaminoethyl (DEAE) groups immobilized on a hydrophilic support material (e.g. DEAE-Sepharose [46], DEAE-Sephacel [47], DEAE-Cellulose [48]). DEAE groups possess a net positive charge at neutral to slightly acidic pH. The columns are generally equilibrated in urea-containing buffers (e.g. 7M urea, 0.05 M sodium acetate, pH 6-7, 0.5% detergent) and eluted with NaCl gradients [49]. Strong anion-exchange chromatographic supports containing quaternary ammonium groups have also been used (Mono Q [47], Q-Sepharose [39,50]). These supports are positively charged independent of pH. These resins are usually used to concentrate PGs by bulk volume loading and step elution in high salt buffers [39,51] (1 to 2 M NaCl).

Size exclusion chromatography (SEC) fractionates PGs based on their hydrodynamic volume. Large PGs will be excluded from the internal pore volume and exit the column first, while smaller PGs will traverse the pores of the stationary phase and hence elute later. SEC employing porous gel beads of polymerized dextran (Superdex, Sephadex), agarose (Sepharose, Sephacryl, Superose), polyacrylamide (Bio-Gel), or silica (Bio-Sil, TSK-Gel) have found applications in proteoglycan purification protocols. Decorin was purified from the culture medium of MRC-5 cells using gradient elution from an anion exchange column (HiTrap Q HP) followed by two successive steps of gel filtration on Superdex 200 columns [52]. Rapid SEC of intact PGs in buffers containing up to 8 M urea was performed with a BioSil Sec-400 column, which is compatible with the high urea concentration and has an upper back-pressure limit of more than 10 MPa [53]. The addition of poly-L-glutamic acid to both sample and mobile phase overcomes the non-specific binding of PGs to the silica matrix [53]. Another way to avoid electrostatic interaction between PGs and the resin is to increase the ionic strength of the eluent to 0.05 M salt. Two tandem Bio-Gel columns (Bio-Gel TSK-60 and Bio-Gel TSK-50) were used to resolve intact cartilage proteoglycans from proteolytic degradation products present in the extract [54]. SEC can also provide an estimation of the weight average (Mw) of proteoglycans, when the column is calibrated with PG standards of known molecular weight. Ohno and co-workers used a Sepharose CL-2B column, equilibrated in 4M GuHCl, to derive the molecular mass of PGs from chicken chondrocytes [55]. Proteoglycans extracted from ovine skeletal muscle were separated into two populations by a Sephacryl S-200HR column [56]. Cooper et al [57] employed a Superose 6 HR column to purify a keratan sulfate/chondroitin sulfate proteoglycan from the culture medium of human embryonal carcinoma stem cells. SEC may also be utilized in place of dialysis to remove salts and/or unincorporated radiolabel [51]. Separation based on surface hydrophobicity is accomplished by either hydrophobic interaction chromatography (HIC) or
reversed-phase chromatography (RPC). In general, HIC resins contain a much lower ligand density as compared to RPC resins. In HIC, elution from the stationary phase (short chains of octyl or phenyl groups) takes place with descending salt gradients or increasing non-ionic detergent gradients [58], while in RPC the bound compounds are eluted with organic solvent. Octyl-Sepharose has been efficiently used for separating small leucine-rich PGs (biglycan, decorin and fibromodulin) from each other [59,60] and also for isolating membrane associated PGs [61-63]. Phenyl-Sepharose has proven successful for fractionation of retinal cell-associated PGs [64]. Reversed phase columns have not generally been used for proteoglycan purification. However, for the separation of the protein cores of small proteoglycans, reversed phase chromatography may be helpful and presents some advantages over other types of chromatography. In our own research we used a C4 reversed phase column in the last step of a decorin purification protocol [65]. Reversed phase chromatography separates decorin from the other proteoglycan protein cores (see Fig. 2) and offers the possibility to profile the proteoglycans from the CS/DS class, gaining more insight into the proteoglycan composition of lung tissue.

![Graph](image)

**Fig. 2.** Separation of decorin core protein (F) from other CS/DS proteoglycan components from a lung extract using a Vydac protein C4 column. (Reprinted from ref. [65] Copyright © 2006, with permission from Elsevier.)

Hydroxyapatite (HAP) presents a complex mosaic of charges (contributed by calcium, phosphate and hydroxyl groups), which give rise to its amphoteric nature. Electrostatic interactions occur between the cationic sites of the mineral (calcium) and the anionic groups of the analyte (carboxyl and sulfate). For chromatography, most columns are equilibrated by extensive washing with phosphate buffer at pH 6.8 [49], which confers a negative net charge to the resin’s surface. Carboxyl groups on the protein core or GAGs are repelled electrostatically from the negatively charged phosphate groups of the column and bind by complexing to calcium. PGs can thus interact with the resin via both the protein cores and the charged GAGs. Protein conformation is an important determinant in HAP interaction, since
only native proteins bind to HAP effectively [66]. Elution of PGs from HAP is performed with an increasing gradient of phosphate. A Bio-Gel HTP hydroxyapatite column was employed to fractionate human aorta PGs [67], porcine bone PGs [37], or rat microvascular endothelial cell PGs [68].

