Annu. Rev. Biochem. 1998. 67:609–52

MATRIX PROTEOGLYCANS: From Molecular Design to Cellular Function

Renato V. Iozzo
Department of Pathology, Anatomy and Cell Biology, and the Kimmel Cancer Center, Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107-6799; e-mail: iozzo@lac.jci.tju.edu

KEY WORDS: glycosaminoglycan, heparan sulfate, dermatan sulfate, chondroitin sulfate, leucine-rich repeat

ABSTRACT

The proteoglycan superfamily now contains more than 30 full-time molecules that fulfill a variety of biological functions. Proteoglycans act as tissue organizers, influence cell growth and the maturation of specialized tissues, play a role as biological filters and modulate growth-factor activities, regulate collagen fibrillogenesis and skin tensile strength, affect tumor cell growth and invasion, and influence corneal transparency and neurite outgrowth. Additional roles, derived from studies of mutant animals, indicate that certain proteoglycans are essential to life whereas others might be redundant.

The review focuses on the most recent genetic and molecular biological studies of the matrix proteoglycans, broadly defined as proteoglycans secreted into the pericellular matrix. Special emphasis is placed on the molecular organization of the protein core, the utilization of protein modules, the gene structure and transcriptional control, and the functional roles of the various proteoglycans. When possible, proteoglycans have been grouped into distinct gene families and subfamilies offering a simplified nomenclature based on their protein core design. The structure-function relationship of some paradigmatic proteoglycans is discussed in depth and novel aspects of their biology are examined.

CONTENTS

INTRODUCTION 610
BASEMENT MEMBRANE PROTEOGLYCANS 610
General Structural Features 611
Genomic Organization and Transcriptional Control 616
Expression and Functional Properties 618

0066-4154/98/0701-0609$08.00

HYALECTANS: PROTEOGLYCANS INTERACTING WITH HYALURONAN AND LECTINS 624
General Structural Features 624
Genomic Organization and Transcriptional Control 629
Expression and Functional Roles 631
SMALL LEUCINE-RICH PROTEOGLYCANS 634
General Structural Features: Three Distinct Classes 635
Genomic Organization and Transcriptional Control 638
Control of Collagen Fibrillogenesis 640
Interactions with TGF-b and the Control of Cell Proliferation and Corneal Transparency 642
PERSPECTIVES 645

INTRODUCTION

This review focuses on areas of proteoglycan biology that are coming under the scrutiny of molecular biology, genetics, and mutant animal studies. Progress has been made in understanding the major biosynthetic pathways by which cells produce proteoglycans, some of the largest and most complex molecular structures in mammalian cells. However, the function of many of these compounds is not understood. Because of space limitations, only the matrix proteoglycans are discussed here. These can be separated into three groups: the basement membrane proteoglycans, the hyalectans—proteoglycans interacting with hyaluronan and lectins, and the small leucine-rich proteoglycans. The structure and function of these and other proteoglycan gene families have been covered in recent reviews (1–11).

The references cited here are not intended to be exhaustive but rather to direct the reader to pertinent primary literature where additional details can be found. In some instances, I have speculated rather freely, while in others I tried to confine such speculations to areas that can be tested experimentally. When possible, I have attempted to connect work on the biology of a specific proteoglycan, or family of proteoglycans, to fundamental cellular processes and pathology. To what extent are common molecular mechanisms involved? What are the underlying mechanisms that regulate the intrinsic function of a given proteoglycan? Is there any common theme in transcriptional regulation of proteoglycan gene expression? Why are so many genes expressed for seemingly identical functions? What is the level of redundancy? What kinds of mechanisms dictate the developmental expression of a specific proteoglycan? These and other important questions are answered in the context of our current knowledge of developmental processes and the extracellular matrix at large.

BASEMENT MEMBRANE PROTEOGLYCANS

Basement membranes are biochemically complex and heterogeneous structures containing laminin, collagen type IV, nidogen, and at least one type of

Table 1 General properties of basement membrane proteoglycans

Proteoglycan	Gene	Chromosomal mapping		Protein core (∼kDa)[a]	Glycosaminoglycan type (number)
		Human	Mouse		
Perlecan	HSPG2	1p36	4, distal	400–467	Heparan/chondroitin sulfate (3)
Agrin	AGRN	1p32-pter	4, distal	250	Heparan sulfate (3)
Bamacan[b]				138	Chondroitin sulfate (3)

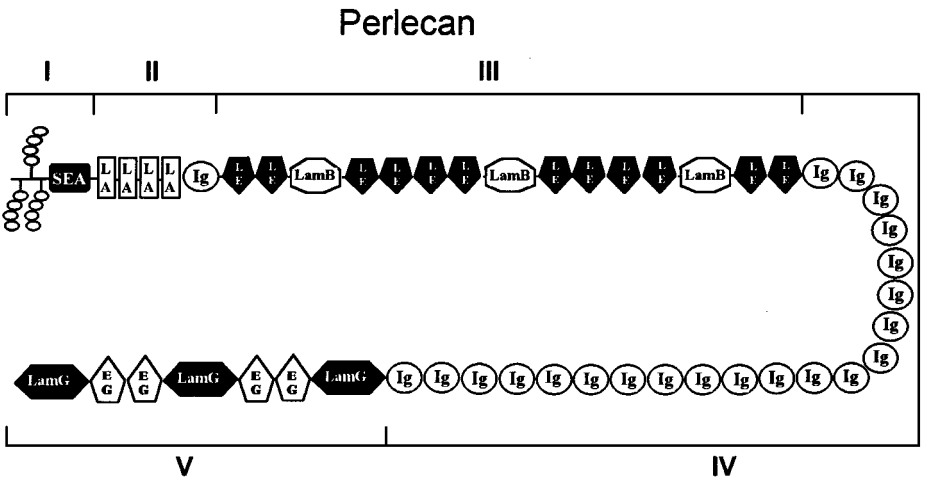
[a]The size of individual protein core does not include any posttranslational modification.
[b]We have no information regarding chromosomal assignment and designation of the bamacan gene. Southern analysis indicates the presence of a single-copy gene in eukaryotic cells (14).

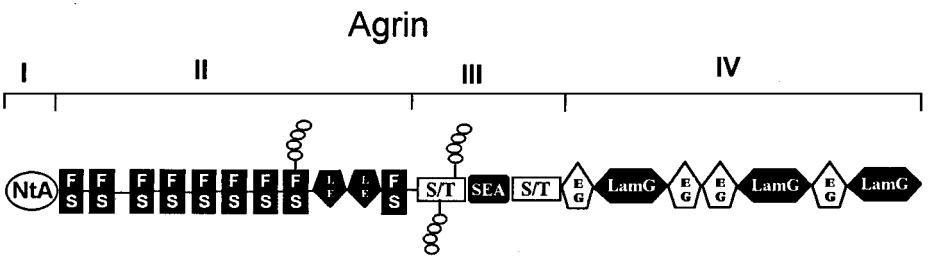
proteoglycan. Three proteoglycans are characteristically present in vascular and epithelial basement membranes of mammalian organisms: perlecan (12), agrin (13), and bamacan (14) (Table 1). The first two carry primarily heparan sulfate side chains, whereas the latter carries primarily chondroitin sulfate. The chimeric structural design of these proteoglycans suggests that they may be involved in numerous biological processes. It is unclear why only these three seemingly diverse molecules are associated with the basement membranes. The following sections discuss the structural and functional properties of these gene products and propose additional functional roles predicted from structural affinities.

General Structural Features

PERLECAN The name derives from its rotary shadowing appearance suggesting a string of pearls. The three glycosaminoglycan side chains are located at one end of the molecule, which also contains numerous globular regions interlinked by rod-like segments (15). The general structural features are shown in Figure 1 and summarized in Table 2. This multidomain proteoglycan is one of the most complex gene products because of its enormous dimensions and number of posttranslational modifications (9, 16–19). It comprises five domains that harbor protein modules used by disparate proteins involved in lipid uptake and metabolism, cell adhesion, and cellular growth. The N-terminal domain I contains three SGD tripeptides, the attachment sites for heparan sulfate chains. This region, which has no internal repeats and is devoid of cysteine residues, is enriched in acidic amino acid residues that facilitate heparan sulfate polymerization. Recombinant domain I can accept either heparan or chondroitin sulfate chains, and this selection of glyconation appears to be cell specific (20–23).

The distal portion of domain I encompasses a SEA module, named after the three proteins—sperm protein, enterokinase, and agrin—in which it was first identified. This module has been proposed to regulate binding to neighboring


Perlecan
I
II
III
SEA
LA
Ig
LF
LamB
LamG
EG
V
IV


Agrin
I
II
III
IV
NtA
FS
LF
S/T
SEA
EG
LamG

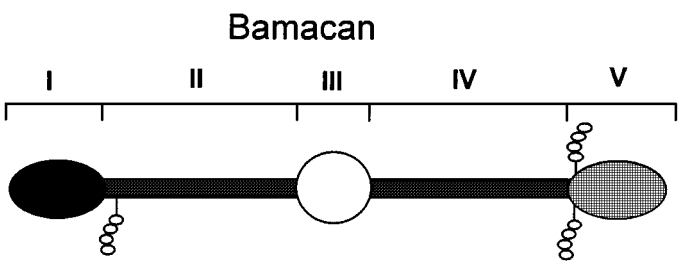

Bamacan
I
II
III
IV
V

Table 2 Structural and functional motifs in the human perlecan proteoglycan

Domain	Coding exons	Domain features	Homology	Proposed function
Signal peptide	1	Hydrophobic	Common type	Signal peptide
I	5	SEA	Sperm protein, enterokinase, agrin	Regulate binding to neighboring carbohydrate moieties
II	3	4 cysteine-rich repeats	LDL receptor	Binding to lipids (?)
IIa	2	1 Ig-like repeat	N-CAM	Spacer
III	27	3 globular and 4 cysteine-rich repeats	Short arm of laminin-1	Binding to fibronectin, interaction with cell surface integrin[a]
IV	40	21 Ig-like repeats[b]	N-CAM	Homophilic interaction (?)
V	16	3 globular and 4 EGF-like repeats[c]	G-domain of laminin-1	Oligomerization, cell binding, neurite outgrowth

[a]The murine species has an Arg-Gly-Asp (RGD) motif that could mediate the binding of perlecan protein core to integrins. However, the human perlecan lacks such a sequence.
[b]Apparently, the murine species has only 14 Ig-like repeats with the possibility of alternatively splice variants in this region. However, no formal evidence for splice variants has been presented so far.
[c]EGF, epidermal growth factor.

carbohydrate moieties (24) and can enhance heparan sulfate synthesis in mouse perlecan (22). Notably, all SEA-containing proteins are glycoproteins or proteoglycans with glycosaminoglycans attached proximally to SEA (24). Secondary structure prediction suggests a conformation of alternating β sheets and α helices (24). Circular dichroism studies of recombinant domain I demonstrate a distinct α helix/β sheet structure of approximately 20 and 60%, respectively

Figure 1 Schematic representation of three major proteoglycans found in basement membranes. The Roman numerals indicate the proposed domains. The symbols and designations of individual protein modules for perlecan and agrin are according to Bork & Patthy (24) with minor modifications: SEA, a module first identified in sperm protein, enterokinase, and agrin; LA, LDL receptor class A module; Ig, immunoglobulin-like repeat typically found in N-CAM; LE, laminin-1 EGF-like; LamB, a globular module similar to that found in the short arm of the α-1 chain of laminin-1; LamG, first identified as the G domain in long arm of the α-1 chain of laminin-1; EG, EGF-like; NtA, N-terminal domain that binds laminin; FS, follistatin-like; ST, serine/threonine rich. The glycosaminoglycan side chains are designated by strings of small circles.

