Published in final edited form as:

Mol Biosyst. 2012 June ; 8(6): 1613–1625. doi:10.1039/c2mb25021g.

Proteoglycan sequence

Lingyun Li^a, Mellisa Ly^a, and Robert J. Linhardt^{a,b}

Robert J. Linhardt: linhar@rpi.edu

^aDepartment of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, 12180, USA; Fax: +1 518-276-3405; Tel: +1 518-276-3404

^bDepartment of Biology, Chemical and Biological Engineering and Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, 12180, USA

Abstract

Proteoglycans (PGs) are among the most structurally complex biomacromolecules in nature. They are present in all animal cells and frequently exert their critical biological functions through interactions with protein ligands and receptors. PGs are comprised of a core protein to which one or multiple, heterogeneous, and polydisperse glycosaminoglycan (GAG) chains are attached. Proteins, including the protein core of PGs, are now routinely sequenced either directly using proteomics or indirectly using molecular biology through their encoding DNA. The sequencing of the GAG component of PGs poses a considerably more difficult challenge because of the relatively underdeveloped state of glycomics and because the control of their biosynthesis in the endoplasmic reticulum and the Golgi is poorly understood and not believed to be template driven. Recently, the GAG chain of the simplest PG has been suggested to have a defined sequence based on its top-down Fourier transform mass spectral sequencing. This review examines the advances made over the past decade in the sequencing of GAG chains and the challenges the field face in sequencing complex PGs having critical biological functions in developmental biology and pathogenesis.

Introduction

Proteoglycans (PGs) are complex glycan–protein conjugates, typically containing various core proteins that are post-translationally modified with one or more glycosaminoglycan (GAG) chains. GAGs are linear, anionic polysaccharides comprised of repeating disaccharide building blocks. The disaccharides can be an amino sugar (*N*-acetyl-D-glucosamine (GlcNAc) or *N*-acetyl-D-galactosamine (GalNAc)) and an uronic acid (D-glucuronic acid (GlcA) or L-iduronic acid (IdoA)) or D-galactose (Gal) that is often modified at different positions with sulfo group (S) substitution. There are four groups of GAGs (Fig. 1A): (1) hyaluronan (HA) is a long polysaccharide with a single non-sulfated disaccharide-repeating unit and no core protein; (2) keratan sulfate (KS) uniquely has a Gal instead of an uronic acid residue and KS is both *O*- and *N*-glycosidically linked to core proteins; (3) chondroitin sulfate (CS, including dermatan sulfate (DS)), in which GlcA has been epimerized to IdoA is *O*-glycosidically linked to various core proteins; and (4) heparan sulfate (HS, including heparin, which has a higher sulfation/disaccharide ratio and a higher IdoA content than HS), the most structurally complex GAG, is *O*-glycosidically linked to various core proteins. The three types of GAGs found in PGs (KS, CS and HS) are

synthesized on a linkage region attached to a core protein through the action of glycosyltransferases and subsequent modification by epimerases and sulfotransferases.^{1–3} PGs are thought to be the most structurally complex biopolymers having GAG chains of different types with variable sequence, length, and occupancy as well as other posttranslational modifications (*i.e.* phosphorylation, *N*-terminal methylation, sulfation and *N*-myristoylation, *etc.*). The diverse structures of PGs make these the most information dense molecules found in nature.

PGs are found on the surface of all animal cells, in intracellular granules of selected cell types, in the basement membranes of various tissues, and in the extracellular matrix (ECM). The GAG chains of PGs are involved in many biological processes including cell–cell and cell–matrix interactions,^{2,4} cell migration and proliferation,^{4,5} growth factor sequestration,² chemokine and cytokine activation, microbial recognition⁶ and tissue morphogenesis during embryonic development.^{2,5} It has been increasingly recognized that specific sequences or motifs with GAGs might be responsible for the regulation of biological activity, however, in only very few cases is there a well-defined structure–activity relationship (SAR). The prototypical example is the heparin antithrombin III (AT) binding pentasaccharide sequence (Fig. 1B), which is critical for heparin's anticoagulant activity.⁷ The discovery of this five sugar sequence responsible for the binding and functional modulation of AT has led researchers to make substantial efforts focused on identifying and specifying the sequences for other GAG–protein interactions.

Although significant advances in the analytical methodology have made the sequencing of DNA and protein a routine task, the sequencing of PG-derived GAG chains still remains very challenging. The tremendous structural heterogeneity of GAGs and the absence of methods for their purification and/or amplification have been major impediments to GAG chain sequencing. Current technology is still unavailable to determine the GAG chain sequence of most types of PGs. The majority of the publications on GAG sequencing describe the structural determination of small GAG-derived oligosaccharides, typically not much larger than deca-saccharides.^{8–12} Successful sequencing of such GAG-derived oligosaccharides has generally been accomplished using techniques ranging from enzymatic degradation followed by polyacrylamide gel electrophoresis (PAGE) or high performance liquid chromatography (HPLC), to mass spectrometry (MS) and multi-dimensional nuclear magnetic resonance (NMR) spectroscopy.^{13–17} A recent break-through study, using a top-down glycomics approach combining PAGE and high-resolution MS, has allowed the sequencing of the structurally simple, short (27–39 saccharide residues) CS-GAG chains of the bikunin PG.¹⁸ The results demonstrate, for the first time, that a heterogeneous PG, bikunin, has a defined sequence, which suggests that even more structurally complex PGs may also have defined sequences.

Over the past decade researchers have developed many approaches for sequencing GAGs, particularly HS and CS, from a variety of different cell and tissue sources. This review briefly introduces the classes of PGs, their preparation, SAR, and biosynthesis before detailed discussion of whether their GAG chains have unique and definable sequences and the different strategies employed for their sequencing. Finally, the future technological and scientific challenges in the area of PG and GAG sequencing are discussed.

PG classification

The PG superfamily is subdivided into three major groups based on the type of the disaccharide building blocks, HS-PGs, CS-PGs and KS-PGs (Fig. 1, 2 and Table 1).¹⁹ Due to their high content of sulfo groups and uronic acids in HS-PGs and CS-PGs, all three groups of PG are strongly negatively charged. HS-PGs, the most structurally diverse type of

PGs found in animal cells, are *O*-sulfo group-substituted, linear chains with 10 to 200 disaccharide units of (GlcA β or IdoA α) 1–4 (GlcNAc/GlcNS α) 1–4, attached through a tetrasaccharide linkage region, GlcA β 1–3 Gal β 1–3 Gal β 1–4 D-xylose (Xyl) β 1-, to different core proteins. Syndecans 1–4, glypicans 1–6, β -glycan, perlecan, agrin, serglycin, and collagen type XVIII are all HS-PGs and contain 1 to 15 *O*-linked HS chains and have a molecular weight ranging from 10 to 400 kDa.^{2,20} CS-PGs are characterized by *O*-sulfo group-substituted, linear chains with 20 to 100 repeating disaccharide units of (GlcA β or IdoA α) 1–3 GalNAc β 1–4, attached through a tetrasaccharide linkage region to various core proteins. The most abundant CS-PGs include aggrecan, versican, neurocan, brevican, decorin and biglycan.²¹ The largest CS-PG, aggrecan, the principal component of cartilage, has an average of 100 CS chains and a protein core of 208–220 kD.¹ KS-PGs GAG chains are characterized *O*-sulfo group-substituted chains formed by repeating disaccharide units of Gal β 1–4 GlcNAc β 1–3 to various core proteins.²² KS-PGs can be classified as either corneal type (KS I, GAG chain *N*-glycosidically linked to an asparagine residue) or skeletal type (KS II, GAG chain *O*-glycosidically linked to serine of a threonine residue). In KS I, the linkage to Asn is a typical biantennary branched, L-fucose (Fuc)-bisected, pentasaccharide with one antenna carrying a KS chain in which most GlcNAc residues and half of the Gal residues are substituted 6-*O*-sulfo groups. The non-reducing end of the KS I is typically capped by a variety of carbohydrate structures, such as neuraminic acid (Neu5Ac), Gal, and GalNAc6S. KS II is characterized by monosaccharide linkage to GalNAc *O*-linked to Ser or Thr residues in the core protein. The chains are shorter and sulfated to a higher extent than in KS I.⁶