Heparan sulfate proteoglycans bind to a large number of proteins and are therefore most frequently isolated by affinity chromatography. Different proteins immobilized to chromatographic resins were employed to affinity purify this class of PGs, antithrombin III [68], TGF-β [69], laminin [70], fibronectin [35] or thrombospondin [71] being only few examples. Monoclonal or polyclonal antibodies immobilized on chromatographic resins can also be used in immunoaffinity purifications of different proteoglycans [72-74]. Another affinity method for the isolation of PGs takes into account the N- or O-linked oligosaccharides present on their protein cores in addition to the GAG chains. Different lectins, with particular specificity for a certain oligosaccharide sequence, may be directed to affinity capture PGs by binding to their specific glycans. Cooper et al. used a peanut lectin affinity column [57] as the first step in their protocol to purify a keratan sulfate/chondroitin sulfate proteoglycan expressed in pluripotent primate stem cells. Brown et al [39] used Concanavalin A as the last step when purifying decorin from bovine cornea.

2.2.3 Specific assays

In any purification procedure it is necessary to have an assay to follow the yield of the component of interest at each step of the protocol. Many assays for PG quantification take advantage of the attached GAGs. For uronic acid containing PGs (HS, CS, DS) the colorimetric carbazole-sulfuric acid assay, as modified by Muir and coworkers [75], can be applied. The method involves the hydrolysis of hexuronic acid derivatives with carbazole and absorbance measurement at 525 nm. The assay is linear between 4-400 µg of standard (D-glucurono-6,3-lactone). Neutral glycans give about 10% interference and heparin and HS produce an anomalously high absorbance.

Another way to quantify PGs is through interaction of their GAGs with cationic dyes. The method involves precipitation of the dye-GAG complexes followed by dissociation of the complexes and resolubilization of the dye. A spectrophotometer is employed to read the absorbance at wavelengths characteristic for each dye. Various dyes, like alcin blue, safranin O, cuprolinic blue or 1,9-dimethylmethylene blue have been employed in assays for GAG quantitation. The same principle can be utilized on a larger scale to selectively precipitate PGs from tissue extracts and to separate them from non-proteoglycan components. McBain and Muller [76] treated cultured cells with cetylpyridinium chloride (CPC) followed by phenol extraction of the pellet (to remove unglycosylated proteins) to yield lipophilic PGs of high density. Bjornsson optimized a method to quantify PGs (1-20 µg GAG) after precipitation with alcin blue from extracts containing GuHCl. Different GAG classes can be specifically precipitated by increasing concentrations of GuHCl in combination with detergent and high salt concentration at low pH [77]. This method was further optimized for a more sensitive quantification of GAGs in biological fluids (blood plasma, urine or wound fluid) [78]. Attention should be paid when choosing the supplier of alcin blue, because the solubility of alcin blue in GuHCl has a major impact on the success of the assay [77]. In all dye-binding assays interferences from other proteins, nucleic acids or hyaluronan are still noticeable to a certain degree. Therefore their application for a sensitive detection is limited, but they can be successfully employed to monitor the enrichment of PGs in fractions of interest along different steps of a purification protocol.

The ability of cationic dyes to bind strongly to GAGs is taken into account when developing specific staining methods to visualize PGs after electrophoretic separation. Alcian
blue or toluidine blue [79] are traditionally employed for staining PGs. McDevitt and Muir [80] developed a composite acrylamide-agarose (1.2% acrylamide, 3% cross linker and 0.6% agarose) gel to separate cartilage PGs. For staining the PGs they used 0.2% toluidine blue in 0.1% acetic acid (alcian blue stains agarose and therefore was not suitable). Gradient SDS-PAGE systems for PGs have also been used, the most common gradients being in the range of 3 to 20% acrylamide [81,82]. The staining methods are characterized by low sensitivity, requiring microgram amounts of each proteoglycan for detection. In order to increase sensitivity, silver enhancement methods have been described [83,84]. Alcian blue is used as the primary staining agent, binding directly to the proteoglycans, with staining being subsequently enhanced by a neutral silver stain. A twofold increase in sensitivity as compared to alcian blue alone is obtained. Nonglycosylated proteins stain weaker than PGs. Radiolabeled PGs separated by composite or SDS-PAGE gels can be visualized by fluorography [85,86].