(23). The N-terminal region of domain I contains six Ser/Thr residues that are substituted with galactosamine-containing oligosaccharides and one N-substituted Asn residue (23).

Domain II comprises four repeats homologous to the ligand-binding portion of the low-density lipoprotein (LDL) receptor (LA module) with six perfectly conserved cysteine residues and a pentapeptide, DGSDE, that mediates ligand binding by the LDL receptor (19). Rotary shadowing electron microscopy of recombinant domain II demonstrates a globular domain connected to a short segment, suggesting that the tandem arrays of LDL receptor class A modules form rod-like elements (25). Further analysis of recombinant domain II shows that it represents an autonomously folding unit within the perlecan structure (25). The function of this domain remains conjectural, and whether this molecule can indeed bind LDL is unknown. However, its location immediately following the heparan sulfate-attachment domain may favor the interaction of perlecan with either lipid bilayers or soluble lipids, or perhaps this region may direct the proteoglycan to the basolateral surfaces of epithelial cells for vectorial deposition into basement membranes (5, 15). Nidogen, a known ligand for perlecan (26, 27), binds to recombinant domain II, though less avidly than to the whole perlecan molecule (25).

Domain III, most homologous to the subdomain IVa and IVb of the short arm of the α-1 chain of laminin-1 (28), harbors three distinct globular domains (LamB modules) connected by short rods, the cysteine-rich epidermal growth factor (EGF)-like regions (LE modules) (29, 30). This domain contains an RGD tripeptide that can promote integrin-mediated cell attachment (31). However, because this sequence is not conserved in the human (32, 33), it is not clear whether additional regions in the protein core may mediate cell attachment.

Domain IV is the largest and the most repetitive since it contains 14 and 21 Ig-like repeats in the murine and human species, respectively. These repeats are similar to those found in members of the Ig gene superfamily such as those described in the neural cell adhesion molecule N-CAM. Thus this domain may be implicated in dimerization or intermolecular self-association.

Domain V has homology to the so-called G domain of the long arm of laminin-1 and has three distinct globular regions (LamG modules) connected by EGF-like repeats (EG modules). Domain V is responsible for self-assembly and may be important for basement membrane formation in vivo (34). This domain contains an LRE tripeptide that is a cell adhesion site for laminin-3 (S-laminin) (35).

The complexity of the perlecan protein core is mirrored by a series of post-translational modifications and additional not-well-understood events that lead to multiple isoforms of the parent molecule. Such biosynthetic events include fatty acylation of the protein core (36), attachment of both chondroitin and

heparan sulfate side chains (37–39), substitution with undersulfated or totally unsulfated heparan sulfate side chains (40), and secretion of the protein core without addition of any glycosaminoglycan side chains (41). These forms of perlecan may be tissue or cell specific; however, their function remains to be fully elucidated.

AGRIN A constituent of the basement membrane that causes aggregation of acetylcholine receptors, agrin is a major heparan sulfate proteoglycan of neuromuscular junctions (13, 42) and renal tubular basement membranes (43). Agrin is a multidomain protein that shares similarities with perlecan and laminin and can be divided into four distinct domains (Figure 1). Domain I harbors the first 130 amino acid residues following the signal peptide. This region has been identified in the chick as a novel module that binds laminin-1 and is thus called NtA, for N-terminal in agrin (44). This domain has been also identified in mouse and human expressed sequence tags and shows a very high conservation among species, up to 90% (44). Thus such a domain is likely to be operational also in mammalian agrin. Alternatively spliced variants may exist that harbor an insertion of approximately 21 amino acid residues (42, 45).

Domain II is characterized by nine follistatin-like repeats, the last of which is interrupted by the insertion of two cysteine-rich, EGF-like modules typically found in laminin β- and γ-chains. This domain may function as a protease inhibitor or may mediate growth-factor binding in vivo (46). If the two EGF-like repeats behave as in laminin, they might also mediate binding of agrin to nidogen. Domain III is highly glycosylated and is characterized by a central SEA module, as in perlecan, flanked by two Ser/Thr-rich regions. In this domain, there are two conserved glycosaminoglycan attachment sites and potential for several O-linked oligosaccharide attachments. Domain IV is the most similar to perlecan and comprises three G modules found in laminin α-chains interrupted by EGF-like repeats. By analogy to laminin and perlecan, this domain may self-assemble. Alternatively spliced variants of domain IV with insertion of small peptide sequences, 4–19 amino acid residues long, have been described (47). These short peptides can significantly influence the binding of agrin to heparin, α-dystroglycan, and the cell surface (46). Moreover, this proteoglycan harbors at least six potential sites for glycosaminoglycan attachment and five N-glycosylation sites in addition to the two Ser/Thr-rich regions described above. Thus agrin may be heterogeneous in tissues.

BAMACAN The presence of chondroitin sulfate–containing proteoglycans in the basement membranes of various tissues is well documented (5). Bamacan was cloned and sequenced recently and is the product of a single gene different from any other basement membrane proteoglycan (14). It is probably the same

proteoglycan that was synthesized in organ cultures of rat parietal yolk sac, the so-called Reichert's membrane (48), which also contained chondroitinase ABC-sensitive glycosaminoglycans (49).

Structurally, bamacan can be subdivided into five domains (Figure 1). Domain I is devoid of cysteine residues, is largely hydrophilic, and has several β turns. Domain II likely assumes a coiled-coil configuration owing to the absence of proline residues. A possible glycosaminoglycan attachment site is also present in this domain. Domain III is a rod-like region with four cysteine residues and a VTxG sequence, which may mediate cell attachment as in thrombospondin-1. Domain IV comprises the second coiled-coil domain very similar in size and overall structure to domain II. As in perlecan, it contains an LRE tripeptide that is a cell adhesion site for laminin-3 (S-laminin) (35). Domain V is also hydrophilic and contains four SG sequences that are potential binding sites for glycosaminoglycan side chains. This region, in addition to potentially interacting with domain I, may undergo heterotypic interactions to incorporate bamacan into the basement membrane (14).

Genomic Organization and Transcriptional Control

Among the basement membrane proteoglycans, only the genomic organization of human perlecan (50) and mouse agrin (51) have been described to date. By comparing these two gene products with other modular proteins, some interesting observations can be made regarding their evolutionary development. The human perlecan (HSPG2) is a single-copy gene located on the short arm of human chromosome 1 at 1p36 (52) and on a syntenic region of mouse chromosome 4 (53). It comprises 94 exons and spans at least 120 kb of continuous DNA (9). The gene duplication theory of molecular evolution is supported strongly by the remarkable conservation of the intron-exon junctions in the various modules of perlecan. For instance, the exon sizes of the LDL receptor–like repeats in domain II are identical to those encoding the ligand-binding region of the LDL receptor with only a few base-pair differences. In other parts of the perlecan gene, some of the repeats of domain IV are almost identical, differing by only two or three nucleotides. Moreover, there is a striking conservation of intron phases among the units encoding the Ig folds. Domain III shows no correlation between exon arrangement and either nominal domain or repeat boundaries. Notably, a similar comparison between the laminin-β1 and -β2 shows considerable divergence between these molecules with no conservation of exon structure or domain location. We can thus assume that the laminin-like region of the perlecan gene might have evolved from an ancestral gene that has undergone extensive rearrangement. As in the case of the other laminin-like region, the genomic organization of perlecan domain V lacks correlation between domain boundaries and exon structure.

The human agrin gene has been mapped to the distal region of the short arm of human chromosome 1, at 1p32-pter, relatively close to the perlecan gene and to the same syntenic region of mouse chromosome 4 (51). The intron-exon pattern of the agrin gene displays a remarkable correspondence to the proposed domain structure of the protein. Once again, the common theme of exon shuffling and duplication seems to have prevailed in the evolution of the agrin molecule. The follistatin repeats 1–7 in the N terminus are all encoded by single exons flanked by phase I introns. These follistatin repeats could have evolved from unequal crossover events after the deletion of the internal intron found between repeats 8 and 9. As in the case of laminin and perlecan, the C-terminal part of agrin shows no significant correlation between exon-intron organization of the gene and the modular organization of the protein. In summary, whereas at the amino acid level these two proteoglycans are related, genomic analysis shows a remarkable divergence.

To date, only the promoter of human perlecan has been sequenced (50) and tested for functional activity (54). Structurally, the human perlecan promoter is enriched in G + C nucleotides, lacks canonical TATA or CAAT boxes, and contains several cognate *cis*-acting elements and palindromic inverted repeats. These features are observed typically in housekeeping and growth factor–encoding genes that are generally devoid of TATA or CAAT boxes and contain multiple transcription initiation sites. The proximal promoter region contains four GC boxes and 15 consensus hexanucleotide-binding sites for the zinc-finger transcription factor Sp1, five of which are located in the first exon.

Another striking feature of the perlecan promoter is the presence of numerous AP2 motifs, eight residing in the first 1.5 kb and two in the most distal areas (54). Notably, the AP2 transcription factors can be suppressed by SV40 T antigen and can confer phorbol ester and cAMP induction (55). These cognate *cis*-acting elements are very likely to be operational in vivo since SV40 T antigen inhibits transcription of perlecan in renal tubular epithelial cells (56), whereas perlecan expression is markedly upregulated by phorbol ester in colon cancer (57) and in erythroleukemia K562 (58) cells, respectively. The perlecan promoter also contains several motifs that bind transcription factors involved in hematopoiesis, including two PEA3 motifs and nine GATA-1 motifs that are involved in erythrocyte differentiation (59). Because perlecan is expressed abundantly in the hematopoietic system (60), these transcription factors may also play a role in regulating perlecan gene expression during bone marrow development and lymphoid organ formation (61).

