Glycoconjugate Journal 19, 227–237, 2003
© 2003 Kluwer Academic Publishers. Manufactured in The Netherlands.

A starting place for the road to function

Jeremiah E. Silbert[1,2] and Geetha Sugumaran[1,3]

[1] *Department of Veterans Affairs Medical Center, Bedford, 200 Springs Road, Bedford, MA 01730, USA,* [2] *Brigham and Women's Hospital, Division of Rheumatology/Immunology/Allergy, Department of Medicine, Harvard Medical School,* [3] *Department of Biochemistry, Boston University School of Medicine*

Published in 2003.

I. Introduction

This special issue of Glycoconjugate Journal, is dedicated to Merton Bernfield, whose major contributions to this area of research were tragically cut short by his untimely death. The issue is intended for each of the contributors to review and discuss their published work and ideas concerning the use of genetic modifications in intact animals, thus providing information regarding the specific functions of proteoglycans and hyaluronan. We have been able to obtain reviews by most but not all of such investigators, and regret that some important work is absent. References cited in this background introduction are mostly reviews from readily accessible journals or Annual Reviews of Biochemistry.

It is clear that hyaluronan and proteoglycans (a special family of glycoproteins consisting of glycosaminoglycans covalently linked to protein) are involved in major diverse aspects of animal biology. Hyaluronan and proteoglycans have extracellular structural roles in cartilage, connective tissue matrix, and basement membrane, while cell surface proteoglycans interact with a host of extracellular substances related to multiple functions, and intracellular proteoglycans have special interactions with specific intracellular substances. The functions of most of the proteoglycans have not been precisely defined, but there are strong presumptions of function that derive from the structures and the known interactions of these compounds with other substances. Often, but not always, the glycosaminoglycan portions of the proteoglycans appear to be the "business ends" that provide the main functional aspects, while the core proteins direct the intracellular and extracellular trafficking and placement during synthesis of the glycosaminoglycan and final positioning in the appropriate locations with orientation for function. The glycosaminoglycans may act as receptors or as recognition sites for active agents or may be the directly active agents. In some cases this is due to the highly charged nature of these glycosaminoglycans, while in other cases it appears to be due to the specific locations of sulfate substituents on the glycosaminoglycans as well as the distribution of glucuronic/iduronic (GlcA/IdoA) residues. These differences have the potential to provide myriads of structures enabling highly specific interactions or modifications of the actions of other substances. The potential variations in structure are much more numerous than those of the shorter non-sulfated oligosaccharides found in the general group of glycoproteins.

Mutations have been described involving proteoglycan synthesis in humans as well as other animals. These include defects in cartilage aggrecan, formation of 3′-phosphoadenosine 5′-phosphosulfate (PAPS), and defective sulfate transport resulting in chick nanomelia, mouse brachymorphia, and human and other animal chondrodysplasias [1]; altered perlecan in Schwartz-Jampel [2] and Silverman-Handmaker syndromes [3]; altered glypican in Simpson-Golabi-Behmel syndrome [4,5]; and multiple mutations in *Drosophila* affecting proteoheparan sulfate [5–9]. There are also mutations of degradation caused by loss of specific degradative enzymes that result in glycosaminoglycan storage diseases known as "mucopolysaccharidoses" [10]. However, as valuable as investigations with natural mutations are, information regarding exact functions in vertebrates has been limited.

Modifications of genes, as presented in this special issue, to eliminate core proteins and enzymes of glycosaminoglycan formation are a welcome addition, providing observed phenotypes as major additions to those of the spontaneous mutations. Genetic modification eliminating the functional presence of a core protein will have effects highly specific to the location and function of the particular proteoglycan. This could be due to loss of the direct action of the protein itself in those tissues that contain the proteoglycan or loss of function due to total loss of the glycosaminoglycan moiety that is no longer present due to absence of the core. The core protein modifications are presented as follows:

1. The first review by Zhuo et al. describes a bikunin knockout, which results in mouse infertility. This is caused by loss of chondroitin to present side proteins to react with hyaluronan in cumulus oophorus.

2. The decorin knockout of Reed and Iozzo results in skin and tendon fragility secondary to defects in collagen fibrillogenesis. This is directly caused by absence of decorin, which normally interacts with collagen and presents chondroitin/dermatan sulfate for bridging between fibrils.
3. The biglycan knockouts of Young et al. develop osteoporosis and multiple musculoskeletal defects, that are more severe with double biglycan/fibromodulin knockouts.
4. Hassell et al. describe multiple neuromuscular effects and skeletal abnormalities in perlecan knockout mice that resembles cartilage and skeletal dysplasia in humans.
5. Watanabe and Yamada have examined link protein knockout mice and shown that they have cartilage defects similar to but milder than the defects in cartilage of animals having a naturally occurring aggrecan knockout.
6. Kao and Liu describe lumican and keratokan knockouts in mice that results in corneal stroma defects ostensibly because of the loss of keratan sulfate for interactions between fibrils.
7. Chakravarti also describes the corneal defects in lumican knockout mice and combines this with a fibromodulin knockout to suggest different growth stages for function of these two proteoglcans.
8. The extensive research on syndecan and syndecan knockouts from the Bernfield laboratiory are described in the review by Bellin et al. This includes mouse knockouts of syndecan-1 and 3 as well as work with *Drosophila.* Unexpected effects on appetite control as well as changes in susceptibility to tumor formation and microbial disease were found with mouse syndecan knockouts. Mert Bernfield was actively involved in every aspect of this work until three weeks before his death, despite the severe physical limitations of his Parkinson's disease—a true indication of his dedication, courage and tenacity.
9. Wilcox-Adelman et al. provide a review of the syndecan family as well as description of syndecan-4 knockout mice. They discuss the multiple interactions that syndecan-4 appears to show *in vitro* and delays in wound healing of the null mice. However, the mice were surprisingly free of overt morphological changes.
10. Ishiguro et al. describe changes in susceptibility to induced nephropathy and endotoxic shock in syndecan-4 null mice, and show that there is an effect on interleukin-1β and TGF-β as a mechanism.
11. Filmus describes the role of glypican in *Drosophila* and zebrafish and relates this to glypican-3 null mice and mutations in humans.