2.3 GAG analysis

The early work of Meyer and others [87] has established the structure of the basic repeating units of each group of sulfated GAGs. Over the years it has, however, become apparent that GAG structure is very heterogeneous with respect to position and distribution of D-glucuronic and L-iduronic acid groups and the content and location of sulfates. Very detailed information about the structure of GAGs is needed in order to elucidate structure-function relationships. PGs are often involved in biological processes necessitating interactions via well defined GAG sequence motives with defined sulfation patterns. This can be illustrated by the fact that pathogens (viruses, parasites, and bacteria) use cell-surface CS/DS and HS with definite sulfation patterns and chain lengths for their attachment to host cells and tissues [88] and that a defined sulfated heparin pentasaccharide unit is critical for the activation of antithrombin III, a major controlling element in the coagulation cascade [89]. The various sulfation patterns in CS have also been suggested as a marker to distinguish osteoarthritis from normal aging [90], since the sulfation pattern of low molecular weight dermatan sulfate plays a key role in its antithrombotic activity in blood vessels.

2.3.1 Intact GAGs

Information about the charge density, polydispersity, molecular size and GAG type can be obtained from the analysis of intact GAGs. Since the detailed structure of glycosaminoglycans cannot be currently assessed at the level of intact proteoglycans, their study involves as a first step the release of the GAG chains from the protein cores. Alkaline β-elimination [91] (used also in the Carlson degradation procedure) and hydrazinolysis [92] are the two most commonly used chemical cleavage methods available for detaching the O-linked GAGs from proteoglycan protein cores. More recently, enzymatic cleavage by cellulases, a class of enzymes with endo-β-xilosidase activity, have also been described to release intact GAG chains attached to peptides [93,94]. The cellulases from A. niger and Patinopecten act on the linkage region (Xyl-Ser) between a core peptide and a GAG chain by hydrolyzing the peptide. Patinopecten endo-β-xilosidase catalyzes hydrolysis of long GAG chains, hardly cleaving short chains, while for A. niger endo-β-xilosidase the enzymatic activity is greater with shorter GAG chains [93].

Once GAGs have been released from the protein cores, the resultant mixture may be fractionated by ion exchange chromatography and subjected to further separation by electromigration methods. All GAG species are amenable to electrophoresis on cellulose acetate membranes [95], agarose gels [96,97], or polyacrylamide gels and classified according
Proteoglycan purification and analysis

to the mobility of standards. Improved detection limit may be achieved by combining different staining agents, like sequential staining with toluidine blue and Stains-All [97], which allows detection of submicrogram quantities of proteoglycans. An example of the separation and detection of a mixture containing 0.5 µg CS and 0.5 µg DS on agarose gel is presented in Fig. 3. Electrophoresis on cellulose acetate membrane is still the first choice for the analysis of clinical samples, due to its simplicity, low cost of analysis, and the ability to process several samples at the same time. Bands are visualized after colorimetric reaction with a cationic dye (mainly toluidine blue) and quantified using image analysis systems. Due to the heterogeneity of GAGs, bands are diffuse and details about minor structural changes or the sites of sulfation cannot be obtained. Microchip electrophoresis with in situ fluorescent detection using ethidium bromide has been proposed as an alternative to cellulose acetate electrophoresis [98], offering the same sensitivity and shortening the analysis time to 150 s. Electromigration procedures employed to analyze GAG mixtures have been recently reviewed [99].

SEC [100] or capillary electrophoresis [101,102] may be also used for intact GAG analysis. Intact HA has been analyzed by CE under both normal and reversed polarity. In normal polarity (phosphate borate buffer, pH 9) it appears as a single broad peak [103], while in reversed polarity under weak acidic condition it migrates as a narrow peak [104]. CE measurements of HA in biological matrices (e.g. synovial fluid) is affected by the interaction with proteins, which may be overcome by the addition of SDS [102], also used for the analysis of long HA derived oligosaccharides [105]. Intact HA was also detected utilizing copper complexes [106], but compared to other copper-GAG complexes, HA gives low detection sensitivity. When separated in a capillary filled with a polyacrylamide gel, HA oligomers show a linear relationship between migration times and degree of polymerization [107].

The requirements for establishing the biological function of GAGs call for very detailed sequence information, which can not be obtained from the analysis of intact GAGs. Therefore, more attention has been recently focused on the analysis of GAG derived oligosaccharides.