In the distal promoter region, perlecan contains a binding site for NF-κB, a factor that has been involved in interleukin-induced transcription of several genes (55). The full-length promoter is quite active in a variety of human cells with various histogenetic backgrounds, as well as in mouse fibroblastic

and melanoma cells (54). Collectively, these data are consistent with the fact that perlecan is ubiquitously expressed (60, 62). The perlecan promoter contains a transforming growth factor (TGF)-β-responsive element with a 5′-TGGCC.$N_{3–5}$.GCC-3′ consensus sequence (54) resembling that described in rat and mouse α2(I) collagen (63, 64), elastin (65), type I plasminogen activator inhibitor (66), and growth hormone (67) genes. This sequence binds to NF-1-like members of transcription factors and is transcriptionally activated by TGF-β (54), thereby validating previous results obtained in human colon carcinoma (68) and murine uterine epithelial (69) cells. In both cases, TGF-β induces the mRNA and protein levels of perlecan. Thus this TGF-β-responsive element may regulate the expression of this proteoglycan at sites of tissue remodeling and tumor stroma formation.

Other factors that regulate perlecan gene transcription are beginning to be elucidated. In F-9 embryonal carcinoma cells, induction of differentiation into the parietal endoderm phenotype occurs with a combination of cAMP and retinoic acid, which results in an induction of perlecan transcription (70). Notably, cAMP alone downregulates perlecan expression in glomerular epithelial cells (71). Another potential regulator of perlecan gene expression is glucose (72). Increased concentrations of glucose correlate with an inhibition of de novo proteoglycan synthesis by glomerular (73) and mesangial cells (72), and these effects may be posttranscriptionally controlled (74). In contrast, long-term exposure of human mesangial cells to elevated concentrations of D-glucose induces perlecan gene expression (75). Moreover, D-glucose can induce dysmorphogenesis of embryonic kidneys at least in part by reducing the expression of perlecan, which would thus act as an essential morphogenetic regulator of extracellular matrix (76, 77).

Expression and Functional Properties

PERLECAN A vast body of evidence that links perlecan to cell differentiation and tissue morphogenesis indicates that this gene product may play an important role in embryogenesis. Earlier studies identified perlecan in preimplantation embryos prior to the formation of a basement membrane (78). In addition to being deposited in the blastocyst interior, perlecan epitopes can be detected on the outer surface of the trophectoderm cells at the time of attachment competence (79), suggesting that perlecan may play a role in attachment of the embryo to the uterine linings.

Increased perlecan expression during blastocyst attachment competence is regulated at multiple levels. It can derive from increased transcription as in normal implantation or from increased translation of preexisting mRNA as in delayed implantation (80). A systematic study of murine embryogenesis has shown that perlecan expression appears early in tissues of vasculogenesis such

as the heart primordium and major blood vessels (81). Subsequently it accumulates in a number of mesenchymal tissues, especially in cartilage undergoing endochondral ossification, where it persists throughout the developmental stages and into adulthood (81). Perlecan expression correlates universally with tissue maturation and is always prominent in the endothelial cell basement membrane of all vascularized organs, particularly the liver, lung, spleen, pancreas, and kidney. Thus perlecan may play important roles not only in the early steps of blood vessel development but also in the maturation and maintenance of a variety of differentiated epithelial and mesenchymal tissues, among which cartilage is prominent.

The distribution of FGF-2 in various basement membranes (82) parallels that observed for perlecan in the mouse embryo (81). This suggests that perlecan plays a key role as a regulator of FGF-2 signaling (see also below) and as a gatekeeper to limit access of growth factors to subjacent target cells (82). Indeed, vascular heparan sulfate proteoglycan can activate or block FGF-2 activity (83) and can control the access of FGF-2 to the underlying vascular smooth muscle cells (84). Moreover, various isoforms of TGF-β (85) as well as platelet factor 4 (86) bind to specific heparan sulfate sequences.

The developmental timing of perlecan expression further suggests that this proteoglycan may play a role in controlling smooth muscle cell replication in vasculogenesis. Heparin-like molecules have long been implicated in vascular smooth muscle cell proliferation (87), and perlecan is a potent growth inhibitor of such proliferation (88). The pattern of perlecan mRNA and protein expression correlates inversely with the degree of smooth muscle cell replication during rat aortic development (89). Therefore, perlecan could modify the behavior of replicative cells by controlling the amount of growth factors involved in vascular morphogenesis (81). The report of a perlecan splice variant that activates FGFs during early neuronal development (90) supports this concept.

Mouse perlecan is capable of downregulating Oct-1, a transcription factor involved in vascular smooth muscle cell growth control (91). Because the addition of soluble heparin does not elicit the same response, it is plausible that the ability of perlecan to alter smooth muscle cell function resides in the coordinated binding of the protein core and heparan sulfate chains (91). The human system is more complex since a variety of proteoglycans, including perlecan synthesized by arterial endothelial cells, do not directly suppress the growth of vascular smooth muscle cells (92). Additional studies are needed to identify the signals involved in perlecan-mediated growth inhibition and changes in gene expression.

The strategic location of perlecan immediately suggests that this gene product may be involved directly in the modulation of cell surface events known to be altered in the multistep process of invasion and metastasis. A cardinal role

for cell-associated heparan sulfate proteoglycans was revealed by the finding that high-affinity receptor binding of FGF-2 is abrogated in mutant cell lines defective in heparan sulfate (93) and in myoblasts depleted of sulfated glycosaminoglycans (94). Bone marrow heparan sulfate proteoglycans, including perlecan, bind growth factors and present them to hematopoietic progenitor cells (95), whereas highly O-sulfated oligosaccharide sequences are required for FGF-2 binding and receptor activation (96, 97). Perlecan is involved directly in this coupling by acting as a low-affinity receptor and as an angiogenic modulator (98).

An emerging body of evidence further supports the notion that perlecan is involved directly in promoting the growth and invasion of tumor cells through its ability to capture and store growth factors (99) by entrapping them within both the basement membrane (100) and the tumor stroma (101). In melanomas a predominance of heparan sulfate proteoglycans at the cell surface is a marker of a more aggressive phenotype (102), and perlecan mRNA and protein levels are notably increased in the metastatic neoplasms (103). Purified perlecan enhances invasiveness of human melanoma cells (104), whereas contact with basement membrane perlecan augments the growth of transformed endothelial cells but suppresses that of their normal counterparts (105). Stable overexpression of an antisense perlecan cDNA in NIH-3T3 cells as well as in human metastatic melanomas leads to reduced levels of perlecan and concurrent suppression of cellular responses to FGF-2 (106).

In contrast, in fibrosarcoma cells, antisense expression of perlecan cDNA causes enhanced tumorigenesis characterized by heightened growth in vitro and in soft agar, increased cellular invasion into a collagenous matrix, and faster appearance of tumor xenografts in nude mice (107). Thus the cellular context is important in mediating perlecan's functions. Indeed, perlecan can behave as either an adhesive or an antiadhesive protein for endothelial and bone marrow cells, respectively (108). Perlecan also inhibits mesangial cell adhesion to fibronectin (109) and is antiadhesive for polymorphonuclear granulocytes (110). Though antiadhesive for hematopoietic and fibrosarcoma cells, perlecan still binds granulocyte/macrophage colony stimulating factor and presents it to the hematopoietic progenitor cells (108).

Large deposits of immunoreactive perlecan are present in the newly vascularized stroma of colon, breast, and prostate carcinomas (19). In tumor xenografts induced by subcutaneous injection of human prostate carcinoma PC3 cells into nude mice, though perlecan was actively synthesized by the human tumor cells, it was clearly deposited along the newly formed blood vessels of murine origin (19). Hence, perlecan deposited by growing tumor cells may act as a scaffold upon which proliferating capillaries grow and eventually form functional blood vessels. The observation that FGF-2 binds to heparan sulfate chains located

in the N-terminal domain of perlecan synthesized by human endothelial cells (111) reinforces the hypothesis that perlecan represents a major storage site for FGF-2 in the blood vessel wall. The release of growth factor/heparan sulfate complexes via controlled proteolytic processing (111, 112) is a physiologic mechanism that disengages biologically active molecules at the site of remodeling and tumor invasion (18).

Another function of perlecan is in regulating the permeability of the glomerular basement membrane (16). It has long been known that removal of heparan sulfate chains increases glomerular permeability to proteins and leads to proteinuria (113). Injection of monoclonal antibodies against heparan sulfate chains derived from glomerular proteoglycans is nephritogenic and can induce a selective proteinuria, likely due to the neutralization of anionic sites on the heparan sulfate (114). Alterations in expression patterns are also noted in both protein core and side chains in various glomerulopathies (115), indicating that perlecan may be involved in several renal pathologies. The urine of patients who have end-stage renal failure and are undergoing hemodialysis contains a fragment of human perlecan derived from the carboxyl end of the protein core (116), further suggesting a role in glomerular filtration.

AGRIN Agrin is among the best-characterized molecules of the synaptic basement membrane. First isolated as a glycoprotein from the basement membrane of the *Torpedo californica* electric organ on the basis of its ability to aggregate acetylcholine receptors, it was subsequently shown to be a key organizer of the postsynaptic apparatus at the neuromuscular junction (117). When added to cultured muscle cells, agrin causes aggregation of acetylcholine receptors and other proteins that are enriched at the neuromuscular junctions (46). It is now recognized that agrin is a heparan sulfate proteoglycan (13) and that the "functionally active" agrin investigated in the past is a proteolytically processed form of the parent molecule, essentially a large C-terminal fragment.

The N-terminal extension of agrin is required for the proper secretion of agrin, and this region is substituted with heparan sulfate side chains (42). Moreover, this N-terminal domain, the NtA domain (Figure 1), mediates binding to laminin-1 (44) as well as to laminin-2 and -4, the predominant laminin isoforms of muscle basement membranes; however, it has no affinity for either perlecan or collagen type IV (44). Thus the specific binding of agrin to laminin may provide the basis for its localization within basement membranes at the neuromuscular junctions (117) and in the renal tubules (43).