In contrast to the specificity in elimination of a core protein, modification of a substituent of glycosaminoglycans by loss of a synthetic enzyme may be reflected on multiple proteoglycans with diverse functions. The modification of a core protein is also likely to be more certain in outcome than loss of an enzyme of glycosaminoglycan synthesis, since a substitute might have limited or no ability to accomplish the same function. Loss of an enzyme of glycosaminoglycan biosynthesis, on the other hand, might be expected to impact multiple proteoglycans having multiple diverse functions, but still might have unexpected or limited effect because of the presence of isoenzymes that can substitute for the same reactions. There are good examples in the modification of glycosaminoglycan-synthesizing enzymes as follows:

1. Berninsone and Hirschberg review the use of *Caenorhabditis elegans* to demonstrate some effects of the loss in heparan sulfate-synthesizing enzymes, with particular attention to nucleotide sugar transport.
2. McDonald and Camenisch describe the growth retardant and cardiovascular defective development related to epithelial-mesenchymal transfomations in hyaluronan synthase-2 null mice.
3. Spicer et al. add to the information concerning hyaluronan synthase-2 and discuss the potential information to be obtained for hyaluronan synthase-3 by use of synthase-3 null mice.
4. Knockout of heparan sulfate 2-O-sulfotransferase is described by Wilson et al., who found failure of kidney formation suggesting that there is a disruption of signaling between ureteric bud and metanephric mesenchyme.
5. Shworak et al. have found that heparan 3-O-sulfate transferase-1 deficient mice have unexpected growth retardation, but no coagulopathy. This is despite the activity of the enzyme in determining the capability of heparan sulfate to affect coagulation by interacting with antithrombin.
6. Studies with *Drosophila*, demonstrating the role of heparan sulfate, is presented by Lin and Perrimon. Several models are described for affecting signaling by heparan sulfate proteoglycans, and loss of UDP-glucose (Glc) dehydrogenase, N-deacetylase/N-sulfotransferase and heparan polymerase enzymes are noted.

II. Glycosaminoglycans

A. Structure [11–17]

Glycosaminoglycans consist of glucosamine (GlcN) or (GalN) residues alternating with another sugar [galactose (Gal), glucuronic acid (GlcA), or iduonic acid (IdoA)] (Figure 1). The GlcN is either N-sulfated (GlcNS) or N-acetylated (GlcNAc) with occasional residues containing free amino groups, while the GalN is always N-acetylated (GalNAc). There may be O-linked sulfate substituents on the hexosamine and/or the uronic acid or Gal. Because the hexosamines alternate with the other sugars, it is convenient to think of the glycosaminoglycans as polymers consisting of repeating identical or similar disaccharide units. Although an individual chain may contain both GlcA and IdoA, no glycosaminoglycan chain has been found to contain both GlcN and GalN or both Gal and uronic acid as part of the repeating portions of the same polymer.

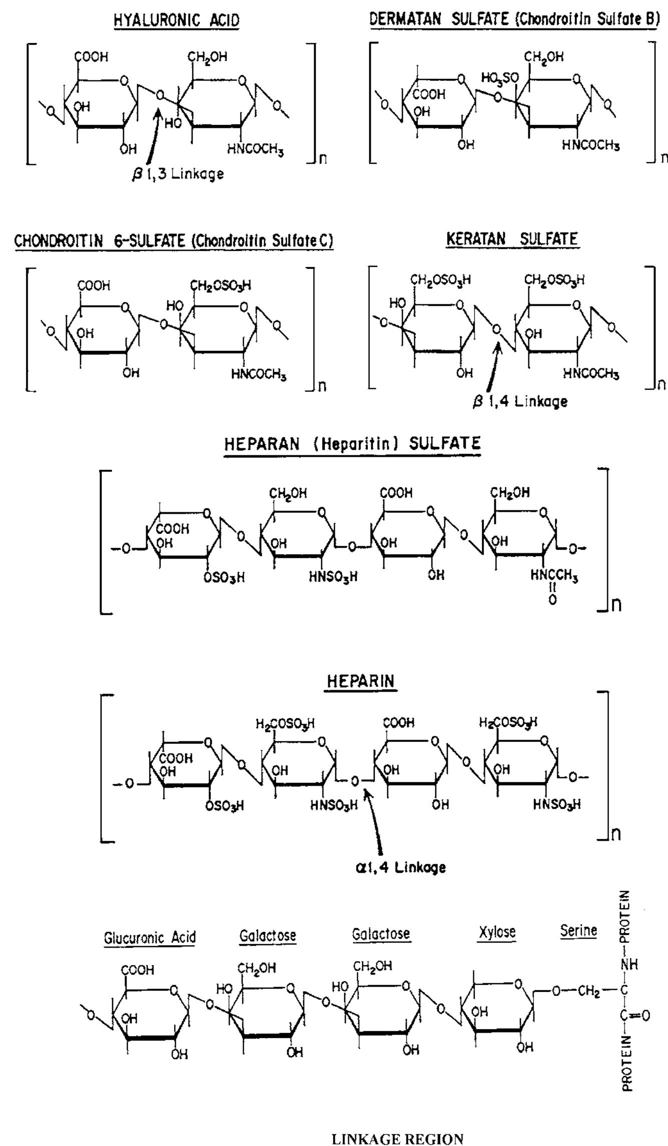

Figure 1. Structures of glycosaminoglycan repeating units and linkage tetrasaccharide.

The trivial names for the naturally occurring glycosaminoglycans are, hyaluronan (hyaluronic acid), chondroitin 4-sulfate (formerly called chondroitin sulfate A), chondroitin 6-sulfate (formerly called chondroitin sulfate C), dermatan sulfate (formerly called chondroitin sulfate B), heparin, heparan sulfate, keratan (poly N-acetyllactosamine), and keratan sulfate. The non-sulfated analogs, chondroitin, dermatan, and heparan, are generally not found in biological materials except at early stages during biosynthesis or when some cultured cells are grown in low sulfate concentrations or in the presence of high concentrations of chlorate [18] that prevents the formation of the sulfate donor, PAPS. Heparan sulfate-like substances are produced to a variable extent by most if not all animal cells, while heparin, chondroitin sulfate, dermatan sulfate, and keratan/keratan sulfate is found mainly in vertebrates. Hyaluronan, heparan and chondroitin, unsulfated and unattached to protein, have been found in some bacteria as well as in vertebrates and other animals.