Preparation and limited availability of PGs

PGs are ubiquitous biomacromolecules found throughout the body of all animals and located intracellularly, on the cell surface and in the ECM. The molecular weight of PGs ranges from tens to millions of kDa. PGs can have one or many highly negatively charged GAG chains attached to their core protein (Table 1). PGs need to be properly and efficiently isolated and purified from cultured cells or tissues to study their structure and biological function. The purification of PGs is often complicated by their large molecular size and charge and necessitates the use of chaotropes or detergents required for their efficient extraction. Several detailed protocols for the preparation of GAGs and PGs for structure analysis have been recently published.^{15,24,25} Slicing, grinding or pulverization of tissue samples is generally necessary for efficient PG extraction. After extraction, GuHCl and other salts present in the extraction buffer are generally removed by buffer exchange, prior to PG or GAG recovery from the extraction buffer using weak (W) anion-exchange (AX) or strong anion-exchange (SAX) chromatography on a column for large volume samples or a spin cartridge for small volume samples. PGs present in biological fluids or secreted into cell culture media can be easily enriched by buffer exchange with CHAPS–urea binding to an SAX spin cartridge, washing with water and elution with 2 M NaCl.²⁶ PGs are often further purified by other standard methods including: size exclusion chromatography (SEC) or ultracentrifugation or membrane filtration; hydrophilic interaction chromatography (HILIC); reversed-phase chromatography (RPC); and antibody affinity chromatography.²⁷

Pure PGs, rarely available in greater than microgram quantities, are often purified in research laboratories at considerable effort for biological studies. Purification from tissues, with few exceptions, is difficult and unsustainable for commercial production. While heparin and CS GAGs are under metric ton-scale production from animal tissues, bikunin is the only PG commercially produced in kilogram quantities from human urine and is used as a pharmaceutical in Japan for the treatment of acute pancreatitis.²⁸ Routine bacterial, yeast and insect cell expression systems cannot be used for producing recombinant PGs because they require complex post-translational modification. Expression in eukaryotic cells is more

complicated, yields considerably smaller amounts of recombinant PGs, and cannot yet permit control of GAG chain modification. Very few recombinant PGs are currently commercially prepared in gram quantities in cell culture, one example is the decorin PG for use in regenerative medicine.²⁹ Thus, the sequencing of PGs has been severely limited by the lack of targets in sufficient amounts and purity.

Structure and activities of PGs

GAG chains can contain 10–200 disaccharide units, which are incredibly, structurally diversified through their biosynthetic pathway. Studies clearly show that GAG structural diversity is strictly controlled and regulated in biological systems, enabling PGs to selectively interact with proteins in a spatially and temporally controlled manner. The activities of PGs cover a wide range of biological processes, including cell proliferation and differentiation,^{2,3,5,6} cell adhesion, transport regulation,^{2,3} blood coagulation,^{2,3,30} angiogenesis,^{32,33} tumor metastasis³² and pathogen invasion.^{19,20,33} Hundreds of PG-binding protein ligands have been identified and the list of GAG-binding proteins, the PG–GAG interactome,^{4,20,32,34} continues to grow. Generally, binding is dominated by ionic interactions between the GAG's negatively charged sulfo/carboxyl groups and basic amino acid residues on the surface of the GAG-binding protein. Non-ionic interactions such as hydrogen bonding and van der Waals interactions can contribute and even dominate GAG–protein binding.³⁵ There are also common linear and topological motifs of basic and hydrophobic amino acid residues present on the surface of proteins that can be predicted to bind GAGs.⁴

The earliest evidence for specificity in GAG-binding to a protein was the establishment of a unique AT-binding sequence (Fig. 1B).³⁶ On binding to heparin AT undergoes a conformational change that accelerates its rate of inhibition of blood serine proteases resulting in much of heparin's anticoagulant activity.³⁰ Extensive SAR studies suggest that while AT–heparin binding is highly specific, there is some allowable variability in chirality and sulfo group substitution in this site.³¹ More active study regarding specificity of GAG structure binding to protein has been conducted under the inspiration of the pentasaccharide specific binding to antithrombin. Many additional binding studies have been undertaken on other GAG-binding proteins.^{32,37} These have generally shown somewhat more limited specificity than observed for the AT–heparin interaction. Often a particular type of sulfo group (*e.g.* 6-*O*-sulfo, or 2-*O*-sulfo, or *N*-sulfo, *etc.*) or a domain structure (*e.g.* highly substituted with sulfo groups in a high sulfo domain) seems to critically contribute to an interaction. To date, there has been no clear evidence that the distinctive sequence specificity observed in the AT–heparin interaction can be generalized to other GAG–protein interactions.^{38,39}

A well-studied example of a GAG–protein interaction with critical biological ramifications involves signaling by the 23-member fibroblast growth factor (FGF) family through their 7 receptor-subtypes (FGFRs). FGFs modulate cellular proliferation and differentiation through a signaling pathway linked to cancer progression and spreading. This signaling modulation requires the formation of a FGF·FGFR·HS-PG complex on the cell surface.⁴⁰ Crystal structure and biochemical experiments suggest a FGF₂·FGFR₂·(HS-PG)₂ complex,⁴¹ with each of the many possible FGF₂·FGFR₂ protein complexes forming an unique, basic, binding site canyon, complementary to HS-PGs having different sequences (<http://www.rcsb.org/pdb/explore.do?structureId=1fq9>). The identification of the HS GAG sequences fitting these binding site canyons is just beginning^{42,43} but offers a rich area of investigation profound in developing a molecular understanding of signal transduction.

In vitro studies on binding thermodynamics (K_d), kinetics (k_{on} , k_{off}) and specificity give an initial assessment of a GAG–protein interaction. Ultimately, *in vivo* assessment of biological activity is critical for developing a complete understanding of protein–GAG interactions. Despite an intensive study of GAG–protein interactions much more research is needed to establish the full SAR of PGs. There still remains a debate about whether specific sequences or overall sulfation density is the primary requirement for PGs binding to ligand proteins to exhibit their biological activities.

Biosynthesis of proteoglycans

If a specific oligosaccharide sequence has a selective biological response, then how does a cell create such a sequence? Multiple enzymes, located in the endoplasmic reticulum (ER) and Golgi, catalyze the assembly of a PG (Fig. 2 and 3). The protein core of PGs is constructed in ribosomes and then translocated to the rough ER to start the GAG chain biosynthesis including initiation, polymerization and modification. In ER, xylosyl transferase initiates the synthesis of the linker tetrasaccharides through adding a Xyl to a specific serine residue of the protein core. Two Gal residues and one GlcA are added in sequence to the Xyl, through the action of galactosyltransferase I, galactosyltransferase II, and glucuronyltransferase I, in Golgi to form the tetrasaccharide linker (GlcA β 1–3Gal β 1–3Gal β 1–4Xyl β 1–O–Ser) (Fig. 3). The fifth saccharide addition determines whether the GAG chain becomes CS/DS or HS/heparin. Adding a GalNAc results in subsequent CS/DS chain elongation while adding a GlcNAc results in HS/heparin chain elongation. In HS/heparin biosynthesis, following the attachment of the first GlcNAc residue, the polymer elongation proceeds through the sequential addition of GlcA and GlcNAc by copolymerases (EXT-1 and EXT-2) to form the heparosan backbone. During this polymerization, chain modifications take place, including *N*-deacetylation/*N*-sulfation, C₅-epimerization and *O*-sulfation, involving *N*-deacetylase/*N*-sulfotransferase (NDST 1 and 2), epimerase and several *O*-sulfotransferases (2-OST, 3-OST 1–7, and 6-OST 1 and 2). There are similar parallel pathways for CS/DS and KS biosynthesis.

PG biosynthesis is very complex involving multiple enzymatic reactions. While most of these biosynthetic enzymes have been cloned and expressed, little is understood about their coordinated action and colocalization. Therefore, the speculation about whether there is a glycan code, *i.e.*, a predetermined sequence of GAG chains, remains difficult to determine because our limited understanding of the control mechanisms in biosynthesis. PG biosynthesis is spatially (tissue) and temporally (developmentally) restricted and is also responsive to environmental signals.

Do PGs have unique and definable sequences

The diverse sulfation pattern of a PG is introduced through extensive modification of the carbohydrate backbone during the biosynthesis. Two major sulfation motifs are found in HS-PGs and CS-PGs, corresponding to high sulfation or low sulfation domains. Evidence suggests that the fine structure within these domains as well as their size and placement of the GAG of PGs is critical for their functions *in vivo*.⁵ There is an open debate over whether PG glycans have a predictable or deterministic sequence.^{18,19,44} While the biosynthetic mechanism, resulting in the introduction of sequence into nucleic acids and proteins, is well understood and involves template driven biosynthesis, there is no comparable understanding of how sequence might be introduced into glycans within the ER and Golgi. Although enzymes can recognize remote sequence, imparting domains, the organized biosynthesis of the multiple domains within a full GAG chain may be beyond enzyme specificity. Despite all that is known about glycan biosynthesis, our understanding is still insufficient to infer a sequence or even to suggest that glycan possesses a sequence.^{18,19,32} Indeed, it is valid to

even question why a glycan would need a definable sequence. GAGs, however, play a complex and even a dominant role in biology. A molecular-level understanding of glycosaminoglycans has begun to emerge, which suggests the importance of fine structure in controlling the biological properties of glycosaminoglycans.