Fig. 3. (A) Separation of 0.5 µg of chondroitin sulfate and 0.5 µg of dermatan sulfate by discontinuous agarose-gel electrophoresis in 0.05 M HCl/0.04 M barium acetate combined with the visualization technique using toluidine blue/Stains-All. (B) Densitometric scan of the gel. (Reprinted from ref. [97] Copyright © 2005, with permission from Elsevier.)
2.3.2 GAG-derived oligosaccharides

GAG structure is traditionally studied by scission of the chains with specific enzymes in order to produce GAG-derived oligosaccharides. A large variety of degrading enzymes for all GAG types is nowadays commercially available. These enzymes are divided into two distinct classes: prokaryotic enzymes, which depolymerize GAGs by an elimination mechanism and eukaryotic enzymes, which act by hydrolysis [108]. Each cleavage, with either class of enzyme, creates a free reducing aldehyde group that can be labeled with fluorescent dyes. The most commonly used enzymes belong to the chondroitin/dermatan lyase and heparin/HS-lyase families. These enzymes act on the β(1→4) bond between hexosamine and uronic acid and produce disaccharides with a double bond between C4 and C5 of uronic acid, which absorbs strongly at 232 nm.

Table 1. Different glycosydases employed for GAG digestion.

<table>
<thead>
<tr>
<th>Enzyme name</th>
<th>Enzyme source</th>
<th>Enzyme substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heparin lyase I</td>
<td>Flavobacterium heparinum</td>
<td>Heparin</td>
</tr>
<tr>
<td>Heparin lyase II</td>
<td>Flavobacterium heparinum</td>
<td>Heparin</td>
</tr>
<tr>
<td>Heparin lyase III</td>
<td>Flavobacterium heparinum</td>
<td>HS</td>
</tr>
<tr>
<td>Chondroitinase ABC</td>
<td>Proteus vulgaris</td>
<td>CS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HA-less effectively</td>
</tr>
<tr>
<td>Chondroitin Lyase AC</td>
<td>Arthrobacter aurescens</td>
<td>CS (glucuronic acid)</td>
</tr>
<tr>
<td></td>
<td>Flavobacterium heparinum</td>
<td>DS (glucuronic acid)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HA</td>
</tr>
<tr>
<td>Chondroitin Lyase B</td>
<td>Flavobacterium heparinum</td>
<td>DS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cleaves amino sugars attached to iduronic acid</td>
</tr>
<tr>
<td>Testicular hyaluronidase</td>
<td>Bovine/sheep testes</td>
<td>Hyaluronan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
</tr>
<tr>
<td>Hyaluronidase</td>
<td>Streptomyces hyalurolyticus</td>
<td>HA</td>
</tr>
<tr>
<td>Hyaluronidase</td>
<td>Streptococcus dyslatea</td>
<td>HA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS unsulfated regions</td>
</tr>
<tr>
<td>Endo-β-galactosidase</td>
<td>E. freudii</td>
<td>KS</td>
</tr>
<tr>
<td>Keratanase I</td>
<td>Pseudomonas species</td>
<td>KS</td>
</tr>
<tr>
<td>Keratanase II</td>
<td>Bacillus species</td>
<td>KS</td>
</tr>
</tbody>
</table>

An overview of the most common enzymes employed for GAG degradation is presented in Table 1. HA, CS and DS are decomposed in disaccharides by chondroitinases (ABC, AC and B). These enzymes have specificities for chondroitin sulfate (ABC and AC) and dermatan sulfate (ABC and B), while chondroitinase ABC cleaves also hyaluronan but less effectively. Testicular hyaluronidase degrades both hyaluronan and chondroitin sulfate while the hyaluronidase from Streptomyces is specific for hyaluronan. Keratan sulfate can be depolymerized by endo-β-galactosidase from E. freudii or keratanases from Pseudomonas or Bacillus species. Heparin and heparan sulfate are cleaved by different heparin liases. Each
Proteoglycan purification and analysis

heparinase can yield a different pattern of products, while their combined use yields disaccharides. In Fig. 4 are depicted the major disaccharide repeating units found in each glycosaminoglycan.

**Heparin**

[Structural diagram of Heparin with major and minor disaccharide sequences]

**Heparan sulfate**

[Structural diagram of Heparan sulfate with major and minor disaccharide sequences]

**Keratan sulfate**

[Structural diagram of Keratan sulfate with major and minor disaccharide sequences]

**Hyaluronic acid**

[Structural diagram of Hyaluronic acid]

**Dermatan sulfate** (Chondroitin sulfate B)

[Structural diagram of Dermatan sulfate]

**Chondroitin-4-O-sulfate** (Chondroitin sulfate A)

[Structural diagram of Chondroitin-4-O-sulfate]

**Chondroitin-6-O-sulfate** (Chondroitin sulfate C)

[Structural diagram of Chondroitin-6-O-sulfate]

Fig. 4. Structures of the major disaccharide repeating units found in different types of glycosaminoglycans. Heparin, HS and KS have diverse sequences where \( X = \text{SO}_3^- \) or H and \( Y = \text{SO}_3^- \), Ac or H. (Reproduced with permission from ref. [134] Copyright © 2002, John Wiley & Sons Limited.)
Compositional analysis of GAGs represents a major methodological challenge in terms of sample complexity and can be compared to a certain extent with peptide mapping of proteins. Differential treatments of the same GAG chain with all the available specific lyases generate oligosaccharide maps, similar to the peptide maps derived from proteins. Recently, Iwafune and coworkers proposed a method for estimating the molecular weight and degree of sulfation of intact GAGs, liberated from PG extracts by digestion with cellulase from *Aspergillus niger* [109]. Size exclusion chromatography on a Shodex OHpak SB-803HQ column was used to estimate the molecular weights of GAG species, while anion exchange on a TSKgel DEAE-5PW column estimates the numbers of sulfates per disaccharide.