Hyaluronan (hyaluronic acid) contains alternating GlcA and GlcNAc saccharides that are β-1,3- and β-1,4-linked respectively (Figure 1), has no sulfate substituents, and is not linked covalently to protein. Chondroitin sulfate glycosaminoglycans consist of alternating GlcA and GalNAc also β-1,3- and β-1,4-linked respectively, with variable amounts and location of sulfation, including non-sulfated GalNAc, GalNAc 4-sulfate, GalNAc 6-sulfate, GlcA 2-sulfate, or combinations of sulfate substitutions on the same saccharides. Depending upon the animal and the tissue source, chondroitin sulfates exhibit a wide range of sulfation and differ in the amounts of 6-sulfate and 4-sulfate in the same glycosaminoglycan chain. In general, the sulfation on a single chain is mostly or entirely 6-sulfate or 4-sulfate, so that chondroitin sulfate is not ordinarily found with equal amounts of both types of sulfate in the same chain. Usually a single GalNAc residue will have only one sulfate, either 4 or 6, but disulfated 4,6 GalNAc residues are found. In addition, 2-sulfated GlcA is found alternating with sulfated or non-sulfated GalNAc. The term dermatan sulfate defines a glycosaminoglycan similar to chondroitin sulfate but containing various amounts of IdoA rather than having GlcA as the only uronic acid. The IdoA residues are ordinarily only found adjacent to 4-sulfated GalNAc and not adjacent to 6-sulfated or non-sulfated GalNAc, while non-sulfated or 6-sulfated GalNAc residues may be found adjacent to GlcA in the same glycosaminoglycan chain. In addition, dermatan sulfate frequently has IdoA 2-sulfate in fairly high amounts.

Heparin mainly consists of disaccharide units of 2-sulfated IdoA alternating with N-sulfated GlcN that is mainly 6-O-sulfated (Figure 1) and may have small amounts of 3-O-sulfate, that is a required component for heparin's anti-coagulant activity. The overall sulfation averages two to two and a half sulfates per disaccharide. Heparan sulfate is similar to heparin except for a high proportion of unsulfated GlcA-GlcNAc disaccharides that are present in heparin to a much lesser degree. The overall degree of sulfation usually averages less than one sulfate per repeating disaccharide unit. The sugars for heparin and heparan sulfate are linked β-1,4- (or α-1-4- if the uronic acid is IdoA) and α-1,4- respectively (Figure 1), in contrast to the β-1,3- and β-1,4-linkages (or α-1-4- if the uronic acid is IdoA) of hyaluronan, chondroitin and dermatan. Despite the similarities in glycosaminoglycan structure of heparin and heparan sulfate, they are different as proteoglycans, since heparin is only found attached to the specific intracellular serglycin protein or as the free glycosaminoglycan chain split from this proteoglycan, while heparan sulfate is found attached to several families of extracellular and cell surface core proteins unrelated to serglycin.

Keratan sulfates are polymers composed of alternating Gal and GlcNAc residues β-1,4- and β-1,3-linked respectively. Most Gal and/or GalNAc are sulfated at the 6-position, with

variable degrees of sulfation along the keratan chains. The first one or two reducing end disaccharides generally are non-sulfated, the next few monosulfated, while the remainder of the chain mostly consists of disulfated disaccharides.

Glycosaminoglycan chains on proteoglycans range in size from approximately 15 disaccharide units to several hundred, while hyaluronan may be as much as several thousand disaccharides in length. The chains of all the glycosaminoglycans are linear (unbranched) and have a non-specific termination, so that some chains may end with uronic acid (or Gal for keratan sulfate) and some with hexosamine and with variable sulfation.

Chondroitin sulfate, dermatan sulfate, heparan sulfate, and heparin are all attached to core proteins at their reducing end through a tetrasaccharide region consisting of GlcA-Gal-Gal-Xyl (Figure 1) with the xylose (Xyl) glycosidically linked to the hydroxyl of Ser adjacent to a Gly in the protein core. The first sugar of the glycosaminoglycan chain (linked to the Gal) always is GlcA, but is considered to be a part of the linkage oligosaccharide, since its addition to the linkage region appears to be catalyzed by a specific enzyme different from the enzyme that is involved in the incorporation of GlcA into the rest of the glycosaminoglycan. The Xyl may be phosphorylated, and one or both Gal residues may be sulfated. In addition to the glycosaminoglycan substituents on proteoglycans, there may be O-linked and N-linked oligosaccharides similar or identical to those of various species of glycoproteins.

Skeletal keratan sulfate and corneal keratan sulfate chains are attached to core protein through oligosaccharides that are O-linked to serine or threonine and oligosaccharides N-linked to asparagine respectively, identical to O-linked and N-linked oligosaccharides of the general class of glycoproteins.

B. Biosynthesis [12,14,19–30]

Detailed biosynthesis of each of the glycosaminoglycans has been reviewed in a single issue of IUBMB Life [21,24–26]. Enzymes involved in polysaccharide synthesis and modification have been cloned from vertebrates, and other sources such as insect (*Drosophila*), *C.elegans* and zebra fish.

Hyaluronan is synthesized on the cell surface from nucleotide sugar precursors that are produced by metabolism from Glc as shown in Figure 2. Glucoamine (GlcN) can also be a precursor, which is helpful experimentally in providing specific radioactively-labeled glycoconjugates such as glycosaminoglycans. However GlcN is not generally a precursor *in vivo*, since it is not ordinarily present in blood or extracellular matrix. The alternate addition of both GlcA and GlcNAc to generate hyaluronan polysaccharide has been shown to be catalyzed by a single hyaluronan synthase protein in both bacteria and vertebrates. Three mammalian synthase genes have been identified (Table 1) and this has been shown to be developmentally regulated.

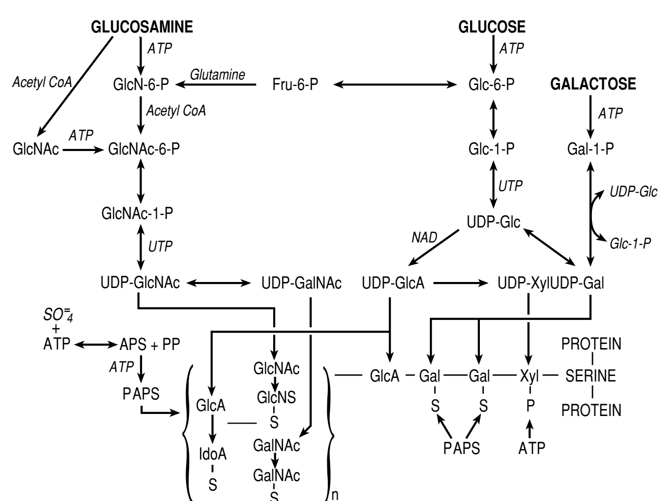

Figure 2. Pathways for precursor formation for biosynthesis of heparin/heparan sulfate and chondroitin/dermatan sulfate.