Chemical and enzymatic degradation to determine GAG structure

PG characterization generally utilizes a bottom-up approach due to the high degree of GAG structural complexity, heterogeneity and polydispersity. Intact polysaccharide chains are chemically or enzymatically depolymerized into smaller oligosaccharides that are then structurally characterized. Chemical degradation typically relies on oxidative deamination by nitrous acid at different pH values.⁴⁵ At pH 1.5 cleavage occurs at GlcNS residues, whereas if the pH $\approx$ 4.0, cleavage occurs at *N*-unsubstituted GlcN. To cleave at GlcNAc (or GalNAc), samples first need to be de-*N*-acetylated with base or hydrazine treatment followed by iodine oxidation before treatment with pH 4.0 nitrous acid. Deamination cleavage alters the structure of the GlcN, thereby producing a 2,5-anhydro-D-mannose residue at the reducing end of the cleaved oligosaccharide, which is then reduced, often with sodium borotritide, to 2,5-anhydro-D-mannitol.⁴⁶ An advantage of deaminative cleavage is that the process preserves the epimerization state of the uronic acid residue. Disadvantages of nitrous acid degradation are that it is difficult to control and it often requires the use of radioisotopes for detectable products.

PG-GAG chains degraded by specific enzymes afford GAG-derived oligosaccharides for structure analysis. Enzymes for degrading all types of GAGs are commercially available. There are two classes of GAG depolymerizing enzymes: prokaryotic lyases and eukaryotic glycosidases.⁴⁷ The cleavage of GAGs affords oligosaccharides with a new free reducing end, which enables the attachment of fluorescent tag for sensitive detection after separation. Among the most commonly used lyases are the HA lyases, CS lyases and HS lyases.⁴⁸ These enzymes β -eliminatively cleave the hexoamine (1 $\rightarrow$ 4) uronic acid bond affording disaccharides and higher oligosaccharides with a C4-5 double bond in uronic acid, which absorbs strongly at 232 nm. Lyases are particularly useful as they are primarily endolytic enzymes, however, information about epimerization state of the uronic acid is lost during enzymatic degradation. Hydrolases, including heparanase, hyaluronidase, keratanase, break GAGs by adding water across their glycosidic linkages. Lysosomal enzymes are primarily exolytic hydrolases (*i.e.* iduronidase, glucuronidase, hexosaminidase, *etc.*) with low pH optima that sequentially catabolize GAGs releasing one saccharide residue at a time from the non-reducing terminus.⁵⁸ Other lysosomal enzymes remove the sulfo groups from GAGs including sulfamidase, glucosamine-3- or 6-sulfatases, galactosamine-4- or 6-sulfatases and glucuronate-2- or iduronate-2-sulfatases. The most common enzymes employed for GAG degradation and their specificities are shown in Table 2.

Combined with high-resolution PAGE, capillary electrophoresis (CE) or HPLC, these endoenzymes and exoenzymes can be used to sequence GAGs and GAG-oligosaccharides. Unfortunately, not all the required enzymes are currently available to determine the full GAG sequence,⁵⁹ many of the available enzymes do not have strict and well-characterized specificity, many enzymes are endolytic and are difficult to use in direct sequencing,⁶⁰ many of the exolytic lyases are not strictly exolytic and can jump over resistant sites, and many of the endolytic enzymes are not purely endolytic, acting preferentially at select sites. Those shortcomings limit the application of an enzymatic approach for direct sequencing of large oligosaccharides.

Gel-based sequencing

Determining GAG structure is a formidable analytical challenge because of their structural complexity, high negative charge, polydispersity and microheterogeneity. Attempts have been made to directly sequence mixtures of intact GAG chains, released from a core protein, using their fluorescently labeled reducing ends as a reading frame. In a top-down approach, partial digestion of an intact, fluorescently end-labeled GAG chain with an endolytic enzyme, similar to the method originally used by Maxam and Gilbert to first sequence DNA,⁶¹ affords a banding pattern on PAGE analysis, from which a short sequence of a GAGs reducing terminal domain can be determined based on known enzyme specificity.⁶² The major limitations of this top down approach are: (1) the precise specificity and action patterns of the lyases are not well established and many of these enzymes cleave at multiple glycosidic linkages; (2) there are insufficient GAG lyases currently available for specific cleavage at all the different glycosidic linkages present in a GAG chain; and (3) the resolution of the PAGE separation is often insufficient the entire size range of the GAG being sequenced.

Middle-down approaches of GAG sequencing commonly involve multiple steps, including partial chemical or enzymatic depolymerization of GAG chains, followed by separation and purification (SEC, SAX-HPLC, PAGE, CE, *etc.*) to obtain a pure oligosaccharide for sequence analysis. Sequential digestion with exoglycosidases, in conjunction with radioactive or fluorescent labeling is then combined with a high-resolution separation^{10,11,63} to sequence the GAG-derived oligosaccharide. Discontinuous PAGE has provided a rapid, high-resolution parallel separation for oligosaccharides since the 1990s.¹⁰ The availability of recombinant, exolytic, lysosomal enzymes that degrade a reducing-end labeled GAG-derived oligosaccharide through the sequential removal of saccharide residues and sulfo groups from its non-reducing end has facilitated sequence analysis. Integral glycan sequencing (IGS),^{9,13} capable of sequencing small amounts (picomoles) of pure HS-derived oligosaccharides, works by fluorescently labeling a pure oligosaccharide at its reducing end, partially cleaving with nitrous acid at low pH, and sequentially processing with various exoenzymes. Based on the enzyme's specificity, researchers can directly read out the oligosaccharide sequence based on the fluorescent banding patterns (Fig. 4A–D). IGS has been successfully applied to sequencing HS polysaccharides that act as specified regulators of FGF signaling.^{8,9}

A similar approach demonstrated on a purified dermatan sulfated-derived oligosaccharide introduces label into the released monosaccharide after exoenzyme treatment and analyzing it using fluorophore assisted carbohydrate electrophoresis (FACE) (Fig. 4D).⁶⁴ The limitations of these middle-down approaches include: (1) no information is provided on the context of placement within the intact GAG chain of the oligosaccharide being sequenced; (2) the maximum size of GAG-derived oligosaccharide that can be purified is usually <octadecasaccharide; and (3) the quality of the sequencing data are dependent on the high specificity of the exoenzymes utilized and the resolution of the PAGE separation through the entire size range of the sequencing experiment.

Nuclear magnetic resonance (NMR) spectroscopy for sequence determination

NMR spectroscopy is sensitive to structural changes making it an important technique for GAG sequence analysis. NMR is typically used for characterization of a single GAG-derived oligosaccharide with >80–90% purity. Because of the low signal dispersion of GAG samples, a high-field strength instrument (>500 MHz) is generally required for structural determination. A ¹H-NMR spectrum is often initially obtained to reveal the number of

monosaccharide residues present and provide a tentative compositional analysis.⁶⁵ ¹H-NMR spectroscopy currently remains one of the best ways of distinguishing IdoA and GlcA residues and assigning anomeric configurations. ¹³C-NMR spectroscopy is particularly important in unambiguously determining which sites are substituted with *N*-sulfo and *O*-sulfo groups along a GAG chain. While less sensitive than ¹H-NMR, ¹³C-NMR offers more dispersion of chemical shift resulting in less signal overlap. In samples in which ¹³C incorporation is possible ¹³C-NMR can also offer a sensitive alternative to ¹H-NMR.

The complete structure characterization of oligosaccharides usually requires multi-dimensional NMR spectroscopy.^{66,67} This is accomplished by a combination of two-dimensional NMR techniques such as correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) for ¹H, which allows assignment of the ¹H signals of individual monosaccharide residues. After this, the heteronuclear single-quantum coherence (HSQC) experiment is often used to extend the proton assignment to the ¹³C spectrum or to the ¹⁵N spectrum.^{68,69} Another key experiment for sequencing is the two-dimensional heteronuclear multiple-bond correlation (HMBC) experiment, which detects a coupling between the anomeric proton and the carbon atom on the opposite side of the glycosidic linkage.