**High performance liquid chromatography (HPLC)**

During the last years improved sensitivity of detection systems for liquid chromatographic and electrophoretic instruments has made separation and quantification of very low amounts of GAGs possible. Different types of chromatography such as normal phase, reversed-phase, ion-pair and ion-exchange have been used for disaccharide and oligosaccharide analysis. Analysis of unsaturated disaccharides is simple and rapid, without the need for derivatization. However, derivatization increases the sensitivity of detection and therefore, derivatization has been proved useful for the detection of both saturated and unsaturated disaccharides. The most common agents employed for pre-column derivatization are: 2-Aminoacridone (2-AMAC) [110], 2-Aminopyridine (2-AP) [111], 2-Aminobenzamide (2-AB) [112], and dansylhydrazine [113].

HPLC methods do not generally provide complete sequence information of the GAG chain, but are able to reveal the sulfation pattern/degree of modification. Quantitative or qualitative changes (molecular size, charge density) can be revealed by HPLC, providing information about the development and progress of a certain disease, for instance is the case in different neoplastic transformations [114,115].

First columns employed for the separation of GAG derived oligosaccharides used amino-bonded silica as packing material and separation was performed isocratically [116]. Nowadays, reversed phase chromatography together with ion-pairing of glycosaminoglycans with hydrophobic amines are more widely used [117,118].

Different GAG chains obtained from human gastric carcinoma (HGC) and human normal gastric mucosa (HNG) were fractionated by anion exchange chromatography (AEC) on DEAE-Sephacel column, revealing the presence of three different populations (I, II, and III) in both normal and cancerous tissue [115] (Fig. 5). Quantitative differences in populations I and II were revealed and subsequent analysis showed a significant increase in HA and CS (population I) and a decrease in HS (population II) in the samples originating from carcinoma tissue.

Ion-pair reversed-phased HPLC, employing binding of a tetrabutylammonium hydrogen sulfate (TBA) cation to the anionic charges of oligosaccharides, offers high separation for all non-, mono-, di- and trisulfated disaccharides [119]. Gel permeation chromatography on a Sepharose CL-6B column has been used to determine the absolute content and the size distribution of digested dermatan sulfate [120].

Endogenous unsaturated disaccharides from human plasma, produced after digestion with chondroitinase ABC were separated and quantified using a Symmetry C-18 analytical column [121]. The separation was performed with gradient elution starting with 96% eluent A (0.8 mM tetrabutylammonium hydrogen sulfate and 2 mM sodium chloride) and 4% acetonitrile, to 47% eluent A, 6% acetonitrile and 47% water. The absorbance of the eluate was monitored at 230 nm.
Fig. 5. Fractionation of glycosaminoglycan chains from human normal gastric mucosa (○) and human gastric carcinoma (●) by anion-exchange on a DEAE-Sephacel column. GAGs from both normal and cancerous tissues are separated in three different populations: I, II and III. (Reprinted from ref. [97] Copyright © 2005, with permission from Elsevier.)

All known unsaturated disaccharides (non-, mono-, and di-sulfated), derived from CS, DS and HA were successfully resolved and detected using HPLC with fluorimetric detection after derivatization with dansylhydrazine [113]. The column used was a µ-Bondapack NH2 (aminopropylmethyilsilyl bonded amorphous silica) with isocratic elution in acetonitrile/100 mM acetate buffer (90/20), pH 5.6. The advantages of the method lie in good resolution, high sensitivity (50 pmol) and a short analysis time (50 min).

**Capillary electrophoresis (CE)**

CE has several advantages over a variety of other analytical methods: high sensitivity, high separation efficiency, on-line detection, simple operation, short analysis time, automated and reproducible analysis, and very low consumption of samples and buffers. The low quantities of samples processed make the method unsuited for preparative work.