Formation of the polysaccharide portions of proteoglycans also proceeds from nucleotide sugars as shown in Figure 2. All reactions prior to addition of sugars to core protein and glycosaminoglycan polymerization take place with soluble enzymes in the cytosol, except for the decarboxylation of UDP-GlcA to form UDP-Xyl, which is within membranes of endoplasmic reticulum or Golgi. The activated form of sulfate, PAPS, is formed from sulfate and ATP in the cytosol.

Following formation and transport [30] of the precursor sugar nucleotides and synthesis of the protein core, synthesis proceeds in separate early Golgi localizations [24,28,29] by stepwise single sugar addition of Xyl from UDP-Xyl to Ser moieties of the core protein. This is followed by single sugar addition of two

Table 1. Mammalian enzymes of hyaluronan biosynthesis

Enzyme name	*Gene name*	*Amino acids*	*Chromosome location*		*Accession number*
			Human	*Mouse*	
Hyaluronan Synthase-1	HAS-1	578 (human) 583 (mouse)	19q13.4	17	U59269 (human) D82964 (mouse)
Hyaluronan Synthase-2	HAS-2	552	8q24.12	15	U54804 (human) U52524 (mouse)
Hyaluronan Synthase-3	HAS-3	554	16q22.1	8	U86409 (human) U86408 (mouse)

Table 2. Enzymes of chondroitin/dermatan sulfate biosynthesis

Name	*Source*	*Amino acids*	*Chromosome location*	*Accession number*
Xyl transferase	Human	827	16p13.1	AJ295748
Gal I transferase	Human	327	5q35.1-q35.3	AB028600
Gal II transferase	Human	329	1p36.3	AF092050
GlcA I transferase	Human	335	11q12q13	AB009598
GalNAc I transferase	Human	532	8	AB071403
Chondroitin Synthase	Human	802	15	AB023207
Chondroitin 6-O-Sulfotransferase	Mouse	472	9	AB008937
	Human	479	N.D.	AB012192
	Human	486	Xp11	AB037187
	Human	530	3q24-3q25	AF083066
	Human	411	11	U65637
Chondroitin 4-O-Sulfotransferase	Mouse	352	N.D.	AB030378
	Human	352	12q23.2-q23.3	
Chondroitin 4-sulfate-6-O-Sulfotransferase	Human	561	10q26	AB062423
Dermatan 4-O-Sulfotransferase	Human	376	N.D.	AF401222
Chondroitin/Dermatan 2-O-Sulfotransferase	Human	406	N.D.	AB020316
Epimerase	None cloned			

Gal residues from UDP-Gal and a GlcA residue from UDP-GlcA for chondroitin/dermatan and heparan/heparin. It is of particular note that formation of the linkage region oligosaccharides is the same for both chondroitin/dermatan and heparin/heparan so that addition of the first hexosamine as GalN or GlcN determines which of these glycosaminoglycans will be formed. The glycosyl transferases involved in the synthesis of the tetrasaccharide linkage common to chondroitin/dermatan sulfate and heparin/heparan sulfate have all been cloned (Table 2). Occasional 2-phosphorylation of the Xyl and 4- or 6-sulfation of the Gal on the linkage region of some chondrotin sulfate is found. The linkage region for keratan/keratan sulfate consists of glycoprotein-like oligosaccharides linked to N-Asn, O-Thr, or O-Ser and presumably the synthesis is by the same pathways as that of other glycoprotein synthesis.

Glycosaminoglycan formation follows in a later portion of Golgi with addition of the alternating saccharides of the glycosaminoglycan chains together with sulfation and/or epimerization of the glycosaminoglycan chains. As with hyaluronan synthases there are two enzyme activities residing on a single protein for GalNAc/GlcA and GlcNAc/GlcA synthases in chondroitin and heparin/heparan sulfate polymerization respectively. 4-Sulfation for chondroitin has been shown to take place while the chain is growing, and this is likely to be the case for GlcA/IdoA epimerization since it has been shown to be linked to 4-sulfation. 6-Sulfation of GalNAc in chondroitin and sulfation and epimerization for heparin/heparan sulfate may also occur as the chains are growing, but this has not been demonstrated as yet. N-Sulfation of heparin/heparan proceeds by N-deacetylation and N-sulfation catalyzed by a single protein that has both enzymatic activities. This is followed by GlcA/IdoA epimerization and 2-O sulfation, 6-O sulfation, and 3-O-sulfation, apparently in that order.

Enzymes for all the steps of polysaccharide formation, sulfation, and epimerization in chondroitin/dermatan (Table 2) and heparin/heparan sulfate (Table 3) synthesis have been cloned [31] with the exception of dermatan GlcA/IdoA epimerase, which has not been purified or cloned. Many of the enzymes have been cloned from several animal sources, and there are many instances of multiple isozymes for synthesizing the same structure. Sulfotransferases for glycosaminoglycans as well as glycoproteins and glycolipids have recently been reviewed [27].

No synthase that catalyzes the alternating transfer of Gal and GlcNAc to form keratan sulfate has been identified to date, although there are speculations that the β-1,4-Gal-1 transferase and β-1,3-GlcNAc transferase enzymes of glycoprotein synthesis may be involved in the synthesis of keratan polymer. Sulfation of both the sugars in keratan sulfate occurs at the 6-position. The sulfotransferases that catalyze the specific sugar sulfation have been cloned and expressed. Interestingly, the chondroitin 6-sulfotransferase that transfers sulfate to GalNAc residues of chondroitin sulfate can also transfer sulfate to Gal residues of keratan sulfate. In addition to this there is a separate sulfotransferase that transfers sulfate only to Gal residues of keratan.

III. Core proteins

Core proteins have been extensively characterized after purification and many have been cloned (Table 3), allowing the grouping of almost all those so far examined into gene families based on similarity of structures and having features that can carry covalently bound glycosaminoglycan chains. Terms for these families of proteoglycans have been and are being introduced by those investigators who initially describe them, and generally reflect their structure, the tissue in which they were initially