Although NMR experiments provide unique molecular-level information required for structure analysis, such measurements are often limited by the experiments' relatively low sensitivity. This is of particular concern for biological samples with limited quantity. Several approaches have been used to improve NMR sensitivity to address this limitation, such as increasing the strength of the static magnetic field to 800 MHz or even 940 MHz. Cryogenically cooled NMR probes and receivers can significantly improve the signal-to-noise ratio (S/N) by 2–4 fold.^{70,71} For compounds with highly soluble samples, such as GAGs, microcoil NMR probes can provide additional sensitivity enhancement. Microcoil NMR also allows easy coupling to separation methods such as HPLC, CE and capillary isotachopheresis (cITP), which has been useful for obtaining ¹H NMR spectra of 1–2 µg of heparin-derived disaccharides and tetrasaccharides.⁷² NMR spectroscopy is still a powerful tool for *de novo* structural characterization of unknown GAG-derived oligosaccharides up to octadecasaccharides, it can make use of stable NMR-sensitive isotopes (*i.e.* ¹³C, ¹⁵N, ³³S) and is nondestructive allowing the same sample can later be used for other destructive analytical methods. Moreover, NMR can also provide information on the conformation and secondary structure of GAGs and their complexes with proteins.⁷³ There are a number of limitations on the use of NMR spectroscopy, such as spectrometer cost, the high-level expertise and time required for interpreting data, resulting in low sample throughput.

X-Ray crystallography for sequence determination

Even though the essential role of protein–GAG interactions in regulation of various physiological processes has been recognized for decades, there are still very limited data available for elucidating the three-dimensional features of GAGs as they interact with their protein ligands. It has not yet been possible to crystallize and solve the X-ray structure of an unbound GAG or GAG-derived oligosaccharides. The limitation is due to the specially challenging GAG heterogeneity and conformational flexibility, the result of which is that GAGs or GAG-derived oligosaccharides do not have a single, well-defined three-dimensional structure in solution. It is, however, possible to lock a GAG-derived oligosaccharide into a single conformation by binding it to its protein ligand.

The first crystal structures published were FGF2–heparin tetrasaccharide and hexasaccharide structures.^{74,75} Currently, more than 60 crystal structures of complexes between protein and glycosaminoglycans are available in the Protein Data bank.⁷⁶ Solutions to 1 Å resolution are possible. Some of these solved structures are quite complex, like the six-component

FGF₂-FGFR₂·(heparin octasaccharide)₂ complex.⁷⁷ Despite the early discovery of the binary AT-pentasaccharide complex, it took many years of refining crystal growth and the use of a synchrotron to obtain the high-resolution 3-D structure of this complex.³⁰ X-Ray crystallography remains limited as a sequencing tool for a number of reasons: (1) GAGs and GAG-derived oligosaccharides have not been crystallized on their own; (2) protein GAG-derived oligosaccharide complexes are difficult to obtain and require multi-milligram amounts of pure GAG-derived oligosaccharide; and (3) crystallography of GAG-protein complexes is a low throughput method, requires a synchrotron, expert experimentalists and structures take a long time to solve and refine.

Mass spectroscopy for sequence determination

MS can provide structural information on elemental composition of intact GAG from molecular ions as well as linkage and sequence information from fragment ions. With advances in spectrometers and techniques for glycan analysis, MS analysis is rapidly becoming the most important tool for GAG characterization and sequence analysis.^{18,19,59,78–83} Combinations of online or offline separations (PAGE, CE, and HPLC) with high-resolution, high-sensitivity MS and tandem MS (MS/MS), small amounts of GAGs present in complex biological samples can now be analyzed to obtain important structural information.^{84,85} Two types of MS soft ionization techniques, matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), are widely used in PG and GAG sequence analysis. Fast atom bombardment (FAB)-MS, the first soft ionization method to play an important role in GAG analysis,⁸⁶ has fallen into disuse because of its complexity and has been largely supplanted by ESI-MS and MALDI-MS. Both MALDI and ESI technologies permit the direct ionization of nonvolatile biomolecules and the transfer of ions into different types of MS and MS/MS analyzers. MS analysis represents an efficient approach for the analysis of many of the structural features present in GAGs and PGs including: high molecular weight and chain length, polydispersity, structural heterogeneity, different types of glycosylation, different glycosidic linkages,⁸⁷ patterns of sulfo group substitution and the presence of epimers within the GAG chain.⁸⁸ It has recently been possible to even obtain complete sequences of individual intact GAG chains.¹⁸

In MALDI-MS experiments, the sample is dried on a metal plate in the presence of a chromophoric matrix until matrix crystals containing trapped sample molecules are formed. Ionization of the sample is effected by energy transfer from matrix molecules that have absorbed energy from laser pulses. Typically, an UV laser is used in most MALDI-MS instruments. MALDI sources are usually attached to time of flight (TOF) mass analyzers that can handle biomolecules having a very wide mass range (from tens of Da to hundreds of kDa) with high throughput. MALDI-TOF-MS has been used for sequencing small heparin-derived oligosaccharides providing fast and accurate mass determination.¹⁴ MALDI-MS based sequencing has several advantages over gel based sequencing (Fig. 4), it does not require tagging, oligosaccharide repurification after the exoenzymes treatment, and while it can be used in conjunction with enzymes, it does not require an analyst to rely on enzyme specificity to deduce a sequence. Direct analysis of large GAG-derived oligosaccharides by MALDI, however, is challenging due to their highly negative charge, low ionization efficiency, and the loss of labile sulfo groups during ionization. Combination of basic peptides such as (Arg-Gly)_n can efficiently improve the ionization and reduce the sulfate loss.⁸⁹ Today, various MALDI matrices were tested to improve the quality of MALDI spectra for GAG analysis. Ionic liquid matrices (ILMs) have a number of advantages over conventional crystalline matrices. It improves the homogeneity of the GAGs-matrix mixture, stability of matrix under high vacuum, and reproducibility from shot-to-shot.⁹⁰ GAG-oligosaccharides have been successfully using 1-methylimidazolium-*R*-cyano-4-hydroxycinnamate and butyl-ammonium 2,5-dihydroxybenzoate although significant sulfate

loss takes place during the ionization process.⁹¹ The ILM, bis-1,1,3,3-tetramethylguanidinium-*R*-cyano-4-hydroxycinnamate, was useful for positive-mode MALDI-TOF-MS analysis of CS-disaccharides at reasonable sensitivities with efficient suppression of sulfate loss.⁹²

ESI is a soft ionization process that generates multiply-charged ions directly from a stream of liquid allowing its convenient interface with online separation methods.⁹³ ESI-MS experiments are often performed using instruments with quadrupole, ion trap and TOF analyzers. Fourier transform (FT)-MS with electron capture dissociation (ECD) in the positive-ion mode or electron detachment dissociation (EDD) in the negative-ion mode and low-cost ion traps with electron transfer dissociation (ETD) represent cutting-edge technologies that can considerably improve the MS analyses of GAGs. In contrast to MALDI, ESI offers the advantages for coupling to microbore or nanobore liquid chromatography (LC) permitting on-line LC/ESI-MS analysis which have been well reviewed by Zaia *et al.* recently.^{84,85} This methodology is particularly useful when a very limited amount of biological mixtures of oligosaccharides needs to be analyzed.

MS direct analysis of intact PGs is an incredibly challenging task because GAGs possess low ionization efficiency, much size heterogeneity and thermal lability.^{79,94} The significant differences in ionization characteristics between GAG and protein need to be carefully considered when attempting to sequence PGs. GAGs are highly negatively charged with tremendous heterogeneity. The preferred method for analysis, after a high-resolution separation, is in the negative-ionization mode. While core proteins are generally sequenced in the positive-ionization mode using traditional, well-established proteomics approaches. When the protein component of a PG is the object of study, GAGs are enzymatically (by GAG lyases or hydrolases (Table 2) or PNGase F for N-linked KS I) or chemically (*i.e.* β -elimination or hydrazinolysis) removed. The core protein is then separated by PAGE or 2D-PAGE, and the resulting band or spot is subjected to the in-gel digestion followed by MS and/or MS/MS analysis for peptide sequencing and protein identification. Linkage region information can be obtained through analysis of the MS and MS/MS of the glycopeptides that originally contained the GAG-chain.^{18,94} In another approach, the PGs mixture can be proteolyzed either by specific or nonspecific enzyme and the resulting peptidoglycosaminoglycans (pGs) can then be separated from unmodified peptides by anion-exchange chromatography and/or size-exclusion chromatography for the LC-MS/MS proteomics analysis.⁹⁵