Agrin exists as a heparan sulfate proteoglycan in a variety of species including mouse, rat, cow, and human (118). Because nonglycosylated splice variants were not detected in any of the tissues investigated (118), it is highly likely that the agrin molecule occurs primarily as a proteoglycan. Moreover, agrin

epitopes have been found in basement membranes of skin, gastrointestinal tract, and heart (119), with a distribution that has no association with synapses. Thus agrin has molecular functions beyond its role in synaptogenesis.

A novel function for agrin is a proposed role in maintaining cerebral microvascular impermeability insofar as agrin accumulates in the brain microvascular basement membranes during development of the blood-brain barrier (120). In both avian and murine brain, agrin accumulates on vessels around the time the vasculature becomes impermeable. Moreover, the agrin isoform that accumulates at such loci lacks the 8- and 11-amino-acid sequences known to confer on agrin high potency in acetylcholine receptor clustering (120).

Of special interest are sites in agrin that undergo alternate mRNA splicing. One site, recently described in the avian form of agrin, is located in the N-terminal end of the molecule (45). In the developing chicken, this variant is expressed primarily by nonneuronal cells such as astrocytes, smooth muscle, and cardiac muscle cells. Its upregulation is consistent with the proportional increase in glial cells during brain development (45). Whereas motor neurons of chicken spinal cord express primarily this splice variant of agrin, muscle cells synthesize primarily agrin lacking the extra sequence (42).

Two additional sites of alternate splicing called a and b in avian species (y and z in rat) are positioned within domain IV of agrin in the second LamG domain and distal to the terminal EGF-like module (Figure 1). When small peptides are inserted in these regions, the overall properties of agrin are modified significantly. Insertion of a four-amino-acid sequence at the y site confers heparin-binding properties (47). When the insertion is present, the binding of agrin to α-dystroglycan is inhibited by heparin.

Considerable excitement followed the finding that agrin is bound by α-dystroglycan, a molecule belonging to the complex that links dystrophin to the cell surface (117). However, it has since been shown that the region of the agrin molecule by which it binds to α-dystroglycan is not essential for acetylcholine aggregating activity. The fact that agrin activity needs protein phosphorylation suggests that agrin may activate a protein kinase.

A significant advance in our understanding of how agrin functions has come from genetic analysis of mutant mice. Researchers have examined mice in which the differentially spliced exons required for the acetylcholine receptor clustering of agrin are deleted. These animals die perinatally and possess no acetylcholine receptor clusters associated with the motor nerves (121). In addition, axons of the motor neurons do not stop growing, and their growth cones fail to differentiate into presynaptic nerve terminals. These genetic experiments establish agrin as a key regulator of synaptogenesis at the neuromuscular junction.

A similar phenotype was obtained in mice harboring a targeted disruption of a gene encoding a muscle-specific receptor tyrosine kinase, called MuSK

(122). The phenotypic similarities between the two conditions indicate strongly that agrin acts via the MuSK receptor. Indeed, agrin induces autophosphorylation of MuSK within minutes, and this phosphorylation is observed only with spliced variants that are also active in acetylcholine receptor aggregation. However, agrin does not appear to bind directly to the MuSK receptor (123), requiring instead an accessory component that has not been identified. Collectively, these results indicate that signal-activated protein phosphorylation, a widespread component of growth-controlled systems, also plays a key part in synapse formation (124).

BAMACAN Among the basement membrane proteoglycans, bamacan's functions are the least well-known. As mentioned previously, bamacan likely represents the proteoglycan that was originally isolated and partially characterized from organ cultures of rat Reichert's membrane (48), the embryonic basement membrane sandwiched between the parietal endoderm and the trophoblast. Biosynthetic experiments revealed the presence of a large, high-density chondroitin sulfate proteoglycan with a core protein of about 135 kDa (48). Subsequent studies using immunohistochemical and histochemical techniques coupled with chondroitinase ABC digestion revealed the presence of two major proteoglycans (49): one sensitive and one resistant to chondroitinase ABC digestion. The latter is probably perlecan.

The concept that basement membranes can contain two or more proteoglycans is supported by the finding that in the EHS tumor matrix, both perlecan and bamacan exist as intrinsic constituents (39). Antibodies against the chondroitin sulfate proteoglycan of Reichert's membrane recognize bamacan in a variety of tissues (5, 125). The deduced protein core structure of bamacan (14) reveals a unique molecule with coiled-coil regions (Figure 1). Because of these unique structural features, bamacan is likely to be an important functional molecule within specialized basement membranes. For example, tissues undergoing morphogenesis appear to lack or express lower amounts of bamacan (5).

Bamacan is immunologically unrelated to perlecan, is apparently regulated during embryonic development, and has been implicated in the pathogenesis of several disease processes in which basement membranes are affected (5). Bamacan is expressed in nearly all of the basement membranes investigated so far (125, 126); however, there are some important exceptions. It is localized in the basement membrane of Bowman's capsule and within the mesangial matrix but is not found in the capillary glomerular basement membrane (125). This observation agrees with the fact that mesangial cells synthesize a spectrum of proteoglycans species that include bamacan (127).

In contrast, perlecan immunostaining is associated with all of the major basement membranes of the kidney. Immunoelectron microscopic studies have

further localized bamacan epitopes to subregions of the mesangium, consistently being found directly subjacent to the lamina of the perimesangial portion of the glomerular basement membrane and the juxtamesangial portion of the capillary endothelial cells. From developmental studies of skin, hair follicle, and kidney, one can infer that this proteoglycan may play a role in basement membrane stability (5). During these developmental processes, the presence of bamacan may inhibit, while the presence of perlecan may enhance, branching morphogenesis. How bamacan interacts specifically with other basement membrane constituents and how it imparts stability is not known.

HYALECTANS: PROTEOGLYCANS INTERACTING WITH HYALURONAN AND LECTINS

Molecular cloning has enabled the identification of a family of proteoglycans that share structural and functional similarities at both the genomic and protein levels. This family currently contains four distinct genes, namely versican, aggrecan, neurocan, and brevican (Figure 2 and Table 3). A common feature of these proteoglycans is their tridomain structure: an N-terminal domain that binds hyaluronan, a central domain that carries the glycosaminoglycan side chains, and a C-terminal region that binds lectins. On this basis, the term hyalectans, an acronym for hyaluronan- and lectin-binding proteoglycans, has been proposed (9). Alternate exon usage occurs extensively, and various degrees of glycanation and glycosylation make these proteoglycans appropriate molecular bridges between cell surfaces and extracellular matrices (128, 129).

General Structural Features

VERSICAN The largest member of the hyalectan gene family, versican (130), is the mammalian counterpart of the so-called PG-M isolated from avian tissue (131). Domain I (Figure 2) contains one Ig repeat followed by two consecutive modules, the link protein modules, which are involved in mediating the binding of proteins to hyaluronan (132). The entire link module is approximately 100 amino acids in length and has a characteristic consensus sequence with

Figure 2 Schematic representation of four members of the hyalectans, the hyaluronan and lectin-binding proteoglycans. The Roman numerals indicate the proposed domains. The symbols and designations are as follows: Ig, immunoglobulin-type repeat; LP, link-protein type module; GAG, glycosaminoglycan-binding domain, where α and β refer to the two alternatively spliced variants encoded by individual exons; EG, EGF-like module; Lectin, C-type lectin-like module; CR, complement regulatory protein; KS, keratan sulfate-attachment module. The glycosaminoglycan side chains are indicated by strings of small circles. The GPI-anchor in brevican is denoted by two filled circles and a curved tail.

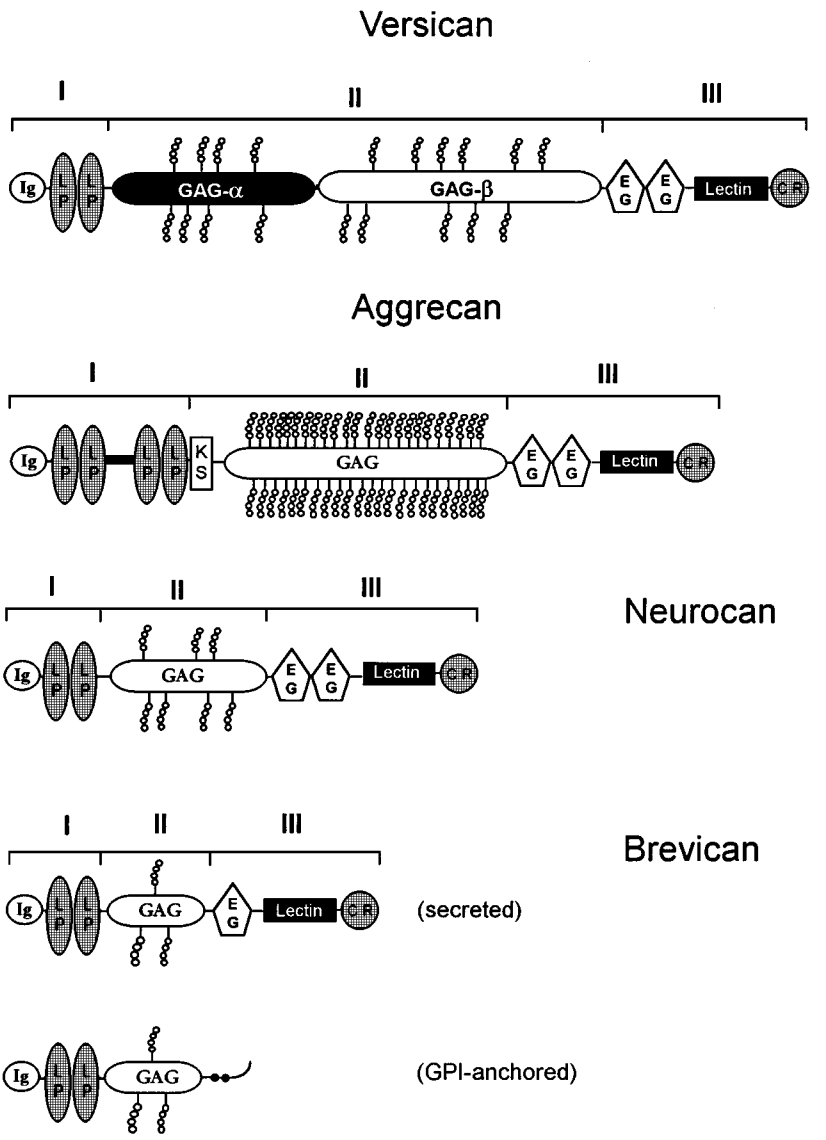

Versican
I
II
III
Ig
LP
LP
GAG-α
GAG-β
EG
EG
Lectin
CR
Aggrecan
I
II
III
Ig
LP
LP
LP
KS
GAG
EG
EG
Lectin
CR
Neurocan
I
II
III
Ig
LP
LP
GAG
EG
EG
Lectin
CR
Brevican
I
II
III
Ig
LP
LP
GAG
EG
Lectin
CR
(secreted)
Ig
LP
LP
GAG
(GPI-anchored)