Table 3. Human enzymes of heparin/heparan sulfate biosynthesis*

Enzyme name	*Gene name*	*Amino acids*	*Chromosome location*	*Accession number*
GlcNAc Transferase-I	GlcNAcT-I (EXTL2)	330	1p11-p12	AF000416
GlcNAc Transferase-II	GlcNAcT-II (EXTL1)	676	1p36.1	U67191
GlcNAc Transferase-I/II	GlcNAcT-I/II (EXTL3)	919	8p12-p22	AF001690
Heparin/Heparan Polymerase	EXT1	746	8q24.1	Q16394
(GlcA-TII/GlcNAc-TII)	EXT2	718	11p11-p12	Q93063
N-Deacetylase/N-Sulfotransferase	NDST-1	882	5q32-q33	U36600
	NDST-2	883	10q22	U36601
	NDST-3	873	4q26-q27	AF07924
	NDST-4	872	4q25-q26	AB036429
Uronosyl-C5-Epimerase	GlcA-epimerase	618	15q23-q24	BAA74859
2-O-Sulfotransferase	HS2ST	356	1p22.2-p31.1	AB024568
6-O-Sulfotransferase	HS6ST-1	401	2q21.1	AB006179
	HS6ST-2	459	Xq26.1	AK027720
	HS6ST-3	468	N.D.	AB024567
3-O-Sulfotransferase	HS3ST-1	307	4p16	NM005114
	HS3ST-2	367	16p12	NM006043
	HS3ST-3A	406	17p12-p11.2	NM006042
	HS3ST-3B	390	17p12-p11.2	NM006041
	HS3ST-4	>250	16p11.2	AF105378
	HS3ST-5	346	11q22.2	AF503292

N.D.—Not determined.
*See Table 1 for linkage region enzymes Xyl transferase, Gal transferase I and II and GlcA transferase I which are in common for both chondroitin/dermatan sulfate and heparin/heparan sulfate.

found, or what is considered to be a major characteristic or function.

Size of the core proteins range from as small as 10 kDa to as large as 400 kDa, and overall size of the proteoglycans from as small as 80 kDa to as large as 3,500 kDa. Core protein sequences from evolutionarily distant animals have been shown to contain some structural features that have been preserved through millions of years of evolution indicating that the core proteins contain functionally important structures in a variety of organisms. Although these proteins may have functions independent of the glycosaminoglycan substituents, essentially all of the molecules carry these chains. Each family appears to be specific in its tissue localization.

Biosynthesis of core proteins does not appear to differ from that of other proteins except for post-translational proteolytic trimming such as that of the large aggrecan core protein, decorin, neurocan, and serglycin. The core proteins of many proteoglycans contain N- and/or O-linked oligosaccharides identical to those of the general class of glycoproteins. It can be assumed that biosynthesis of these oligosaccharides is identical to that of other glycoproteins. Compared to the general class of glycoproteins, which also may contain Ser-Gly dipeptides, only a small number of proteins become substituted with glycosaminoglycans. The mechanism of selection of these proteins for subsequent glycosaminoglycan attachment is unclear, but may be related to some as yet undetermined common structural feature.

Proteoglycans can best be grouped as (A) extracellular, (B) cell surface, and (C) intracellular, since their protein structures, presumed functions, and the role of glycosaminoglycan components are specific to these sites. The proteoglycans in each one of these groupings can be classified further into different families, each consisting of structurally related proteins. In addition a number of "part-time" proteoglycans that are either extracellular, cell surface, or intracellular are found in various locations, with much less indication of the role for their glycosaminoglycan components. These are also found without any bound glycosaminoglycans and do not appear to be members of any of the proteoglycan gene families. They do not resemble one another in any consistent manner regarding their structure as proteoglycans, their locations, or their presumptive functions.

A. Extracellular proteoglycans (Table 4A)

1. Hyalectin family [32–36]

These large (100 to 370 kD) chondroitin sulfate-containing proteoglycans, all capable of interacting with hyaluronan, include versican, neurocan, and brevican as well as aggrecan.

Aggrecan is the major proteoglycan of cartilage, the largest of all proteoglycans, typically containing as many as 100 chondroitin sulfate chains of 20–60 kDa, about 30 shorter keratan sulfate chains, and a small number of N-linked oligosaccharides, all attached in a "bottle brush" configuration. Chondroitin sulfate chains containing variable proportions of 4-sulfated,

Table 4. Classification of core proteins

	Core protein (∼kDa)	GAG type (number of chains)	Chromosomal mapping	
			Human	Mouse
A. Extracellular matrix				
Hyalectans				
Aggrecan	220	CS (∼100)	15q26	7
Versican	265–370	CS/DS (10–30)	5q13.2	13
Neurocan	136	CS (3–7)	19	8
Brevican	100	CS (1–3)	1q25-q31	3
SPACRCAN or IPM 150	160	CS (3)	3q11.2	9
SLRPs				
Class I				
Decorin	40	CS/DS (1)	12q23	10
Biglycan	40	CS/DS (2)	Xq28ter	X
Class II				
Fibromodulin	42	KS (2–3)	1q32	1
Lumican	38	KS (3–4)	12q21.3-22	10
Keratocan	38	KS (3–5)	12q	10
PRELP	44	KS (2–3)	1q32	1
Osteoadherin/modulin	42	KS (2–3)		
Class III				
Epiphycan	35	CS/DS (2–3)	12q21	10
Osteoglycin	35	KS (2–3)	1q31-32	1
Oculoglycan/Opticin	48	?	1q31	1
Basement membrane				
Perlecan	400–467	HS/CS (3)	1p36	4,distal
Agrin	250	HS (3)	1p32-pter	4,distal
Bamacan	138	CS (3)	10q25	19,distal
Type XVIII collagen	180	HS (1)	21q22.3	10
Part-time				
Phosphacan	173	CS (3–4)		
Endocan (ESM-1)	22	CS/DS (1)		
Testican	44	CS/HS (2)	5q31	
SV2	80	KS (?)		
Type IX collagen	60–80	CS	1p32-33	4
Clq inhibitor	30	CS (1)		
Colony-stimulating factor	43	CS (1)		
Appican	140–250	CS (1)		
Phosphacan	170	CS (3–4)		
Lubricin	120	CS/KS		
Superficial zone protein	45	CS/KS	Iq25	
B. Cell surface				
Glypican family				
Glypican-1	60–70	HS (2)	2q35-37	
Glypican-2/Cerebroglycan	60–70	HS		
Glypican-3	60–70	HS	Xq26	
Glypican-4	60–70	HS	Xq26	
Glypican-5	60–70	HS	13q31N-32	
Glypican-6	60–70	HS	13q31-32	
Syndecan family				
Syndecan-1	31 (80)	HS/CS (3–5)	2p23-24	Proximal 12
Syndecan-2	20 (48)	HS (2)	8q23	Proximal 15
Syndecan-3	38 (120)	HS/CS (3–5)	1p32-36	Distal 4
Syndecan-4	20 (35)	HS (2)	20q12-13	2

(*Continued on next page.*)

Table 4. (*Continued*)

	Core protein (~kDa)	GAG type (number of chains)	Chromosomal mapping	
			Human	Mouse
Part-Time				
Betaglycan	100–120	CS/HS		
NG2	300–330	CS (1)		
Thrombomodulin	70	CS		
CD44/epican	110	CS/HS		
FGFR-2	145	HS		
Transferrin receptor	90	HS		
C. Intracellular				
Serglycin	10–15	Heparin/CS (7–15)	10q22.1	
Part-time				
Invariant chain	38	CS (1)		