The MS analysis of intact GAGs released from PGs remains incredibly difficult. However, domain sequencing of HS and CS/DS GAGs is a widely-applied strategy for GAG characterization.^{59,84,96} Novel MS/MS technique development focuses mainly on FT-ICR MS instrumentation in negative-ion mode for GAGs sequence analysis. Various MS fragmentation techniques (*i.e.* collisionally induced dissociation (CID), high energy collisionally induced dissociation (HCD), ETD, ECD, and EDD) have been used to obtain information rich fragmentation, especially cross-ring fragmentation.^{83,88,97–99} The stereochemistry of the uronic acid in CS and HS epimers can be distinguished by MS/MS.^{87,100,101} Fragmentation by CID has been extensively studied on disaccharides since they are commercially available as pure standards.¹⁰⁰ A heparin/HS oligosaccharide sequencing tool, called HOST, has been developed to help heparin sequencing using enzymatic digestion and ESI-MSⁿ approach. The HOST program helps in automating the interpretation of the MSⁿ data generated from GAGs, provides a practical methodology for the future analysis of heparin/HS oligosaccharides of unknown structure. Typically, CID MS/MS fragmentation is not efficient to generate cross-ring fragmentation, especially for highly sulfate oligosaccharides. ETD and EDD are more gentle fragmentation methods that preserve labile modifications on the precursor molecule. A doubly sulfated tetrasaccharide

has been fragmented by EDD and shown to produce more cross-ring cleavages than observed by CID.⁹⁷ The rich information contained EDD MS/MS can also be used to distinguish GlcA from IdoA in GAG-derived tetrasaccharides.¹⁰² Most recently, negative electron transfer dissociation (NETD)/ETD was applied to analyze the glycosaminoglycans. Similar to EDD, NETD is also able to distinguish the IdoA from GlcA epimers in HS tetrasaccharides and suggests that a radical intermediate plays an important role in distinguishing these epimers.⁸³

FT-MS in negative-ion mode has also been found to be an important tool in longer chain oligosaccharides MS and MS/MS analysis which enable us to assign the fragments to a meaningful structure.^{18,103} The analysis of intact pGs, prepared through proteolysis of a PG, was the first demonstration of a top-down approach for GAG sequencing (Fig. 5).¹⁸ Bikunin pG, recovered from the therapeutic urinary bikunin PG, was fractionated by continuous elution preparative PAGE to obtain a simplified mixture for FT-MS and FT-ICR-MS analysis using a nanospray-ionization. Chain length and molecular mass are not the major limitations in the applications of FT-MS for sequencing GAGs, as single chains of heparosan up to 80 saccharide units have been analyzed.¹⁰⁴ A major limitation in obtaining excellent MS data in GAG sequencing is difficulties associated with the high-resolution, high throughput purification of GAG chains for sequencing. However, even with excellent spectral data, the manual interpretation of the MS and MS/MS data sets remains a daunting task. There is publicly available glycomics software, such as GlycoWorkbench, which can assist the manual interpretation of MS and MS/MS data.^{105,106} The absence of seamless glycobioinformatics for direct interpretation of a GAG sequence still represents a major impediment in the high throughput sequencing of GAGs.

Computer simulation for sequence determination

Computer modeling has been applied to simulate a GAG sequence as a number chain and then to compare a simulated GAG sequence to an actual GAG sequence.^{34,107,108} Similarly a computer can simulate the breakdown of these number chains and compare these simulated data to the breakdown of a GAG on treatment with a polysaccharide lyase.¹⁰⁹ As our understanding of Golgi-based biosynthesis of GAGs improves, it might be possible to use a computer-based number chain building model, based on the rules of Golgi-based biosynthesis, and test the outcomes of these simulations by simulating either their chemical, enzymatic or mass spectral fragmentation. Such simulations could then be compared to experimental sequencing data. Refinement of the model might then lead to an iterative improvement in our understanding of GAG biosynthesis and hence an improvement in our understanding of GAG and PG sequences.

Future challenges in PGs sequencing

The future prospects for PG sequencing brightened considerably this past year with the first sequencing of the simplest PG, bikunin, using top-down FT-MS analysis.¹⁸ This simple PG, with a single, short, structurally simple, polydisperse GAG chain, showed a single sequence motif that held through a wide range of chain sizes. It remains to be seen whether single sequence motifs will be observed for far more complex PGs. But since all PGs are biosynthesized through a similar pathway in the ER and Golgi it is now reasonable to hypothesize that, like the GAG chains of bikunin, the GAG chains of other PGs may also have a defined sequence.

Future challenges include broadening the number of sequenced PGs from one simple PG, bikunin to larger, more structurally complex PGs, such as decorin, a readily available dermatan sulfate PG. Decorin still has a single GAG chain but has both uronic acid epimers and considerably longer chains. Other PG sequencing targets should include members of the

KS and HS families but these are difficult to obtain in the multi-milligram quantities required to undertake GAG chain sequencing. Thus, more effort will be required for recovery of intact PGs from tissues and for their recombinant overexpression from cell culture to improve their availability. Moreover, these targets typically contain multiple sites for GAG chain attachment and would require the separation of GAG chains attached to different sites and their individual sequencing in parallel. Major technical challenges remain in the purification of individual or simple mixtures of chains for top-down sequencing approaches. New separation technology will need to be developed for both offline and online, high-resolution, high-throughput GAG chain purification. An improved understanding of PG and GAG biosynthesis and biosynthetic control will be needed to limit the number of possible sequence permutations to make sequencing of large, and highly complex GAGs possible. Computer simulation and automation¹⁰⁷ could also assist in improving our understanding of biosynthesis and sequence. An increased number of GAG-degrading enzymes with well-established specificity and action patterns would greatly assist gel-based middle-down sequencing efforts. Improvements in FT-MS and MS/MS methods will certainly be needed to enable the determination of highly sulfated GAG chains of high molecular mass, high structural complexity and dispersivity. Improved glycobioinformatics software will be essential in making the interpretation of complex spectral data sets efficient and practicable. Improvements in NMR and X-ray crystallography will be needed to convert sequence data into conformational data that will allow us to better understand the protein–GAG interactome. Ultimately, high-throughput *in vitro* and *in vivo* biological evaluation of PGs and GAGs having defined sequences will be needed to understand the biological ramifications of sequences and to establish well-defined SAR. Despite the long list of challenges the future of sequencing and proteoglycomics is a bright one.

Acknowledgments

The authors thank the National Institutes of Health for supporting this work in the form of grants GM38060, HL096972, HL094463, and GM090127.

References

1. Silbert JE, Sugumaran G. IUBMB Life. 2002; 54(4):177–186. [PubMed: 12512856]
2. Eskoand SB, Selleck JD. Annu Rev Biochem. 2002; 71:435–471. [PubMed: 12045103]
3. Nairn AV, Kinoshita-Toyoda A, Toyoda H, Xie J, Harris K, Dalton S, Kulik M, Pierce JM, Toida T, Moremen KW, Linnhardt RJ. J Proteome Res. 2007; 6(11):4374–4387. [PubMed: 17915907]
4. Capila I, Linnhardt RJ. Angew Chem, Int Ed. 2002; 41(3):391–412.
5. Couchman, JR. Transmembrane Signaling Proteoglycans. In: Schekman, R.; Goldstein, L.; Lehmann, R., editors. Annual Review of Cell and Developmental Biology. Vol. 26. Annual Reviews; Palo Alto: 2010. p. 89–114.
6. Gesslbauer B, Rek A, Falsone F, Rajkovic E, Kungl AJ. Proteomics. 2007; 7(16):2870–2880. [PubMed: 17654462]
7. Petitou M, van Boeckel CAA. Angew Chem, Int Ed. 2004; 43(24):3118–3133.
8. Kreuger J, Salmivirta M, Sturiale L, Gimenez-Gallego G, Lindahl U. J Biol Chem. 2001; 276(33):30744–30752. [PubMed: 11406624]
9. Turnbull JE. Methods Mol Biol. 2001; 171:129–139. [PubMed: 11450223]
10. Lee KB, Alhakim A, Loganathan D, Linnhardt RJ. Carbohydr Res. 1991; 214(1):155–168. [PubMed: 1954629]
11. Rudd PM, Dwek RA. Curr Opin Biotechnol. 1997; 8(4):488–497. [PubMed: 9265730]
12. Merry CLR, Lyon M, Deakin JA, Hopwood JJ, Gallagher JT. J Biol Chem. 1999; 274(26):18455–18462. [PubMed: 10373453]
13. Turnbull JE, Hopwood JJ, Gallagher JT. Proc Natl Acad Sci U S A. 1999; 96(6):2698–2703. [PubMed: 10077574]