Table 3 General properties of the hyalectans, proteoglycans interacting with hyaluronan and lectins

Proteoglycan	Gene	Chromosomal mapping		Protein core (∼kDa)[a]	Glycosaminoglycan type (number)
		Human	Mouse		
Versican	CSPG2	5q13.2	13	265–370	Chondroitin/dermatan sulfate (10–30)
Aggrecan	AGC1	15q26	7	∼220	Chondroitin sulfate (∼100)
Neurocan	NCAN		8	∼136	Chondroitin sulfate (3–7)
Brevican	BCAN	1q25–q31[b]	3	∼100	Chondroitin sulfate (1–3)

[a]The size of individual protein core does not include any posttranslational modification.
[b]The mouse brevican gene is closely linked to the osteocalcin gene on chromosome 3 and is flanked by markers that are all located in a syntenic region of the long arm of human chromosome 1, at 1q25–31 (150). Thus, it is likely that the human BCAN homologue also maps to this region of human chromosome 1.

four disulfide-bonded cysteine residues. The solution structure of a homologous link module was elucidated recently and consists of two α helices and two antiparallel β sheets arranged around a large hydrophobic core (133). A hydrophobic/hydrophilic region is proposed to mediate hyaluronan binding in the various members of the link module superfamily since there is conservation of the hydrophobic and some of the charged residues. Notably, there is a striking similarity between the link module and the C-type lectin domain (133). These domains have identical topologies and a similar organization of their secondary structure. The major difference is in the length of the loop. The long loop that contains the Ca^{2+}-binding residues in the lectin domain is missing in the link module.

Because there is no evidence that the binding of hyaluronan to the link protein module requires Ca^{2+}, it is quite likely that this structural difference is involved directly in dictating the differential affinity of the two domains for carbohydrates. The structural similarities between the link module and the C-type lectin domain suggest that they have a common evolutionary origin, even though the sequence similarity between the two is not as evident (132). Both recombinant versican and a truncated form of versican containing domain I bind to hyaluronan with a K_d of approximately 4 nM, in the same range as aggrecan (134). This observation suggests that versican may form a molecular link between lectin-containing glycoproteins at the cell surface and extracellular hyaluronan. Because hyaluronan is bound to the cell surface via its CD44 receptor, versican may also stabilize a large supramolecular complex at the plasma membrane zone.

Domain II consists of two large subdomains, designated GAG-α and GAG-β (135), that are encoded by two alternatively spliced exons (136). These regions lack cysteine residues and contain as many as 30 potential consensus sequences for glycosaminoglycan attachment as well as several binding sites for N- and O-linked oligosaccharides.

There are at least four possible splice variants of mammalian versican. The largest one, designated V0, contains both GAG-α and GAG-β and has an estimated M_r of approximately 370 excluding the signal peptide and the possible posttranslational modifications. The other three variants contain only GAG-β (V1 variant), GAG-α (V2 variant), or neither region (V3 variant). In birds, there is an additional exon, designated PLUS, in the N-terminal region corresponding to the mammalian GAG-α (137). This exon can be alternatively spliced giving rise to two additional isoforms. No corresponding region was found in the mammalian genome, suggesting a divergence of this domain during evolution. Sequence homology, however, indicates that the PLUS domain of avian versican may correspond to the keratan sulfate attachment region of aggrecan (137).

Domain III contains a series of structural motifs including two EGF-like repeats, a C-type lectin domain, and a complement regulatory protein-like module. These motifs are characteristically observed in the selectins, adhesion receptors regulating leukocyte homing and extravasation at inflammatory sites. The C-type animal lectins harbor a Ca^{2+}-dependent carbohydrate-recognition motif that is highly conserved among species (9, 138). The recombinant C-terminal portion of avian versican can bind heparin, heparan sulfate, and simple carbohydrates (139), whereas the recombinant human lectin domain can bind fucose and GlcNAc as well as tenascin-R (140). The binding is Ca^{2+} dependent and is abrogated by deglycosylation of tenascin-R. Other C-type lectin domains may have different saccharide-binding specificity, a mechanism that may provide additional specialized and refined functions for the hyalectans (138).

AGGRECAN The overall organization of aggrecan, the main proteoglycan of cartilaginous tissues, is similar to that of versican with a few exceptions. Domain I contains four link protein-like modules in addition to the Ig-like repeat (Figure 2). These modules form two globular domains also known as G1 and G2 (141). An interglobular region has a rod-like structure and contains cleavage sites for proteases involved in the degradation of aggrecan (8). The function of the G2 domain is poorly understood insofar as this region does not mediate the binding to hyaluronan. Immediately following the G2 domain is a relatively small region that contains numerous keratan sulfate consensus sequences. This domain has few similarities among species, and its size also varies in different species (142).

Domain II is the largest domain of aggrecan and contains the glycosaminoglycan-binding region. This domain is similar in size to the GAG-β of versican but harbors many more consensus sequences for glyconation. Thus a fully glycosylated aggrecan may contain up to 100 chondroitin sulfate chains. The human glycosaminoglycan-binding domain harbors a polymorphism that depends on a high degree of sequence conservation (143). A variable number of tandem repeats generates at least 13 different alleles in the general population with repeats numbering between 13 and 33. This could lead to a great variation in the degree of glycosylation and ultimately charge (sulfation) of the parent molecule within cartilage.

The structural motifs in domain III are very similar to those of versican; however, there is evidence that in both the human and murine species, the EGF repeats can be alternatively spliced (144). As in the case of versican, the lectin-like domain of aggrecan can bind simple sugars in a Ca^{2+}-dependent manner (145). Thus aggrecan may also be bridging or interconnecting various constituents of the cell surface and extracellular matrix via its terminal domains.

NEUROCAN The third member of the hyalectan gene family is neurocan, a chondroitin sulfate proteoglycan originally cloned from rat brain (146). Rotary shadowing electron microscopy of tissue-derived neurocan reveals two globular domains connected by a central rod of 60–90 nm (147), in agreement with the organization derived from protein sequencing and cDNA cloning (Figure 2). As with other members of the hyalectan gene family, neurocan has an N-terminal region with all the typical arrangements found in link protein. Recombinant neurocan domain I interacts with hyaluronan in gel permeation assays, and the isolated, retarded preparations contain supramolecular complexes of hyaluronan and globular profiles of domain I (147). Thus all domain Is likely mediate hyaluronan binding in vivo. Domain II contains at least seven potential binding sites for glycosaminoglycans and has no significant homology to any other protein.

Recombinant studies have demonstrated that glycanation is restricted solely to this central domain even though other portions of the molecule contain SG repeats (147). The C-terminal domain is again very similar to that of the other members of the hyalectan gene family with approximately 60% identity between the rat neurocan and human versican and aggrecan. Although not formally demonstrated, this domain may mediate the binding of neurocan to a variety of brain glycoproteins including Ng-CAM, N-CAM, and tenascin (discussed below).

BREVICAN Brevican is the most recently discovered member of this class of proteoglycans and takes its eponym from the Latin word *brevis* (short) because

it has the shortest glycosaminoglycan-binding region (Figure 2) (128, 148). Within conserved domains, sequence homology with the other members is relatively uniform (55–60%). However, domain II shows little homology to the corresponding regions of the other hyalectans. This domain is also characterized by a relatively high content of glutamic acid, including a sequence of eight consecutive residues. Such a cluster of acidic residues, which is also present in the link protein-like module of versican, may mediate binding to cationic proteins and, perhaps, minerals.

Similarly to neurocan, brevican exists in vivo either as a full-length proteoglycan or as a proteolytically processed form lacking the GAG-binding region and the N-terminal domain I. Proteolytic cleavage of rat brevican occurs at the same site as bovine brevican within an amino acid sequence that is similar to the aggrecanase cleavage site in aggrecan (149); this site is also conserved in mouse brevican (150). Domain III is also organized similarly to the other hyalectans. However, unlike the other members, brevican contains only one EGF-like repeat, which shows high sequence similarity to the second EGF repeat of versican and neurocan (128).

In addition to the secreted species of brevican, an isoform of rat brevican encoded by a shorter 3.3 kb mRNA is bound to the plasma membrane via a GPI anchor (151). Immunochemical and biochemical data demonstrate that both soluble and GPI-anchored forms of rat brevican exist and that the latter derives from alternatively processed transcripts. The glypiation signal is encoded by a DNA segment that is removed as an intron from the 3.6 kb-transcript encoding the secreted (larger) form of brevican (151). The GPI-anchored form contains no EGF, lectin, or CRP motifs but contains a stretch of hydrophobic amino acids resembling the GPI-anchor. When transfected into human epithelial cells, this form of brevican localizes to the cell surface and is cleavable by the PI-specific phospholipase C (151). Interestingly, the nucleotide sequence of the rat GPI-anchored species is nearly identical to a sequence immediately distal to exon 8 of the mouse brevican gene (150), which is followed by a potential polyadenylation signal. Collectively, these data indicate that the GPI-anchored brevican variant derives from lack of splicing at the 3′ end of exon 8.

Genomic Organization and Transcriptional Control

Analysis of the genomic organization of the hyalectan genes indicates that they are modular and have utilized exon shuffling and duplication during evolution to permit progressive refinement of protein function. The versican (CSPG2) gene (136) maps to human chromosome 5q13.2 (152) and to a syntenic region of mouse chromosome 13 (153). The rat (154), mouse (142, 155), and human (156) aggrecan (AGC1) genes have been fully sequenced and assigned to human chromosome 15q26 (157) and mouse chromosome 7 (158), whereas

neurocan (NCAN) and brevican (BCAN) have been mapped to mouse chromosomes 8 (159) and 3 (150), respectively. This diverse chromosomal location not only suggests an early divergence of the hyalectan genes during evolution but also indicates a significant evolutionary pressure to maintain the overall structure of these modular proteoglycans.

A salient feature of the hyalectans is a remarkable conservation of the exon/intron junctions. For example, the hyaluronan-binding region in these genes is encoded by four exons with identical conservation of the exon size and exon/intron phases (9). All of the introns flanking the individual modules are phase I introns, thereby allowing alternative splicing to occur without altering the reading frame. A typical example is the splicing of the GAG-α and GAG-β modules of versican. Alternate exon usage is indeed a widely used mechanism for increasing coding diversity within genes coding for extracellular matrix proteins (160, 161).