The driving forces in CE are the electroosmotic flow (EOF) and electrophoresis. For GAG analysis CE in both normal (positive) and reversed (negative) polarity may be used. Normal polarity employs a basic or neutral buffer and requires sample application at the anode and detection at the cathode. Negatively charged species are prevented from migrating out of the capillary by the dominant EOF. In reversed polarity, an acidic buffer is used and the sample is applied at the cathode and detected at the anode. At pH < 3 the silanol residues of the internal wall of the capillary begin to lose their negative charges resulting in the reduction of the electroosmotic flow. Therefore the dominant force in the separation becomes electrophoresis and only anionic species migrate through the capillary towards the detector, making this method useful for the analysis of acidic carbohydrates [122]. Improved resolution is obtained in reversed polarity as compared with normal polarity. Sulfated GAGs develop a high electrophoretic mobility (EM) even at low pH, which may be reduced by adding ion-pairing reagents to the operating buffer.
CE has been successfully applied to the separation, characterization (in terms of the average number of sulfates per repeating unit), quantitation, and size determination of intact GAGs or oligosaccharides derived after enzymatic digestion. The possibility for size determinations are an advantage over the conventional chromatographic techniques.

The first application of CE to the analysis of disaccharides derived from CS, DS, and HA was simultaneously reported in 1991 by Al-Hakim and Linhardt [123] and Carney and Osborne [105], using normal polarity in a basic borate buffer. The method was later successfully applied to disaccharides derived from heparin and HS [124]. Under normal polarity, improvement in resolution was obtained by applying micellar electrokinetic capillary electrophoresis in basic borate buffer using cetyltrimethylammonium bromide as ion-pairing and micelle-forming agent [125]. Oligosaccharides from HA were well separated according to size by CE in a 0.1 M Tris-0.25 M borate buffer, pH 8.5, containing 10% polyethylene glycol (PEG) 70000, with detection at 200 nm [126].

Capillaries filled with polyacrylamide gel were used to separate CS-derived oligosaccharides produced by limited digestion of GAGs with different lyases [25] and these gels may prove effective in determining the degree of polymerization [127]. Separation of CS and DS was achieved by reversed-polarity CE with sensitive detection at 240 nm, based on the formation of a copper (II) complex in copper sulfate solution of pH 3.0 [106]. For the separation of non-, mono-, and disulfated disaccharides derived from CS/DS chains, the conditions were optimized during the last years by using pyrazolone derivatives [128] (in 100 mM borate buffer at pH 9.0) permitting an efficient and complete separation. Analysis of underivatized CS disaccharides from mink skin is possible using micellar electrokinetic capillary chromatography (MECC) in the presence of cetyltrimethylammonium bromide, pH 7.0, with direct UV detection at 232 nm [125]. CS and HS derived di- and oligosaccharides were fluorescent labeled with 2-AMAC, separated by CE and detected by UV at 254 nm. The sensitivity was limited, due to the lack of a fluorescent detector [110]. Digestion of DS with chondroitinases ABC, AC and B, results in unsaturated disaccharides having various sulfation positions and sizes. Using reversed polarity all of them can be separated in one run, using fused-silica capillary and phosphate buffer (pH 3.0). Detection at 232 nm offers sharp and symmetrical peaks and concentration as small as tens of pmol/l of each disaccharide can be reliably detected [129].

Various synthetic heparins and their fragments obtained by heparinases action were analyzed in phosphate buffer at pH 3.0 [101]. The method could be useful for assessing the purity of the low-molecular weight heparins. CE separation in a reversed polarity mode, using phosphate buffer at pH 3.5 can separate all twelve known isomeric disaccharides within 15 min [130].

CE with indirect UV detection overcomes the difficulties encountered with the analysis of carbohydrates that lack chromophores in their structures [127]. The principle of this detection consists in chromophore displacement by the analyte molecule, resulting in negative peaks. Buffers of benzene-1,2,4,-tricarboxylic acid and of 5-sulfosalicylic acid are generally used [131].

On-line CE-MS has been performed using different ionization techniques: electrospray, fast atom bombardment, and time-of-flight. The use of positive polarity CE with negative MS ionization seems to be the best approach to analyze complex oligosaccharide mixtures [132]. The applications of CE coupled to different MS detection for the analysis of biomolecules has been recently reviewed [133].

More details about capillary electrophoresis for the analysis of GAGs and GAG derived oligosaccharides may be found in several reviews [134,135], which have covered only this technique.
Fluorophore assisted carbohydrate electrophoresis (FACE)

A recently developed technique based on slab gel electrophoresis is fluorophore assisted carbohydrate electrophoresis or FACE. The technique was originally developed by Jackson and Williams [136,137] for the analysis of glycoprotein-derived carbohydrates labeled with 8-aminonapthalene-1,3,6 trisulphonate (ANTS) and further adapted to the analysis of GAG derived disaccharides and oligosaccharides. HA and CS from cartilage were digested with hyaluronidase/chondroitinase, labeled with AMAC [119,138] or ANTS [138] and subjected to FACE. This method permits separation of charged or uncharged mono- or oligosaccharides with high resolution and can detect single hydroxyl anomeric differences between mono- and oligosaccharides with otherwise identical molecular weight, charge and sequence. In addition to separating low molecular weight oligosaccharides, an HA polymer preparation with an average molecular weight of several tens of thousands was separated into each chain length component [138]. The sensitivity of the method is in the ng range and several samples can be analyzed in the same gel.