14. Venkataraman G, Shriver Z, Raman R, Sasisekharan R. *Science*. 1999; 286(5439):537–542. [PubMed: 10521350]
15. Volpi N, Linhardt RJ. *Nat Protocols*. 2010; 5(6):993–1004.
16. Zaia J. *Chem Biol*. 2008; 15(9):881–892. [PubMed: 18804025]
17. Saad OM, Leary JA. *Anal Chem*. 2005; 77(18):5902–5911. [PubMed: 16159120]
18. Ly M, Leach FE, Laremore TN, Toida T, Amster IJ, Linhardt RJ. *Nat Chem Biol*. 2011; 7(11):827–833. [PubMed: 21983600]
19. Ly M, Laremore TN, Linhardt RJ. *OMICS*. 2010; 14(4):389–399. [PubMed: 20450439]
20. Sarrazin S, Lamanna WC, Esko JD. *Cold Spring Harbor Perspect Biol*. 2011; 3(7):a004952.
21. Schaefer L. *J Am Soc Nephrol*. 2011; 22(7):1200–1207. [PubMed: 21719787]
22. Heinegard D. *Int J Exp Pathol*. 2009; 90(6):575–586. [PubMed: 19958398]
23. *Essentials of Glycobiology*. 2. Cold Spring Harbor, NY: 2009.
24. Didraga M, Barroso B, Bischoff R. *Curr Pharm Anal*. 2006; 2(4):323–337.
25. Skidmore MA, Guimond SE, Dumax-Vorzet AF, Yates EA, Turnbull JE. *Nat Protocols*. 2010; 5(12):1983–1992.
26. Zhang FM, Sun PL, Munoz E, Chi LL, Sakai S, Toida T, Zhang HF, Mousa S, Linhardt RJ. *Anal Biochem*. 2006; 353(2):284–286. [PubMed: 16529709]
27. Hascall VC, Calabro A, Midura RJ, Yanagishita M. *Guide to Techniques in Glycobiology*. 1994; 230:390–417.
28. Michalski C, Piva F, Balduyck M, Mizon C, Burnouf T, Huart JJ, Mizon J. *Vox Sang*. 1994; 67(4):329–336. [PubMed: 7535497]
29. Goldoni S, Owens RT, McQuillan DJ, Shriver Z, Sasisekharan R, Birk DE, Campbell S, Iozzo RV. *J Biol Chem*. 2004; 279(8):6606–6612. [PubMed: 14660661]
30. Jin L, Abrahams JP, Skinner R, Petitou M, Pike RN, Carrell RW. *Proc Natl Acad Sci U S A*. 1997; 94(26):14683–14688. [PubMed: 9405673]
31. Linhardt RJ. *J Med Chem*. 2003; 46(13):2551–2564. [PubMed: 12801218]
32. Kreuger J, Spillmann D, Li JP, Lindahl U. *J Cell Biol*. 2006; 174(3):323–327. [PubMed: 16880267]
33. Turnbull JE. *Biochem Soc Trans*. 2010; 38:1356–1360. [PubMed: 20863313]
34. Gandhi NS, Mancera RL. *Chem Biol Drug Des*. 2008; 72(6):455–482. [PubMed: 19090915]
35. Desai BJ, Boothello RS, Mehta AY, Scarsdale JN, Wright HT, Desai UR. *Biochemistry*. 2011; 50(32):6973–6982. [PubMed: 21736375]
36. Lindahl U, Backstrom G, Thunberg L, Leder IG. *Proc Natl Acad Sci U S A*. 1980; 77(11):6551–6555. [PubMed: 6935668]
37. Imberty A, Lortat-Jacob H, Perez S. *Carbohydr Res*. 2007; 342(3–4):430–439. [PubMed: 17229412]
38. Ashikari-Hada S, Habuchi H, Kariya Y, Itoh N, Reddi AH, Kimata K. *J Biol Chem*. 2004; 279(13):12346–12354. [PubMed: 14707131]
39. Powell AK, Yates EA, Fernig DG, Turnbull JE. *Glycobiology*. 2004; 14(4):17R–30R.
40. Yayon A, Klagsbrun M, Esko JD, Leder P, Ornitz DM. *Cell (Cambridge, Mass)*. 1991; 64(4):841–848.
41. Schlessinger J, Plotnikov AN, Ibrahim OA, Eliseenkova AV, Yeh BK, Yayon A, Linhardt RJ, Mohammadi M. *Mol Cell*. 2000; 6(3):743–750. [PubMed: 11030354]
42. Thao KNN, Raman K, Tran VM, Kuberan B. *FEBS Lett*. 2011; 585(17):2698–2702. [PubMed: 21803043]
43. Zhang FM, Zhang ZQ, Lin XF, Beenken A, Eliseenkova AV, Mohammadi M, Linhardt RJ. *Biochemistry*. 2009; 48(35):8379–8386. [PubMed: 19591432]
44. Desai UR, Wang HM, Linhardt RJ. *Arch Biochem Biophys*. 1993; 306(2):461–468. [PubMed: 8215450]
45. Bienkowski MJ, Conrad HE. *J Biol Chem*. 1985; 260(1):356–365. [PubMed: 3965453]
46. Shively HE, Conrad JE. *Biochemistry*. 1976; 15(18):3932–3942. [PubMed: 9127]

47. Medeiros MGL, Ferreira T, Leite EL, Toma L, Dietrich CP, Nader HB. *Comp Biochem Physiol, B: Biochem Mol Biol.* 1998; 119(3):539–547. [PubMed: 9734337]
48. Xiao ZP, Tappen BR, Ly M, Zhao WJ, Canova LP, Guan HS, Linhardt RJ. *J Med Chem.* 2011; 54(2):603–610. [PubMed: 21166465]
49. Peterson SB, Liu J. *J Biol Chem.* 2010; 285(19):14504–14513. [PubMed: 20181948]
50. Linhardt, RJ. Analysis of Glycosaminoglycans with Polysaccharide lyases. In: Varki, A., editor. *Current protocols in Molecular Biology.* Wiley Interscience; Boston: 1994.
51. Knudson W, Gundlach MW, Schmid TM, Conrad HE. *Biochemistry.* 1984; 23(2):368–375. [PubMed: 6421315]
52. Fukuda MN. *Curr Protocols Mol Biol.* 2001; 17(17B):6–13.
53. Hanson SR, Best MD, Wong CH. *Angew Chem, Int Ed.* 2004; 43(43):5736–5763.
54. Freeman C, Hopwood J. *Adv Exp Med Biol.* 1992; 313:121–134. [PubMed: 1442257]
55. Gu KN, Linhardt RJ, Laliberte M, Gu KF, Zimmermann J. *Biochem J.* 1995; 312:569–577. [PubMed: 8526872]
56. Pervin A, Gu K, Linhardt RJ. *Appl Theor Electrophor.* 1993; 3(6):297–303. [PubMed: 8199222]
57. Uchimura K, Morimoto-Tomita M, Rosen SD. *Methods Enzymol.* 2006; 416:243–253. [PubMed: 17113870]
58. Freeman, C.; Hopwood, J. Lysosomal Degradation of Heparin and Heparan-Sulfate. In: Lane, DA.; Bjork, I.; Lindahl, U., editors. *Heparin and Related Polysaccharides.* Vol. 313. Plenum Press Div Plenum Publishing Corp; New York: 1992. p. 121-134.
59. Laremore TN, Ly M, Zhang ZQ, Solakyildirim K, McCallum SA, Owens RT, Linhardt RJ. *Biochem J.* 2010; 431:199–205. [PubMed: 20707770]
60. Gu KN, Liu J, Pervin A, Linhardt RJ. *Carbohydr Res.* 1993; 244(2):369–377. [PubMed: 8348558]
61. Maxamand W, Gilbert AM. *Methods Enzymol.* 1980; 65(1):499–560. [PubMed: 6246368]
62. Liu J, Desai UR, Han XJ, Toida T, Linhardt RJ. *Glycobiology.* 1995; 5(8):765–774. [PubMed: 8720074]
63. Jackson P. *Biochem J.* 1990; 270(3):705–713. [PubMed: 2241903]
64. Dasgupta F, Masada RI, Starr CM, Kuberan B, Yang HO, Linhardt RJ. *Glycoconjugate J.* 2000; 17(12):829–834.
65. Chuang WL, Christ MD, Rabenstein DL. *Anal Chem.* 2001; 73(10):2310–2316. [PubMed: 11393857]
66. Pervin A, Gallo C, Jandik KA, Han XJ, Linhardt RJ. *Glycobiology.* 1995; 5(1):83–95. [PubMed: 7772871]
67. Griffin CC, Linhardt RJ, Vangorp CL, Toida T, Hileman RE, Schubert RL, Brown SE. *Carbohydr Res.* 1995; 276(1):183–197. [PubMed: 8536254]
68. Zhang ZQ, McCallum SA, Xie J, Nieto L, Corzana F, Jimenez-Barbero J, Chen M, Liu J, Linhardt RJ. *J Am Chem Soc.* 2008; 130(39):12998–13007. [PubMed: 18767845]
69. Langeslay DJ, Beni S, Larive CK. *Anal Chem.* 2011; 83(20):8006–8010. [PubMed: 21913677]
70. Serber Z, Richter C, Moskau D, Bohlen JM, Gerfin T, Marek D, Haberli M, Baselgia L, Laukien F, Stern AS, Hoch JC, Dotsch V. *J Am Chem Soc.* 2000; 122(14):3554–3555.
71. Spraul M, Freund AS, Nast RE, Withers RS, Maas WE, Corcoran O. *Anal Chem.* 2003; 75(6):1536–1541. [PubMed: 12659219]
72. Korir AK, Almeida VK, Malkin DS, Larive CK. *Anal Chem.* 2005; 77(18):5998–6003. [PubMed: 16159133]
73. Pellegrini L, Burke DF, von Delft F, Mulloy B, Blundell TL. *Nature.* 2000; 407(6807):1029–1034. [PubMed: 11069186]
74. Faham S, Hileman RE, Fromm JR, Linhardt RJ, Rees DC. *Science.* 1996; 271(5252):1116–1120. [PubMed: 8599088]
75. Plotnikov AN, Schlessinger J, Hubbard SR, Mohammadi M. *Cell (Cambridge, Mass).* 1999; 98(5):641–650.