6-sulfated GalNAc, and some non-sulfated residues are concentrated in a central polypeptide region of the core protein where there may be more than 100 Ser-Gly dipeptides for attachment. The keratan sulfate chains are concentrated on a region immediately N-terminal to the chondroitin sulfate attachment sites, and smaller number of glycoprotein-like N-linked oligosaccharides are scattered on less well defined areas of the core protein. Globular N-terminal G1 and G2 domains adjacent to the keratan sulfate-binding region are capable of binding non-covalently to hyaluronan. This allows as many as 100 proteoglycan molecules to form an aggregate on a single hyaluronan molecule constituting a total molecular mass as high as 100,000–200,000 kDa. Binding is stabilized by a specific glycoprotein (similar in sequence to the N-terminal domain of aggrecan) that interacts with the hyaluronan and the binding region of the core protein. It is generally accepted that these largest of all proteoglycans function as a cushion to external pressure with concomitant slow decrease in the hydrated volume of the chondroitin sulfate chains; release of pressure results in gradual rehydration of glycosaminoglycan with reconstitution of original volume. The highly anionic nature serves as a barrier or filter of charged molecules.

Versican (named because of versatile, complex structure) is found as a major component in most extracellular matrix. This proteoglycan, containing 10–30 chondroitin/dermatan sulfate chains with little or no keratan sulfate; can be aggregated with hyaluronan.

Neurocan (named because it is found in nerve tissue) has N-, and C-terminal domains similar to those of aggrecan and versican. It contains chondroitin sulfate chains.

Brevican is a similar shorter proteoglycan found with or without chondroitin sulfate.

2. Leucine-rich core protein family (SLRP) [35,36–38]

These are major components of the interstitial matrix produced by fibroblasts and other cells. Core proteins are small, with several leucine-rich motifs. Characteristically they undergo proteolytic processing following synthesis, with removal of an additional small peptide from the N-terminus.

Decorin (name derived from characteristic binding to and thus "decorating" collagen) is an ubiquitous connective tissue matrix proteoglycan with one to three N-linked and/or O-linked oligosaccharides and a single chondroitin/dermatan sulfate glycosaminoglycan chain as large as 80 kDa attached at Ser-4. It binds to a specific region on type I collagen fibrils with the single dermatan sulfate chain interacting with dermatan sulfate chain of another decorin molecule bound to a neighboring collagen fibril [39]. The interaction between these two dermatan sulfate chains is thought to provide adherence between collagen fibrils and to regulate the spacing between them, thus affecting connective tissue integrity in skin and other tissues.

Biglycan is an ubiquitous component of connective tissue matrix, named because of two chondroitin/dermatan sulfate chains. It binds to TGF-β and to other proteins through core protein rather than the glycosaminoglycan chains.

Fibromodulin (named because of an effect on collagen fibrillogenesis) is keratan sulfate-containing and found in cartilage, tendon, and sclera. It has been reported to attach to TGF-β by means of its core protein.

Lumican (named because of its presence in cornea) is related to transparency. In particular, the keratan sulfate structure that it contains is considered to be important to the function, since opacities in corneal macular dystrophies were found to correlate with a deficiency in keratan sulfate or a lack of sulfation of keratan. Furthermore, corneal transparency in developing chick embryos was shown to correlate with the sulfation of keratan. Collagen fibrillogenesis was shown to be inhibited by lumican core protein. Lumican from aorta has been reported to contain only non-sulfated keratan.

3. Basement membrane family [40,41]

The most prominent of these is perlecan, so named because rotary shadowing micrographs resemble a string of pearls. It is the largest core protein (400 kDa) with domain I of five unique

sequences containing heparan sulfate and chondrotin/dermatan sulfate. Domains II, III, IV, and V are similar to the low density lipoprotein (LDL)-receptor, short arm of laminin A-chain, neuronal cell adhesion molecule (N-CAM), and the globular C-terminal region of laminin A-chain respectively.

Perlecan is the major proteoglycan of basement membranes, and plays an important part in filtration of charged molecules. For example it has been shown that removal of heparan sulfate from glomerular basement membrane by use of degradative enzymes results in loss of a barrier to anionic substances such as anionic ferritin. Perlecan in other basement membranes has also been shown to bind basic fibroblast growth factor (bFGF) and interferon gamma, with the heparan sulfate chains as the active factors in this binding. Since the binding of growth factors to heparan sulfate may protect them from proteolytic degradation, this may be a mechanism for storage. Proliferation of arterial smooth muscle cells can be inhibited by the heparan sulfate chains on perlecan, suggesting that perlecan may be involved in the regulation of smooth muscle growth. This regulation may involve the endocytosis and nuclear transport of the heparan sulfate chains. Other basement membrane core proteins include Agrin, Bamacan, and Type XVIII collagen.

In addition to hyalectans, SLRPs, and basement membrane proteoglycans, there are many extracellular part time proteoglycans as listed in the table.

B. Cell surface proteoglycans (Table 4B)

These appear to serve as receptors for growth factors and other components of the extracellular matrix, for cell-matrix and cell-cell interactions, and as receptors for other cell-cell interaction molecules. Heparan sulfate glycosaminoglycans are the usual interactive agents, although chondroitin sulfate may be present in addition to the heparan sulfate and the core proteins may also have a primary function.

1. Syndecan family [13,15,42,43]

These are four similar proteoglycans that are inserted into the plasma membrane via a highly conserved hydrophobic transmembrane domain, leaving a small domain in the cytoplasm. The extracellular domain has an extended configuration with attachment sites for heparan sulfate near the N-terminus away from the cell surface, and may have attachment sites for chondroitin sulfate near the cell surface. They received the name because of a presumptive function in "binding together" of extracellular matrix and cytoskeleton. Syndecan-1 is found on epithelial cells, syndecan-2 (originally called fibroglycan) on endothelial cells, syndecan-3 on cells of the nervous system, and syndecan-4 (also called amphiglycan and ryudocan presumptively for anticoagulant activity) on vascular epithelial cells.

Syndecan-1 and -3 are found with both heparan sulfate and chondroitin sulfate on the same core protein, while syndecan-2 and -4 have only heparan sulfate. *Drosophila* syndecan with heparan sulfate has an extensive homology to vertebrate syndecan.