Another intriguing property of versican and aggrecan is the large size of the spliced exons encoding the central domain. They vary from 3 to 5.3 kb. This observation is in contrast to the average size of exons ($\sim$150 bp) for a variety of genes and suggests that these genes have evolved a way to bypass the rules of exon definition. In neurocan, the central nonhomologous domain is encoded by two distinct exons of relatively smaller dimension ($\sim$0.6 and 1.2 kb) (159). An established notion is that intervening sequences function by limiting amplification of genomic DNA that needs to maintain a constant size. Thus genes that code for proteins in which the strict dimension and copies of the repeats are not demanded are likely to harbor large exons. A potential advantage for large exons is an improved RNA processing (161).

Another striking conservation of the exon/intron organization occurs in the 3′ end of the hyalectan genes. The selectin region is encoded by six exons with identical size and phasing in versican, aggrecan, and neurocan. This arrangement should allow alternative splicing of the EGF and CRP repeats. However, splicing events of this region have been observed only in aggrecan (162). Whether this specialized splicing within the selectin domain has any biological significance awaits further experimentation. Notably, murine brevican is the first hyalectan where an intron bordering a domain module is not a phase I intron, thereby preventing the alternative splicing of the CRP domain (150).

The promoter sequences of the human versican (136), rat and mouse aggrecan (142, 154), and mouse neurocan (159) and brevican (150) genes have been sequenced. However, functional studies are available for only the human versican and rat aggrecan promoters. The human versican promoter is active in both squamous carcinoma cells and in embryonic lung fibroblasts (136). It contains a TATA box and numerous cognate *cis*-acting elements that could drive the tissue-specific expression of versican. For example, a cluster of AP2-binding

sites acts as an enhancer element since the presence of AP2 markedly increases the level of versican promoter activity (61). AP2 is expressed in the neural crest and its derivatives, and versican, and perhaps other members of the hyalectan gene family, may be regulated by AP2 in the central nervous system. Also present in the human versican promoter are binding sites for CCAAT-binding transcription factor, Sp1, CEBP, and several cAMP-responsive elements. Of note, a CCAAT-binding transcription factor has been previously involved in TGF-β-induced upregulation of type I plasminogen activator (66), and it may mediate the established upregulation of versican by TGF-β (163).

In contrast to versican, both the rat (154) and mouse (142) aggrecan promoters lack a TATA box and display multiple transcription start sites. They are relatively enriched in G + C, similar to many housekeeping and growth factor–encoding genes. Accordingly, several Sp1-binding sites are scattered throughout the promoter. There are, however, some similarities with versican insofar as the proximal promoter region of rat aggrecan also harbors several AP2-binding sites. There are conserved sequences common to link protein, aggrecan, and type II collagen (142, 154). These sequences include a potential binding site for NF-κB, a factor known to interact with various cytokines that may also play a role in cartilage differentiation. Since these genes are expressed primarily in cartilaginous tissues, these regions may universally mediate chondrocyte gene expression.

The neurocan putative promoter region contains a TATA box and several cognate *cis*-acting elements commonly found in the other hyalectan genes (159) including several Sp1, AP2, AP1, and glucocorticoid-responsive elements. Partial characterization of the 5′ flanking region of mouse brevican shows no TATA or CAAT boxes but reveals a relatively high G + C content, numerous transcription start sites, and several *cis*-acting elements potentially involved in the regulation of neural gene expression (150), consistent with its restricted tissue distribution.

Expression and Functional Roles

Based on the combination of the structural domains summarized above, the hyalectans' major functional roles would be to bind complex carbohydrates such as hyaluronan at their N termini and less complex sugars at their C termini. Less clear is the role of the central nonhomologous regions insofar as they can be substituted with as few as 2 or as many as 100 glycosaminoglyan chains. Via multiple isoforms generated by alternative splicing, these central domains would provide a means to introduce glycosaminoglycans into various extracellular matrices. The different spliced variants of hyalectans are often expressed in distinct spatial and temporal patterns. For example, the V0 and V1 variants of versican are present in fibroblasts, chondrocytes, hepatocytes, and smooth

muscle cells of the aorta and myometrium (135). In contrast, keratinocytes express only versican V1.

Chondroitin sulfate proteoglycans have been implicated in the regulation of cell migration and pattern formation in the developing peripheral nervous system. Versican is selectively expressed in embryonic tissues that act as barriers to neural crest cell migration and axonal outgrowth (164). Versican interferes with the attachment of embryonic fibroblasts to various substrata, including fibronectin, laminin, and collagen. Therefore, the expression of versican within barrier tissues may be linked to guidance of migratory neural crest cells and outgrowing axons (164). In contrast, a report suggests that aggrecan, but not versican, inhibits avian neural crest cell migration (165). In muscular arteries versican-specific epitopes are restricted to the tunica adventitia (166), but in the media and the split elastic interna of atherosclerotic lesions, versican is prominent. The latter observation supports a role for versican in the development of atherosclerotic lesions (167). Of relevance, matrilysin, a matrix metalloproteinase, is expressed by macrophages at sites of potential rupture in atherosclerotic lesions and specifically degrades versican (168). Versican has also been implicated in retaining hyaluronan in mouse cumulus cell-oocyte complexes (169) and could play a role in hair follicle development and cycling (170).

Versican is upregulated in smooth muscle cells treated with platelet-derived growth factor (PDGF) and TGF-β (163, 171, 172). The PDGF-mediated induction of versican gene expression is abrogated by genistein, a broad inhibitor of tyrosine kinase activity (173). Versican can also be induced by a cocktail of growth factors including EGF, TGF-β, and PDGF (174), suggesting that there are independent and synergistic signal-transducing pathways regulating its expression. Abnormal versican expression has been demonstrated in human colon cancer where it occurs as a consequence of hypomethylation of its control genomic regions (175, 176), and enhanced deposits of versican have been observed in the stroma of various tumors, particularly in hyaluronan-rich regions (177, 178). These findings support the observation that abrogation of versican expression by stable antisense transfection can revert the malignant phenotype (179). A link between versican expression and cellular growth is provided by the observation that ectopic expression of the retinoblastoma (RB) protein, a potent tumor suppressor, in RB-negative mammary carcinoma cells induces a relatively small number of genes including the two proteoglycans versican and decorin (180). The RB-induced versican transcript was approximately 5 kb, a size compatible with a selective upregulation of the V2 splice variant containing the GAG-α module (136, 181).

The functional properties of aggrecan reside in the two structural features summarized above: the high concentration of chondroitin sulfate side chains

and the formation of large supramolecular aggregates with hyaluronan (8). In cartilage, each aggrecan monomer occupies a large hydrodynamic volume, and when subjected to compressive forces, water is displaced from individual monomers. This swelling of the tissue is dissipated readily when the compressive forces are removed and the water molecules are siphoned back into the tissue. The function of aggrecan in cartilage development and homeostasis is demonstrated by the phenotype of two mutant animals—the nanomelic chicken (182, 183) and the cartilage matrix-deficient (*cmd*) mouse (158). Nanomelia, a recessively inherited connective tissue disorder of the chicken in which very low levels of aggrecan mRNA are found, affects cartilage formation. A single base mutation leads to a premature truncation of the protein core, which lacks the C-terminal globular (G3) domain, and the truncated precursor accumulates in the endoplasmic reticulum (182). In agreement with these findings, aggrecan mRNA is detectable in the nuclei of nanomelic chondrocytes but is greatly reduced in the cytoplasm (183).

In the *cmd* mice, a 7-bp deletion in exon 5 of the aggrecan gene has been discovered (158). The mutation occurs in the G1 domain and leads to a termination codon within exon 6, resulting in a truncated polypeptide of approximately 36 kDa. Although heterozygous *cmd* mice appear normal, the homozygous mice die soon after birth due to respiratory failure. Both the nanomelic chicken and the *cmd* mouse are characterized by shortened limbs attesting to the importance of aggrecan in the space-filling role in cartilage.

The functional roles of the two brain proteoglycans, neurocan and brevican, are less well understood. Neurocan binds with a relatively high affinity ($K_d \sim 6$ nM) to the neural cell adhesion molecules Ng-CAM and N-CAM, inhibits their homophilic interactions, and blocks neurite outgrowth (184, 185). Moreover, neurocan interacts with tenascin (186) and axonin-1 (187). Some of these interactions may be confined to restricted areas or to a relatively brief developmental stage, and the multiplicity of potential ligands may provide a mechanism for fine tuning of various regulatory processes of neurocan (187).

Neurocan is developmentally regulated (188, 189), and astrocytes, in addition to neurons, may be a cellular source of neurocan in the brain (190). In situ hybridization and immunohistochemical studies have demonstrated that neurocan transcripts are distributed widely in pre- and postnatal brain but not in other organs (191). The adult form of neurocan is generated by a developmentally regulated proteolytic processing of the larger species predominant in the early postnatal brain (146). The adult protein core, mostly devoid of glycosaminoglycan side chains, is formed by proteolytic cleavage of the C terminus.

Brevican is probably the most abundant hyalectan in the adult brain (128). Two species, a 145-kDa and an 80-kDa band representing the full-length and the N-terminally truncated isoforms of brevican, respectively, represent the

largest fraction of whole brain extracts that binds to DEAE-cellulose columns (128). The content of the 80-kDa species increases in the later stages of development, suggesting that the proteolytic processing of brevican may also be developmentally regulated.

Finally, what is the function of a GPI-anchored form of brevican? The obvious answer is that this isoform may link hyaluronan to the cell surface of a specific subset of neurons, or it may induce vectorial insertion of brevican into finite subdomains of the plasma membrane, since the axonal membranes of neurons are equivalent to the apical membranes of polarized epithelia (128).