76. Berman HM, Bhat TN, Bourne PE, Feng ZK, Gilliland G, Weissig H, Westbrook J. *Nat Struct Biol.* 2000; 7:957–959. [PubMed: 11103999]
77. Saxena K, Schieborr U, Anderka O, Duchardt-Ferner E, Elshorst B, Gande SL, Janzon J, Kudlinzki D, Sreeramulu S, Dreyer MK, Wendt KU, Herbert C, Duchaussoy P, Bianciotto M, Driguez PA, Lassalle G, Savi P, Mohammadi M, Bono F, Schwalbe H. *J Biol Chem.* 2010; 285(34):26628–26640. [PubMed: 20547770]
78. Sisu E, Flangea C, Serb A, Zamfir AD. *Amino Acids.* 2011; 41(2):235–256. [PubMed: 20632047]
79. Laremore TN, Leach FE, Amster IJ, Linhardt RJ. *Int J Mass Spectrom.* 2011; 305(2–3):109–115. [PubMed: 21860600]
80. Huang RR, Pomin VH, Sharp JS. *J Am Soc Mass Spectrom.* 2011; 22(9):1577–1587. [PubMed: 21953261]
81. Bin Oh H, Leach FE, Arungundram S, Al-Mafraji K, Venot A, Boons GJ, Amster IJ. *J Am Soc Mass Spectrom.* 2011; 22(3):582–590. [PubMed: 21472576]
82. Bielik AM, Zaia J. *Int J Mass Spectrom.* 2011; 305(2–3):131–137. [PubMed: 21860601]
83. Wolff JJ, Leach FE, Laremore TN, Kaplan DA, Easterling ML, Linhardt RJ, Amster IJ. *Anal Chem.* 2010; 82(9):3460–3466. [PubMed: 20380445]
84. Zaia J. *Mass Spectrom Rev.* 2009; 28(2):254–272. [PubMed: 18956477]
85. Staples GO, Bowman MJ, Costello CE, Hitchcock AM, Lau JM, Leymarie N, Miller C, Naimy H, Shi XF, Zaia J. *Proteomics.* 2009; 9(3):686–695. [PubMed: 19137549]
86. Linhardt RJ, Wang HM, Loganathan D, Lamb DJ, Mallis LM. *Carbohydr Res.* 1992; 225(1):137–145. [PubMed: 1633599]
87. Zaia J, Li XQ, Chan SY, Costello CE. *J Am Soc Mass Spectrom.* 2003; 14(11):1270–1281. [PubMed: 14597117]
88. Wolff JJ, Laremore TN, Aslam H, Linhardt RJ, Amster IJ. *J Am Soc Mass Spectrom.* 2008; 19(10):1449–1458. [PubMed: 18657442]
89. Juhasz P, Biemann K. *Carbohydr Res.* 1995; 270(2):131–147. [PubMed: 7585697]
90. Tholey A, Heinze E. *Anal Bioanal Chem.* 2006; 386(1):24–37. [PubMed: 16830111]
91. Laremore TN, Murugesan S, Park TJ, Avci FY, Zagorevski DV, Linhardt RJ. *Anal Chem.* 2006; 78(6):1774–1779. [PubMed: 16536411]
92. Laremore TN, Zhang FM, Linhardt RJ. *Anal Chem.* 2007; 79(4):1604–1610. [PubMed: 17297962]
93. Smith RD, Loo JA, Edmonds CG, Barinaga CJ, Udseth HR. *Anal Chem.* 1990; 62(9):882–899. [PubMed: 2194402]
94. Chi LL, Wolff JJ, Laremore TN, Restaino OF, Xie J, Schiraldi C, Toida T, Amster IJ, Linhardt RJ. *J Am Chem Soc.* 2008; 130(8):2617–2625. [PubMed: 18247611]
95. Olson SK, Bishop JR, Yates JR, Oegema K, Esko JD. *J Cell Biol.* 2006; 173(6):985–994. [PubMed: 16785326]
96. Staples GO, Zaia J. *Glycobiology.* 2009; 19(11):1346–1346.
97. Wolff JJ, Amster IJ, Chi LL, Linhardt RJ. *J Am Soc Mass Spectrom.* 2007; 18(2):234–244. [PubMed: 17074503]
98. Zaia J, Costello CE. *Anal Chem.* 2003; 75(10):2445–2455. [PubMed: 12918989]
99. Leach FE, Wolff JJ, Xiao ZP, Ly M, Laremore TN, Arungundram S, Al-Mafraji K, Venot A, Boons GJ, Linhardt RJ, Amster IJ. *Eur J Mass Spectrom.* 2011; 17(2):167–176.
100. Saad OM, Leary JA. *Anal Chem.* 2003; 75(13):2985–2995. [PubMed: 12964742]
101. Schenauer MR, Meissen JK, Seo Y, Ames JB, Leary JA. *Anal Chem.* 2009; 81(24):10179–10185. [PubMed: 19925012]
102. Wolff JJ, Chi LL, Linhardt RJ, Amster IJ. *Anal Chem.* 2007; 79(5):2015–2022. [PubMed: 17253657]
103. Laremore, TN.; Leach, FE.; Solakyildirim, K.; Amster, IJ.; Linhardt, RJ. *Glycosaminoglycan Characterization by Electrospray Ionization Mass Spectrometry Including Fourier Transform Mass Spectrometry.* In: Fukuda, M., editor. *Methods in Enzymology, Glycomics.* Vol. 478. Elsevier Academic Press Inc; San Diego: 2010. p. 79–108.

104. Ly M, Wang ZY, Laremore TN, Zhang FM, Zhong WH, Pu D, Zagorevski DV, Dordick JS, Linhardt RJ. *Anal Bioanal Chem.* 2011; 399(2):737–745. [PubMed: 20407891]
105. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. *J Proteome Res.* 2008; 7(4):1650–1659. [PubMed: 18311910]
106. Tissot B, Ceroni A, Powell AK, Morris HR, Yates EA, Turnbull JE, Gallagher JT, Dell A, Haslam SM. *Anal Chem.* 2008; 80(23):9204–9212. [PubMed: 19551986]
107. Spencer JL, Bernanke JA, Buczek-Thomas JA, Nugent MA. *PLoS One.* 2010; 5(2):e9389. [PubMed: 20186334]
108. Cardin AD, Weintraub HJR. *Arteriosclerosis (Dallas).* 1989; 9(1):21–32.
109. Linhardt RJ, Cohen DM, Rice KG. *Biochemistry.* 1989; 28(7):2888–2894. [PubMed: 2742816]

Biographies

Lingyun Li is a research scientist at Rensselaer Polytechnic Institute. Dr Lingyun Li received his PhD from the Chemistry and Chemical Biology Department/Barnett Institute at Northeastern University (NEU) under the supervision of Prof. Barry Karger. His research at NEU focused on quantitative proteomics analysis using nano-LC LTQ-FT MS. On his graduation in 2007, Dr Li joined the Genzyme Corp. as a staff scientist focusing on drug discovery and glycolipids biomarker analysis utilizing LC/MS/MS. In 2011, Dr Li joined Professor Linhardt's group. His current research focuses on glycosaminoglycan structure analysis combining various separation technologies with high-resolution mass spectrometry.

Mellisa Ly is a postdoctoral fellow at Agilent Technologies, Inc. Dr Ly graduated in 2011 from Rensselaer Polytechnic Institute, where her analytical glycomics interest and expertise grew by glycosaminoglycan sequencing. During her doctoral studies at RPI, she published 13 peer-reviewed publications.

Robert J. Linhardt is Ann and John Broadbent, Jr. '59 Senior Constellation Professor of Biocatalysis and Metabolic Engineering at the Rensselaer Polytechnic Institute. His honors include the American Chemical Society Horace S. Isbell, Claude S. Hudson and Melville L. Wolfrom Awards, the AACP Volwiler Research Achievement Award, USP Award for Innovative Response to a Public Health Challenge, is an AAAS Fellow, and one of the Scientific American Top 10. His research focuses on glycobiology, glycochemistry and glycoengineering. Professor Linhardt has published over 550 peer-reviewed manuscripts and holds over 50 patents.

Fig. 1.