2. Glypican family [6,9,15]

This is so named because the C-terminus of the cysteine-rich core protein is attached to cells through a glycophosphatidylinositol moiety [22]. It can be released from cell surface by specific phopholipases. In addition to glypican, which is found on a variety of cell types, it includes cerebroglycan, which is found in nerve tissue.

In addition to the syndcan and glypican proteoglycans, there are many part time cell surface proteoglycans as listed in the table.

C. Intracellular proteoglycans (Table 4C) [44]

Serglycin, the only intracellular proteoglycan described to date, is found in secretory granules of mast cells, basophils, natural killer cells, and less well-described locations in nuclei, and possibly other organelles. The core is the smallest of all the proteoglycans and contains a continuous stretch of 9–24 protease-resistant Ser-Gly repeats, which provide the name. In connective tissue mast cells serglycine has 7–15 heparin glycosaminoglycan chains, 100 or more disaccharides in length. However, oversulfated chondroitin is the glycosaminoglycan in mucosal mast cells. The glycosaminoglycan component binds serine proteases and vasoactive amines in the granules, and that appears to be its sole function as evidenced by knockout of mouse heparin N-deacetylation/N-sufation [45].

In addition to serglycin, the intracelllular invariant chain protein is a part time proteoglycan.

IV. Structure/function interactions [46–56]

Many proteins have been found to interact with heparin and heparan sulfate, apparently due to interaction of highly acidic IdoA-GlcNS disaccharide residues with a variety of amino acid sequences and secondary protein structures. Earlier work suggested a function in angiogenesis, cell attachment, cell regulation, morphogenesis and remodeling. It was clear, however, that in general the binding was not solely due to ionic interactions since other highly anionic polysaccharides, such as chondroitin sulfate or pentosan polysulfate, do not behave in the same manner. It is now known that distinct structures within the heparin or heparan sulfate chains provide the specific interactions and the presence of IdoA, can assume a greater number of conformational states than GlcA, enables the glycosaminoglycan chains to be more flexible. Development in the sequencing and classification of core proteins, together with some fine structure information about sulfate substituents and uronic acid epimerization, has provided an extensive list of potential structure/function relationships for heparan sulfate deduced from *in vitro* systems, as well as some structure/function information concerning chondroitin/dermatan sulfate and keratan sulfate.

Heparin is the best example of glycosaminoglycan function or action based upon a specific oligosaccharide sequence. The well known and extensively utilized anticoagulant activity of this glycosaminoglycan has been shown to be based upon the binding and activation of antithrombin III, and requires a specific pentasaccharide containing an unusual 3-O-sulfated GlcNS present in small numbers in a minority of the glycosaminoglycan chains. However heparin's activity and widespread use as an anticoagulant, does not appear to relate to its intracellular function *in vivo*.

Understanding the specific function of proteoglycans has been quite limited until the advent of gene characterization and subsequent demonstration of phenotypes resulting from genetic manipulation. Examination of genetically modified organisms as presented in the following 17 reviews are dramatic and frequently surprising examples of the new information that is being obtained by these techniques.

Acknowledgment

Supported by Department of Veterans Affairs, Medical Research Service.

References

1 Schwartz NB, Domowicz M, Chondrodysplasias due to proteoglycan defects, *Glycobiol* **12**, 57R–68R (2002).
2 Nicole S et al., Perlecan, the major proteoglycan of basement membranes, is altered in patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia), *Nat Genet* **26**, 480–3 (2000).
3 Arikawa-Hirasawa E, Wilcox WR, Le AH, Silverman N, Govindraj P, Hassell JR, Yamada Y, Dysegmental dysplasia, Silverman-Handmaker type, is caused by functional null mutations of the perlecan gene, *Nat Genet* **27**, 431–4 (2001).
4 Pilia G, Hughes-Benzie RM, MacKenzie A, Baybayan P, Chen EY, Huber R, Neri G, Cao A, Forabosco A, Schlessinger D, Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome, *Nat Genet* **12**, 241–7 (1996).
5 Filmus J, Selleck SB, Glypicans: Proteoglycans with a surprise, *J Clin Invest* **108**, 497–501 (2001).
6 Perrimon N, Bernfield M, Specificities of heparan sulphate proteoglycans in developmental processes, *Nature* **404**, 725–8 (2000).
7 Princivalle M, deAgostini A, Developmental roles of heparan sulfate proteoglycans: A comparative review in *Drosophila*, mouse and human, *Int J Dev Biol* **46**, 267–75 (2002).
8 Lander AD, Selleck, The elusive functions of proteoglycans: *In vivo* veritas, *J Cell Biol* **148**, 227–32 (2000).
9 Hacker U, Lin XH, Perrimon N, The *Drosophila* sugarless gene modulates wingless signaling and encodes an enzyme involved in polysaccharide biosynthesis, *Development* **124**, 3565–73 (1997).
10 Neufeld EF, Muenzer J, The mucopolysaccharidoses, in *Metabolic and Molecular Bases of Inherited Disease*, edited by Scriver CR, Beaudet AL, Sly WS, Valle D (McGraw-Hill, New York, 1995), pp. 2465–94.
11 Kjellen L, Lindahl U, Proteoglycans: Structures and interactions, *Annu Rev Biochem* **60**, 443–75 (1991).
12 Silbert JE, Kokenyesi R, Bernfield M, Proteoglycans: A special class of glycoproteins, in *Glycoproteins II*, edited by Montreuil J, Vliegenthart JFG, Schachter H (Elsevier, Amsterdam, 1997), pp. 1–31.
13 Carey DJ, Syndecans: Multifunctional cell-surface co-receptors, *Biochem J* **327**, 1–16 (1997).
14 Lindahl U, Kusche-Gullberg, Kjellen L, Regulated diversity of heparan sulfate, *J Biol Chem* **273**, 24979–82 (1998).
15 Bernfield M, Gotte M, Park PW, Reizes O, Fitzgerald ML, Lincecum J, Zako M, Functions of cell surface heparan sulfate proteoglycans, *Annu Rev Biochem* **68**, 729–77 (1999).
16 Funderburgh JL, Keratan sulfate: Structure, biosynthesis and function, *Glycobiol* **10**, 951–8 (2000).
17 Esko JD, Lindahl U, Molecular diversity of heparan sulfate, *J Clin Invest* **108**, 169–73 (2001).
18 Humphries DE, Sugumaran G, Silbert JE, Techniques for decreasing the sulfation of proteoglycans produced by cultured cells, *Methods Enzymol* **179**, 428–34 (1989).
19 Weigel PH, Hascall VC, Tammi M, Hyaluronan synthases, *J Biol Chem* **272**, 13997–4000 (1997).
20 Spicer AP, McDonald JA, Characterization and molecular evolution of a vertebrate hyaluronan synthase gene family, *J Biol Chem* **273**, 1923–32 (1998).
21 Itano N, Kimata K, Mammalian hyaluronan synthases, *IUBMB Life* **54**, 195–9 (2002).
22 Udenfriend S, Kodukula K, How glycosyl-phosphatidylinositol-anchored proteins are made, *Annu Rev Biochem* **64**, 563–91 (1995).
23 Rosenberg RD, Shworak NW, Liu J, Schwartz JJ, Zhang L, Heparan sulfate proteoglycans of the cardiovascular system: Specific structures emerge but how is synthesis regulated, *J Clin Invest* **99**, 2062–70 (1997).
24 Silbert JE, Sugumaran G, Biosynthesis of chondroitin/dermatan sulphate, *IUBMB Life* **54**, 177–86, (2002).
25 Sugahara K, Kitagawa H, Heparin and heparan sulfate biosynthesis, *IUBMB Life* **54**, 163–75, (2002).
26 Funderburgh JL, Keratan sulphate biosynthesis, *IUBMB Life* **54**, 187–94, (2002).
27 Fukuda M, Hiraoka N, Akama TO, Fukuda MN, Carbohydrate-modifying sulfotransferases: Structure, function, and pathophysiology, *J Biol Chem*, **276**, 47747–50 (2001).
28 Silbert JE, Sugumaran G, Intracellular membranes in the synthesis, transport, and metabolism of proteoglycans, *Biochim Biophys Acta* **1241**, 371–84 (1995).
29 Colley KJ, Golgi localization of glycosyltransferases: More questions than answers, *Glycobiol* **7**, 1–13 (1997).
30 Hirschberg CB, Robbins PW, Abeijon C, Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus, *Annu Rev Biochem* **67**, 49–69 (1998).
31 Esko JD, Selleck SB, Order out of chaos: Assembly of ligand binding sites in heparan sulfate, *Annu Rev Biochem* **71**, 435–71 (2002).
32 Schwartz NB, Pirok EW, Mensch JR, Domowicz MS, Domain organization, genomic structure, evolution and regulation of expression of the aggrecan gene family, *Prog Nucleic Acid Res Mol Biol* **62**, 177–225 (1999).
33 Luo W, Guo C, Zheng J, Chen TL, Wang PY, Vertel BM, Tanzer ML, Aggrecan from start to finish, *J Bone Miner Metab* **18**, 51–6 (2000).