Spectroscopic techniques applied to GAGs

Fully characterization of GAGs requires detailed information about the fine structural composition of the different constitutional units of the polymer. Although chromatographic and electrophoretic techniques used for the separation/detection can provide a rough estimation of the structure/identity by comparison with known standards, more powerful techniques are necessary to fully sequence the compositional units. For this purpose Nuclear Magnetic Resonance (NMR) and mass spectrometry are the techniques that offer the best capabilities.

1H-NMR and 13C-NMR have been used for the characterization of oligosaccharides from chondroitin/dermatan sulfates [139,140] and heparan sulfate [141,142]. However, a major drawback of this technique is the high amount of saccharide required (around 5 mg) with enough purity.

Mass spectrometry has become in the last years the most important tool for GAG characterization. This is due mainly to the high sensitivity (low femtomol range) as well as the possibility of on-line coupling with separation techniques such as liquid chromatography [117,143] or capillary electrophoresis. The way mass spectrometry techniques have been applied to GAG analysis has evolved during the years, following the technical developments in the field. First attempts for MS analysis of oligosaccharides used fast atom bombardment (FAB) ionization, where the oligosaccharides were ionized as salt adducts. The method was able to detect up to 10 nmol octasaccharides derived from heparin [144,145]. However, considerable loss of SO3 took place, making difficult the determination of degree of sulfation.

It must be born in mind that the sulfate groups attached to GAG oligosaccharides are very labile constituting a big challenge for MS analysis. In order to prevent their loss mild and soft ionization techniques are preferred. MALDI-TOF mass spectrometry has been used to determine the sequences of purified GAG molecules [146]. However, sulfated GAGs ionize poorly when mixed directly with any of the matrixes commonly used for MALDI-TOF MS [147]. A strategy followed to circumvent this fact has been paring the oligosaccharide with a basic peptide, to form a neutral complex which ionizes readily allowing to determine the mass of the peptide-oligosaccharide complex [147]. The oligosaccharide sequence can be figured out indirectly by measuring the mass before and after partial enzymatic or chemical digestion of the complex. A clear limitation of this methodology is the impossibility to perform direct structural analysis. Although the oligosaccharide composition can be deduced from the molecular weight alone, the structural information is limited because stereoisomers and
positioned isomers cannot be distinguished. Moreover, the procedure is not quantitative. These restrictions together with the necessity to purify GAG oligosaccharides prior to MALDI-TOF make this technique better suited for a first evaluation of chain length distribution before applying other methodologies [148]. Undoubtedly electrospray (ESI) ionization is nowadays the most widely chosen approach for GAG analysis. Sulfated GAG oligosaccharides are easily detected using electrospray (ESI) mass spectrometry in the negative ion mode.

ESI-MS/MS was used to characterize KS oligosaccharide mixtures without prior purification or chromatographic separation [149]. Nevertheless, when using flow injection analysis extensive optimization of the MS parameters is required to distinguish between isomers. ESI spectra of sulfated GAGs are quite complex (when compared with other biomolecules such as peptides) due to several factors: a) high degree of heterogeneity of chain lengths in biological samples; b) formation of adducts between the sulfate groups and different cations to varying degrees; c) the molecules are fragile and may fragment during ionization loosing sulfate and d) a range of charge states? can be detected for each molecule [150]. Therefore there is a need for up-front separation of the oligosaccharides generated in enzymatic GAG digests before sequence assignment by MS/MS, in order to reduce the complexity of the resulting mixtures and to separate the isomers. For this purpose, CE and different types of liquid chromatography can be interfaced with MS. At first sight CE is the ideal separation technique for GAG derived sulfated saccharides due to their ionic character and polarity. Fractions separated by CE have been further characterized off-line by negative ESI/MS [151]. However, on-line CE/MS have been scarcely used in GAG analysis. Known limitations are the lack of sensitivity due to the inherent low loading capacity of CE and the loss of resolution in the interface with the mass spectrometer when using any sheath liquid. Useful application to real complex and low concentrated biological samples can only be achieved using sheathless approaches [152,153].

A SEC step prior to LC/MS analysis efficiently simplifies the composition of GAG entering the mass spectrometer [150,154], affording compositional mapping of oligosaccharides generated by hyaluronidase digestion of CS. Adding 10 mM HCl to the solvent eliminates sodium adduction and permits analysis of disaccharides produced by chondroitinase ABC digestion and determination of ratios of sulfation positions using multiple reaction monitoring.