SMALL LEUCINE-RICH PROTEOGLYCANS

The family of small leucine-rich proteoglycans (SLRPs) contains at least nine distinct products encoded by separate genes (Table 4). They were previously termed nonaggregating or small dermatan-sulfate proteoglycans because of their inability to interact with hyaluronan or because of their type of glycosaminoglycans, respectively. Based on their protein and genomic organization, three classes of SLRPs can be easily identified (Table 4 and Figure 3). Essentially, they are all characterized by a central domain containing leucine-rich

Table 4 General properties of small leucine-rich proteoglycans

Proteoglycan	Gene	Chromosomal mapping		Protein core (∼kDa)[a]	Glycosaminoglycan type (number)
		Human	Mouse		
Class I					
Decorin	DCN	12q23	10	40	Dermatan/chondroitin sulfate (1)[b]
Biglycan	BGN	Xq28	X	40	Dermatan/chondroitin sulfate (2)
Class II					
Fibromodulin	FMOD	1q32		42	Keratan sulfate (2–3)
Lumican	LUM	12q21.3–22	10	38	Keratan sulfate (3–4)
Keratocan				38	Keratan sulfate (3–5)
PRELP[c]	PRELP	1q32		44	Keratan sulfate (2–3)
Osteodherin				42	Keratan sulfate (2–3)
Class III					
Epiphycan	DSPG3	12q21		35	Dermatan/chondroitin sulfate (2–3)
Osteoglycin	OG			35	Keratan sulfate (2–3)

[a]The size of individual protein core does not include any posttranslational modification.

[b]Adult chicken cornea contains decorin with keratan sulfate side chains. Also in avian decorin there is the possibility of two glycosaminoglycan side chains.

[c]The inclusion of PRELP in this category is still preliminary because in most cases PRELP appears as a glycoprotein rather than a proteoglycan.

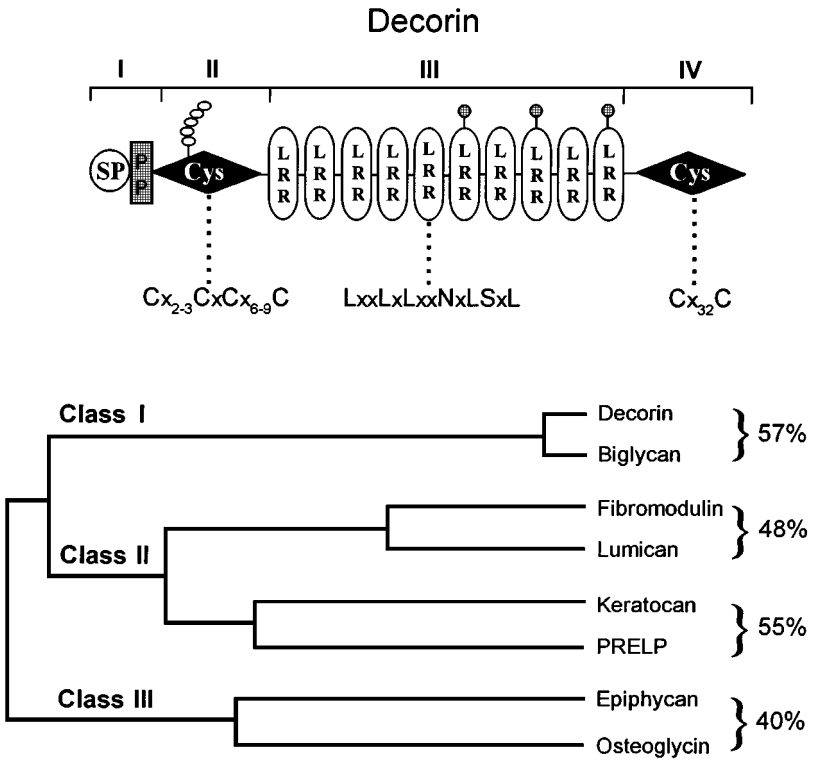


Figure 3 Schematic representation of the structural features of decorin, the prototype member of the small leucine-rich proteoglycans (*top*) and sequence-based evolutionary tree (*bottom*) of the various members of this gene family. The Roman numerals at the top indicate the proposed domains of decorin. The symbols and designations are as follows: SP, signal peptide; PP, propeptide; Cys, cysteine-rich region; LRR, leucine-rich repeat. The consensus sequences for the N-terminal and C-terminal cysteine-rich regions, as well for the leucine-rich repeats are also shown. The glycosaminoglycan side chain and potential N-linked oligosaccharides are indicated by strings of small circles or by a single circle, respectively. In the dendrogram obtained with the program CLUSTAL, branch lengths (*horizontal lines*) are proportional to evolutionary distances. The percent identity in amino acid sequence for each pair of related proteoglycans is indicated in the right margins.

repeats (LRR) flanked at either side by small cysteine-clusters. The prototype member, decorin, is shown in Figure 3 (*top*).

General Structural Features: Three Distinct Classes

Analysis by multiple sequence alignments identifies three classes of SLRPs and two subfamilies. Class I comprises decorin and biglycan, which show the highest homology (57% identity). Class II includes fibromodulin (192), lumican (193), keratocan (194), and PRELP (195). Two subclasses can also

be detected (Figure 3): Fibromodulin and lumican comprise the first subclass (48% identity), and keratocan and PRELP comprise the second subfamily (55% identity). The inclusion of PRELP as a Class II proteoglycan is still preliminary because there is only weak evidence that PRELP is indeed a proteoglycan. In most tissues, PRELP is a glycoprotein. Recently, a novel member of the Class II SLRP gene family has been cloned and named osteoadherin (RD Heinegård, personal communication). As in the case of the other members of Class II, osteoadherin is substituted with keratan sulfate and contains tyrosine sulfate in its N-terminal end. However, osteoadherin has a larger C-terminal extension and is primarily expressed in bone.

Class III comprises epiphycan and osteoglycin. Epiphycan (196, 197), also known as PG-Lb, derives its name from its highly selected expression in epiphyseal cartilage. Based on the dendrogram shown in Figure 3, Class III proteoglycans diverged from the common precursors of Class I and II before the presumed gene duplication and independent evolution of the latter two subfamilies.

The overall structural characteristics of the SLRPs are similar although there are some interesting variations. In the prototype SLRP decorin, four domains can be identified: domain I, which contains the signal peptide and a propeptide; domain II, which contains four evenly spaced cysteine residues and the glycosaminoglycan attachment site; domain III, which contains the LRRs; and domain IV, which contains a relatively large loop with two cysteine residues. Interestingly, only decorin and biglycan contain a propeptide (198, 199). While the signal peptide targets the nascent core protein to the rough endoplasmic reticulum, less clear is the function of the propeptide. This sequence is highly conserved across species (200), and the propeptide may function as a recognition signal for the first enzyme (xylosyltransferase) involved in the biosynthesis of glycosaminoglycans (201).

When constructs containing deletions of the decorin propeptide are transfected into mammalian cells, the secreted proteoglycans are substituted with shorter glycosaminoglycan chains (202). Thus deletion of the propeptide may lower the affinity for xylosyltransferase or may induce a more rapid transition through the Golgi compartment. This situation may also occur in biglycan because a recombinant molecule that lacks the propeptide results in a protein core devoid of glycosaminoglycans (203). The presence of the pro-biglycan species has been documented in keratinocytes (204) and endothelial cells (205). In contrast to Class I, Class II SLRPs appear to be proteolytically processed following removal of their signal peptide and thus are devoid of the propeptide (206, 207).

Domain II is the negatively charged region carrying sulfated glycosaminoglycans in decorin, biglycan, and epiphycan and carrying sulfotyrosine in all of the rest, with the exception of PRELP (10). Decorin and biglycan both have

a region that contains a series of acidic amino acids followed by a consensus GAG attachment domain with potential attachment sites at Ser^4, or Ser^5 and Ser^{11}, respectively (208). Site-directed mutagenesis of the serine residue has demonstrated the requirement of a specific amino acid sequence to direct proper O-glycosylation of decorin (209). Biglycan and decorin can occur as either a monoglycanated (208) or biglycanated (210) species. All Class II and III SLRPs contain consensus sequences for tyrosine sulfation, with the exception of PRELP. This sequence is characterized by at least one tyrosine residue followed by an acidic amino acid (Asp or Glu).

Domain II also contains a cluster of highly conserved cysteine residues with a general consensus $Cx_{2-3}CxCx_{6-9}C$, where x is any amino acid and the subscripts denote the number of intervening residues (Figure 3). Interestingly, within each subclass there is identical spacing of the intervening amino acid residues. For example, in Class I the four cysteine residues are spaced by 3, 1, and 6 residues; in Class II, the spacing is a 3, 1, 9 pattern; whereas Class III follows a 2, 1, 6 pattern (10). Not only is the spacing of the cysteine residues conserved within each class of SLRPs, but also the nature of the intervening amino acid is maintained.

Domain III comprises 10 tandem LRRs, with the exception of epiphycan and osteoglycin, which contain only six repeats. In the LRR consensus sequence, that is, LxxLxLxxNxLSxL, L is leucine, isoleucine, or valine, and S is serine or threonine (10). If the consensus for LRRs is interpreted with less stringency, then there could be two additional LRRs, one in the N-terminal and one in the C-terminal end of the molecule. As discussed below, the major function of these LRRs is to bind and interact with other proteins.

The level of complexity increases if one considers the substitution of the protein core with several N-linked oligosaccharides. For example, human decorin has three potential sites and biglycan has two. The variability in the number and complexity of oligosaccharide chains, which are of both the high-mannose and the complex type, could modulate some of the functions of these gene products. A proposed role of the N-linked oligosaccharides in decorin and biglycan is to retard self-aggregation, thereby favoring interactions with cell surface proteins and/or other extracellular matrix constituents.

In Class II proteoglycans, one to several asparagine residues can be substituted with N-linked keratan sulfate. At least in the case of fibromodulin, all four Asn residues in domain III can be acceptors (211). Moreover, polylactosamine, essentially an unsulfated keratan sulfate, can be found in both fibromodulin (212) and keratocan (194). Lumican and keratocan have a characteristic unlike other SLRP members: In the cornea, they are both keratan sulfate–carrying proteoglycans; in other tissues, these molecules occur as poorly sulfated or unsulfated glycoproteins (194).

The function of the C-terminal domain of the SLRPs is the least characterized so far. It comprises about 50 amino acid residues with considerable similarity among the various members. However, there are several stretches that show a profound divergence. Domain IV contains two cysteines spaced by 32 intervening amino acids, except in keratocan and PRELP, where an insertion of 7 and 8 amino acids, respectively, has occurred. For bovine biglycan, a disulfide bond has been demonstrated in this region (208). Thus a large loop of approximately 34 amino acid residues would be formed at the C end of the SLRPs. Notably, reduction and alkylation of decorin and lumican (213, 214) abolishes their ability to interact with collagen. Thus the disulfide bonding at the C-terminal end may also be crucial in maintaining the fibrillogenesis-controlling activity of SLRPs.