(A) Repeating disaccharide units of the various GAG families (adapted and modified from <http://www.ncbi.nlm.nih.gov/books/NBK1900/>). X = SO₃⁻ or H and Y = SO₃⁻ or Ac, “/” means “or” and “±” means “with/without”. (B) AT pentasaccharide binding sequence within heparin with structural variants shown and the critical 3-*O*-sulfo group shown in red.

Fig. 2.

A diagram of various proteoglycans in a cell and ECM is shown (adapted and modified from <http://www.ncbi.nlm.nih.gov/books/NBK1900/>). Protein cores (yellow), HS/HP GAG chains (red), CS/DS GAG chains (green), and KS GAG chains (blue).

Fig. 3. Biosynthesis of HS/heparin, CS/DS, KSI and KSII (adapted and modified from <http://www.ncbi.nlm.nih.gov/books/NBK1900/>). XylT-I, xylosyltransferase I; GalT-I, galactosyltransferase I; GalT-II, galactosyltransferase II; GlcAT-I, glucuronosyltransferase I; GalNAcTI, GlcNAc transferase I; ChSy1,2,3, chondroitin synthase 1,2,3; CS4ST, chondroitin sulfate GalNAc-4-*O*-sulfotransferase; CS6ST, chondroitin sulfate GalNAc-6-*O*-sulfotransferase; DSEpi, dermatan sulfate glucuronosyl C5 epimerase; DS4ST, dermatan sulfate GalNAc-4-*O*-sulfotransferase; DS2ST, dermatan sulfate uronyl-2-*O*-sulfotransferase; GlcNAcT I, GlcNAc transferase I; EXT1–EXT2, copolymerase complex (GlcNAc transferase–glucuronosyltransferase); NDST, GlcNAc *N*-deacetylase/*N*-sulfotransferase; C5epi, uronosyl C5 epimerase; 2OST, uronyl-2-*O*-sulfotransferase; 6OST, glucosamine-6-*O*-sulfotransferase; 3OST, glucosamine-3-*O*-sulfotransferases; Gal6ST, galactosyl-6-*O*-sulfotransferase; GlcNAc6ST, GlcNAc-6-*O*-sulfotransferase.

Fig. 4.

Principles of gel-based sequencing (adapted from ref. 13 and 57). (A) Fluorescence detection of different amounts of a 2-aminobenzoic acid (2AA)-tagged heparin tetrasaccharide subjected to electrophoresis on a 33% Tris-acetate mini-gel. (B) Exosequencing of a 2AA-tagged heparin tetrasaccharide with lysosomal enzymes and separation of the products on a 33% Tris acetate mini-gel (15 pmol per lane). After the exoenzyme treatments, the band shifts shown indicate the structure of the nonreducing-end disaccharide unit (lane 1, untreated). (C) IGS performed on a purified heparin hexasaccharide with the combinations of partial HNO₂ and exoenzyme treatments indicated (lane 1, untreated, 25 pmol; other lanes correspond to ~200 pmol per lane of starting sample for partial HNO₂ digest). (D) Polyacrylamide gel electrophoresis of a DS oligosaccharide fluorescently labeled with 2-aminoacridone (AMAC). Lane 1 contains a mixture of AMAC-labeled monosaccharide standards; lane 2 contains disulfated trisaccharide; lanes 3–5 contain disulfated trisaccharide treated with one or multiple exoenzymes followed by AMAC-labeling.

Fig. 5. FT-ICR-MS analysis of a bikunin pG fraction. (A) FT-ICR negative-ion mass spectrum of 5.80 kDa MR fraction by PAGE with 18 isobars and 63 parent-ions; (B) deconvolution of spectrum A; (C) CID FT-ICR-MS/MS spectra of parent-ion $m/z = 917.38$ ($z = 6$) and annotated fragment-ions providing sequence with dp27-5-Ser fragmentation pattern assigned from the spectrum. (Adapted from ref. 18 with permission.)

Table 1Properties of common PGs^{19,20,23}

PG class	PG	Number/type of GAG chains	Core protein size/kDa
HSPG	Glypican 1–6	1–3 HS	57–69
	CD44v3	1 HS	37
	Perlecan	1–4 HS	400
	Agrin	2–3 HS	212
	Collagen XVIII	1–3 HS	150
HS/CSPG	Syndecans 1–4	1–3 CS, 1–2 HS	31–45
	β -Glycan	1–2 HS/CS	110
	Neuropilin-1	1 HS/CS	130
	Serglycin	10–15 heparin/CS	10–19
CSPG	Bikunin	1 CS	18
	Decorin	1 DS/CS	36
	Biglycan	1–2 DS/CS	38
	Neuroglycan C	0–1 CS	~60
	Leprecan	1–2 CS	82
	Phosphacan	2–5 CS	175
	Thrombomodulin	1 CS	58
KSPG	Lumican	KS I	37
	Keratocan	KS I	37
	Fibromodulin	KS I	59
	Mimecan	KS I	25
	SV2	KS I	80
	Claustrin	KS II	105
	Aggrecan	KS II	200

Table 2

Enzymes used to examine GAG structure

GAG digestion enzymes	Enzyme specificity		Action pattern	Ref.
	Substrate	Specificity		
Heparin lyase I	Heparin/HS	GlcNS $\pm$ 6Sa 1–4 IdoA2Sa	Endolytic	44
Heparin lyase II	Heparin/HS	Linkages containing 1–4 IdoA $\pm$ 2Sa/ GlcA β 1–4	Endolytic	44
Heparin lyase III	HS	GlcNAc α /GlcNS $\pm$ 6Sa 1–4 IdoA α / GlcA β	Endolytic	44
Heparanase	HS	At GlcA β 1–4 GlcNS $\pm$ 3S $\pm$ 6Sa or GlcA β 1–4 GlcNAc with a GlcA2S β (not IdoA2Sa) in proximity	Endolytic	49
Chondroitin lyase ABC	CS/DS, HA with low efficiency	GalNAc $\pm$ 4S $\pm$ 6S β 1–4 IdoA α /GlcA $\pm$ 2S β 1–3 or HA	Endolytic	50
Chondroitin lyase ABC	CS/DS, HA with low efficiency	GalNAc $\pm$ 4S $\pm$ 6S β 1–4 IdoA α /GlcA $\pm$ 2S β 1–3 or HA	Exolytic	50
Chondroitinase lyase AC-I	CS/DS/HA	GalNAc $\pm$ 4S $\pm$ 6S β 1–4 GlcA $\pm$ 2S β 1– 3 or HA	Endolytic	50
Chondroitinase lyase AC-II	CS/DS/HA	GalNAc $\pm$ 4S $\pm$ 6S β 1–4 GlcA $\pm$ 2S β 1– 3 or HA	Exolytic	50
Chondroitinase lyase B	DS,	GalNAc $\pm$ 4S β 1–4 IdoA α 1–3	Endolytic	50
Hyaluronan lyase	HA or CS unsulfated regions	GlcNAc β 1–4 GlcA β 1–3	Endolytic	50
Hyaluronidase (mammalian)	HA or CS	GlcNAc β 1–4 GlcA β 1–3 and GalNAc $\pm$ 4S $\pm$ 6S β 1–4 GlcA $\pm$ 2S β 1–3	Endolytic	51
<i>Endo</i> - β -galactosidase	KS	Gal β 1–4 GlcNAc $\pm$ 6S β 1–3	Endolytic	52
Keratanase	KS	Gal β 1–4 GlcNAc $\pm$ 4S $\pm$ 6S β 1–3 Nonsulfated linkages are resistant	Endolytic	52
Sulfamidase (lysosomal)	Heparin/HS	Removes <i>N</i> -sulfo from NRE	Exolytic	53
Glucosamine-3/6-sulfatase (lysosomal)	Heparin/HS	Removes 3- <i>O</i> or 6- <i>O</i> sulfo from terminal GlcNAc	Exolytic	53
Glucuronate/iduronate-2-sulfatase (lysosomal)	Hep/HS CS/DS	Removes the 2- <i>O</i> sulfo from terminal GlcA/IdoA	Exolytic	53
Galactosamine-4/6-sulfatase (lysosomal)	CS/DS	Removes 4- <i>O</i> /6- <i>O</i> sulfo from terminal GlcNAc	Exolytic	53
Iduronidase/glucuronidase/hexosaminidase (lysosomal)	Hep/HS CS/DS	Removes non-sulfated terminal IdoA/GlcA/HexNAc	Exolytic	54
Glycuronidase (bacterial)	Hep/HS CS/DS	Removes non-reducing end C4-5 unsaturated uronate from oligosaccharides	Exolytic	55
Sulfoesterases 2/3/4/6 (bacterial)	Hep/HS CS/DS	Acts to remove specific sulfo groups from disaccharides	Exolytic	56
Sulf-1/2	HS	Removes 6- <i>O</i> -sulfo	Endolytic	57