34 Watanbe H, Yamada Y, Kimata K, Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cortical structure and function, *J Biochem* **124**, 687–93 (1998).
35 Iozzo RV, Murdoch AD, Proteoglycans of the extracellular environment: Clues from the gene and protein side offer novel perspectives in molecular diversity and function, *FASEB J* **10**, 598–614 (1996).
36 Iozzo RV, Matrix proteoglycans: From molecular design to cellular function, *Annu Rev Biochem* **67**, 609–52 (1998).
37 Hocking AM, Shinomura T, McQuillan DJ, Leucine-rich repeat glycoproteins of the extracellular matrix, *Matrix Biol* **17**, 1–19 (1998).
38 Fransson L-A, Belting M, Jonsson M, Mani K, Moses J, Oldberg A, Biosynthesis of decorin and glypican, *Matrix Biol* **19**, 367–76 (2000).
39 Scott JE, Extracellular matrix, supramolecular organisation and shape, *J Anat* **187**, 259–68 (1995).
40 Iozzo RV, Cohen IR, Grassel S, Murdoch AD, The biology of perlecan: The multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices, *Biochem J* **302**, 625–39 (1994).
41 Dunlevy JR, Hassell JR, Heparan sulfate proteoglycans in basement membranes, in *Proteoglycans: Structure, Biology and Molecular Interactions*, edited by Iozzo RV (Marcel Dekker, Inc., New York, 2000), pp. 275–326.
42 Bernfield M, Kokenyesi R, Kato M, Hinkes MT, Spring J, Gallo RL, Lose EJ, Biology of the syndecans: A family of transmembrane heparan sulfate proteoglycans, *Annu Rev Cell Biol* **8**, 365–93 (1992).
43 Zimmerman P, David G, The syndecans, tuners of transmembrane signaling, *FASEB J* **13**, S91–S100 (1999).
44 Sali A, Matsumoto R, McNeil HP, Karplus M, Stevens RL, Three-dimensional models of four mouse mast cell chymases. Identification of proteoglycan binding regions and protease specific antigen epitopes, *J Biol Chem* **268**, 9023–34 (1997).
45 Humphries DE, Wong GW, Friend DS, Gurish MF, Qiu WT, Huang C, Sharpe AH, Stevens RI, Heparin is essential for the storage of specific granule proteases in mast cells, *Nature* **400**, 769–72 (1999).
46 Park PW, Reizes O, Bernfield M, Cell surface heparan sulfate proteoglycans: Selective regulators of ligand-receptor encounters, *J Biol Chem* **275**, 29923–6 (2000).
47 Filmus J, Glypicans in growth control and cancer, *Glycobiol* **11**, 19R–23R (2001).
48 Iozzo RV, San Antonio JD, Heparan sulfate proteoglycans: Heavy hitters in the angiogenesis arena, *J Clin Invest* **108**, 349–55 (2001).
49 Gallagher JT, Heparan sulfate: Growth control with a restricted sequence menu, *J Clin Invest* **108**, 357–61 (2001).
50 Tammi MI, Day AJ, Turley EA, Hyaluronan and homeostasis: A balancing act, *J Biol Chem* **277**, 4581–4 (2002).
51 Day AJ, Prestwich GD, Hyaluronan-binding proteins: Tying up the giant, *J Biol Chem* **277**, 4585–8 (2002).
52 Turley EA, Noble PW, Bourguignon LYW, Signaling properties of hyaluronan receptors, *J Biol Chem* **277**, 4589–92 (2002).
53 Toole BP, Wight TN, Tammi MI, Hyaluronan-cell interactions in cancer and vascular disease, *J Biol Chem* **277**, 4593–6 (2002).
54 Iozzo RV, The biology of the small leucine-rich proteoglycans: Functional network of interactive proteins, *J Biol Chem* **274**, 18843–6 (1999).
55 Vlodavsky I, Friedmann Y, Molecular properties and involvement of heparanase in cancer metastasis and angiogenesis, *J Clin Invest* **108**, 341–7 (2001).
56 Bame KJ, Heparanases: Endoglycosidases that degrade heparan sulfate proteoglycans, *Glycobiol* **11**, 91R–98R (2001).