Strong anion exchange LC although commonly used in fractionation of GAG of oligosaccharide mixtures is difficult to interface with MS and a good alternative has been found using ion-pairing reversed phase LC. This type of chromatography has been applied to the analysis of heparin oligosaccharides combined with ESI-MS. The necessity of utilizing volatile pairing agents has led to the use of alkyl ammonium acetates or dibutyl ammonium acetate [118], instead of tetraalkyl ammonium salts (which provides excellent chromatography). Even with this, the achievement of reasonable ionization of heparin was critical. Improvement of the method was obtained by introducing an on-line cation exchange step (using a Dionex ASRS-Ultra II Column) before the mass spectrometer [155], which removed the ammonium ions.

Easy and quantitative direct analysis of disaccharides released enzymatically from proteoglycans has been achieved in our group using porous graphitized carbon as reversed phase chromatographic stationary phase and on-line negative ESI/MS detection [143]. The high resolution power of this material allowed distinguishing between isomers differing in the sulfate position (4 or 6) utilizing a simple mobile phase composed of water/acetonitrile in 0.1% formic acid (see Fig. 6). A correct assignment of the sulfation position is obtained from the fragmentation spectra (MS²), based on the ratio of the fragment ions (Fig. 7). Graphitized carbon columns have been used also for the separation and characterization of larger oligosaccharides fragments derived from HS, heparin, HA and KS [156].
Proteoglycan purification and analysis

Fig. 6. LC-MS separation of positional isomers of standard CS disaccharides using a graphitized carbon column. (A) standard disaccharides sulfated in position 4; (B) standard disaccharides sulfated in position 6. (Reprinted from ref. [142] Copyright © 2005, with permission from Elsevier.)

Fig. 7. MS/MS spectra corresponding to each isomer of both disaccharides (sulfated in position 4 or 6) presented in Fig 6. The ratios of the different fragment ions depend on the position of the sulfate group, the most intense fragment ions being at m/z = 300.1 (sulfate in position 4) and m/z = 282.0 (sulfate in position 6). (Reprinted from ref. [142] Copyright © 2005, with permission from Elsevier.)

2.4 Conclusions and future perspectives

In the past decades the study of PGs has gained increasing importance due to their involvement in a variety of biological processes. This class of macromolecules has captured the attention of many researchers, which has triggered significant methodological developments in the analytical techniques employed for PG characterization. Both protein
cores and GAG chains are investigated in order to gain a deeper understanding of the structure-activity relationships as well as the biological role of a given PG.

The polyanionic nature of proteoglycans is the most important molecular property, which is taken into account when developing purification strategies. In addition, tissue localization, molecular size, and solubility of a given PG are important features when developing a purification strategy. Although many techniques are at the researcher’s disposal, the most generalized approach involves PG extraction with a chaotropic agent (most commonly 4 M GuHCl) followed by anion exchange and size exclusion chromatography. Nondenaturing buffers are required when the interest lies in studying structure-activity relationships in more detail and when preserving the biological function of the isolated PG is required. Advances in the chemistry of chromatographic supports have enabled the application of HPLC for the isolation and fractionation of PGs with an increased resolving power. Reversed phase chromatography and affinity chromatography using immobilized lectins result in high purity PGs when used as the last purification steps.

An increasing number of derivatizing agents facilitate more sensitive UV or fluorescent detection of GAG derived oligosaccharides in chromatographic and electrophoretic techniques.

Characterization of oligosaccharides derived enzymatically from different GAGs is an important step toward the complete understanding of the relation between their chemical structure and biological functions. During the past decade, MS has become the key methodology in the structural elucidation of GAGs. Technical developments in MS, especially the soft ionization of thermolabile biomolecules, have allowed studying complex sulfated oligosaccharides, as for example derived from highly sulfated heparins. Based on fragmentation patterns, detailed structural information about the number and position of sulfate groups can be determined and thus correlated to biological activity.

Methodological advances continue to drive the field of PG purification and characterization. For example, the development of monolithic chromatography materials with freely accessible throughpores, overcomes some of the limitations associated with packed columns when it comes to throughput and diffusion limited mass transfer. Monolithic columns offer high efficiency, fast separation, and high binding capacity at a lower pressure compared to conventional HPLC resins. For all these considerations we think that monolithic columns may find more widespread applicability for the fractionation of intact proteoglycans.

Implementation of microfluidic LC-MS systems (such as the recently introduced chip LC-MS system) operated at flow rates of some hundred nanoliters per minute promises to increase sensitivity further and will likely also enter the area of glycosaminoglycan analysis. This will permit analysis of biological samples where previously sample amount and concentration were the limiting factors.
References