Genomic Organization and Transcriptional Control

The various members of the SLRP gene family map to relatively few chromosomes (Table 4). Decorin, lumican, and epiphycan map to the long arm of chromosome 12 between 12q21 and 12q23 (215–217) and in the corresponding syntenic regions of mouse chromosome 10 (200). In contrast, biglycan maps to Xq28 (218), and fibromodulin and PRELP map to the long arm of chromosome 1 at 1q32 (219). The pairwise similarity at the protein level is also reflected at the genomic level. For example, both members of Class I, decorin and biglycan, are encoded by eight distinct exons with very similar intron/exon boundaries (216, 220), and these features are very well conserved in the mouse (200, 221). A unique feature of the human decorin gene is the presence of alternatively spliced leader exons, Ia and Ib, encoding 5′-untranslated sequences (216). The region corresponding to exon Ib is not found in the mouse (200), suggesting that the decorin gene has undergone significant recombination during evolution (10).

In contrast to Class I, members of Class II SLRPs are encoded by three exons, with all ten LRRs encoded by a single large exon (222, 223). In these genes, the introns are positioned at identical sites, just proximal to the translation initiation and termination codons, respectively (223). There is no published information regarding the genomic organization of class III SLRPs; however, a preliminary study suggests that their intron/exon structure follows a distinct pattern (224), further stressing their belonging to a distinct subfamily.

So far, only the promoter regions of decorin and biglycan have been sequenced and shown to be functionally active (61). Decorin has a complex promoter in the region flanking exon Ib (225). The discovery of two leader exons in the 5′-untranslated region has suggested that a two-promoter system and alternative splicing could be responsible for the presence of heterogeneous transcripts from a single gene (216). However, no functional activity for the region flanking exon Ia was found (225). In contrast, strong basal and inducible

promoter activity was detected using the approximately 1-kb region 5′ to exon Ib (225). This promoter can be divided arbitrarily into two main regions: a proximal promoter of approximately 188 bp and a distal promoter of approximately 800 bp. The proximal promoter region contains two functional TATA boxes and a CAAT box (225). Moreover, it contains two tumor necrosis factor-α (TNF-α)-responsive elements that mediate the binding of TNF-α-induced nuclear proteins and a consequent transcriptional repression of the decorin gene (226). The proximal promoter region contains also a canonical and a functional AP1–binding site, a bimodal regulator of decorin gene expression, which allows both repression by TNF-α and induction by interleukin-1 (IL-1) (227).

These transcriptional data thus provide a molecular mechanism for the previous observations that decorin expression can be induced by either IL-1 (228) or IL-4 (229). The distal promoter of decorin harbors a number of cognate *cis*-acting factors including AP1, AP5, and NF-κB; several direct repeats; and a TGF-β-negative element. The latter element has been found in a variety of proteinases that are involved in the degradation of collagens and proteoglycans during remodeling and suggests a link between suppression of matrix-degrading enzymes and downregulation of decorin gene expression (230). The distal promoter also contains a long stretch of homopurine/homopyrimidine residues. When contained in a supercoiled plasmid, this sequence is sensitive to endonuclease S1, an enzyme that digests preferentially single-stranded DNA. Moreover, this sequence can upregulate a minimal heterologous promoter (225). Thus this region may adopt an intramolecular hairpin triplex and may regulate in vivo the transcription of decorin.

In contrast to decorin, the promoter of the human biglycan gene contains neither a TATA nor a CAAT box but is G + C rich (220) with a 66% overall G + C content and two clusters of G + C that reach 73 and 87%, respectively. These features have been conserved in the mouse promoter, which contains an overall G + C content of approximately 60% (221). The human biglycan promoter contains numerous Sp1-binding sites and several transcription initiation sites. In contrast to decorin, the biglycan promoter is highly conserved in the mouse, particularly in the proximal region where two AP2 and two Sp1-binding sites are perfectly maintained (221). In both the human and mouse biglycan promoters there are numerous motifs that could potentially bind members of the *Ets* family of oncogenes (PU-boxes and PEA3 motifs). These factors can transcriptionally activate B cells and macrophages (231).

The human promoter contains at least five IL-6-responsive elements, TNF-α-responsive elements similar to those found in the decorin gene promoter, and a binding site for the liver-specific transcription factor C/EBP (232). The latter is interesting because it may explain why biglycan expression is induced in the liver during the transformation of the fat-storing cells into myofibroblast-like

cells (233). Biglycan can be transcriptionally induced in MG-63 osteosarcoma cells by forskolin and 8-Bromo-cAMP (234), and the biglycan promoter is upregulated by exogenous IL-6 and downregulated by TNF-α (232). That the putative binding sites for the corresponding cytokines are functionally active is demonstrated in human endothelial cells whose biglycan expression is downregulated by TNF-α (235).

Decorin is generally downregulated by TGF-β, whereas biglycan is upregulated in various cells and organisms (61, 163, 236–239). In contrast, dexamethasone increases decorin production and also prevents the TGF-β-elicited downregulation of decorin gene expression (240). In contrast to decorin where TNF-α and TGF-β are additive in their inhibitory action, TNF-α counteracts the effects of TGF-β and IL-6 in controlling biglycan gene expression. Thus these two important members of the SLRP gene family can be diversely regulated at the site of injury by a finely balanced release of active cytokines. A unique feature of the biglycan gene, being located on the X chromosome, is that it should follow the rules of X chromosomal inactivation. However, biglycan does not follow the expected conventional correlation between gene dosage and expression rate (241). That is, biglycan behaves as a pseudoautosomal gene even though in somatic cell hybrid experiments, biglycan undergoes X chromosomal inactivation. These data suggest that there may be an additional regulatory gene(s) that controls the transcriptional activity of biglycan (241).

Control of Collagen Fibrillogenesis

Collagen fibril formation is a self-assembly process that has been investigated in vitro for over four decades. Although it is clear from fibril reconstitution experiments that information to build periodic fibrils resides in the amino acid sequence of the fibril-forming collagens, other molecules have been identified that modulate the assembly process (242). Both the kinetics of assembly and the ultimate fibril diameters are modulated by these factors, and both acceleration and inhibition of fibril formation have been reported.

After the discovery that the interaction of dermatan sulfate proteoglycans with collagen causes increased stability of collagen fibrils and a change in their solubility (243, 244), it was demonstrated that various members of the SLRP gene family interact directly with fibrillar collagens (245). The orthogonal position of the proteoglycans would facilitate proper spacing of the collagen fibrils during axial growth or perhaps during lateral fusion (see below). At least three members of the SLRP gene family, namely decorin (246), fibromodulin (247, 248), and lumican (214), can delay fibril formation in a dynamic fibrillogenesis assay. Neither removal of the glycosaminoglycan chain nor the 17-amino acid N-terminal portion of the decorin protein core significantly alters collagen fibrillogenesis (249), indicating that the decorin/collagen interactions require the remaining

protein core (250). Corneal and scleral SLRPs also retard fibrillogenesis in vitro and reduce the size of the formed fibrils (251), in agreement with data generated from analysis of developing avian tendon where a decrease in fibril-associated decorin is necessary for fibril growth during tissue maturation (252).

The three-dimensional (3D) model of human decorin (253) and the ultrastructural observation that several SLRPs are horseshoe shaped (254) predict a close interaction between decorin and fibrillar collagen. This close interaction would help stabilize fibrils and orient fibrillogenesis. The 3D model of decorin, based on the crystal structure of the porcine ribonuclease inhibitor (255), predicts an arch-shaped structure with the inner concave surface lined by β-strands and the outer convex surface formed by α-helices. This model would allow easy access of interactive proteins to the inner surface. Another interesting feature of the decorin model is that the three N-linked oligosaccharides and the single glycosaminoglycan chain at Ser^7 are all positioned on one side of the arch-shaped structure, and the glycosaminoglycan chain is relatively free to project away from the protein core. This arrangement would also account for the proposed function of the glycosaminoglycan chain to maintain interfibrillary space (256). The inner surface of the arch-shaped decorin molecule contains a series of charged residues that would complement the charges in a corresponding amino acid stretch within the collagen type I triple helix, the proposed binding site for decorin (257).

Of course, the data generated with the decorin model should be interpreted cautiously because the homology between ribonuclease inhibitor and decorin is relatively low, particularly at the N- and C-terminal ends. However, even if the protein is folding quite differently in these regions, it has been shown that the central domain, which should fold as the ribonuclease inhibitor does, is the essential part for modulating collagen fibrillogenesis (258–260).

Genetic evidence for a role of decorin in maintaining collagen fibrillogenesis has been provided by the phenotype of decorin null animals. The mice carrying a homozygous disruption of the decorin gene grow normally to adulthood; however, they manifest a phenotype characterized by increased skin fragility (261). When samples of skin from the wild-type and decorin null animals were subjected to biomechanical testing, the latter samples exhibited a markedly reduced tensile strength that could be associated with an abnormal collagen fiber formation. Ultrastructural analysis of skin revealed bizarre and irregular collagen morphology with coarser and irregular fiber outlines in the decorin null specimens (Figure 4). Although the mean cross-sectional diameter of the fibers did not vary significantly between the wild-type and decorin null animals, the latter showed a markedly increased range with profiles varying between 40 and 260 nm. Scanning-transmission electron microscopy of isolated collagen fibers revealed that the wide variation in range was not the result of multiple

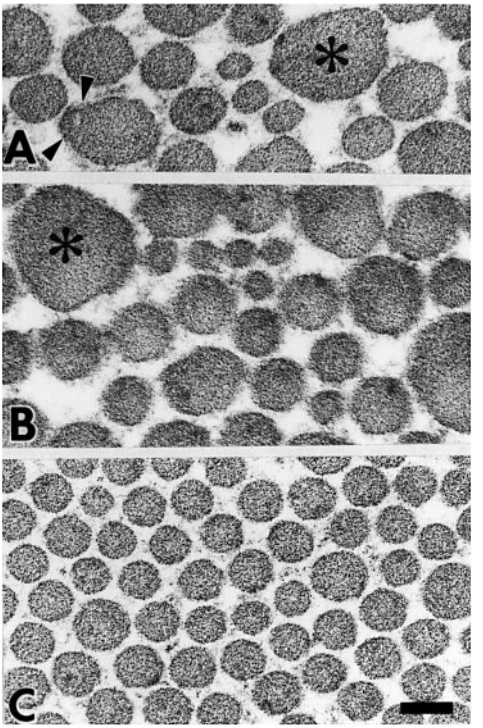


Figure 4 Ultrastructural appearance of dermal collagen from the skin of decorin null (*A* and *B*) and wild-type (*C*) mice. Notice the larger and irregular cross-sectional profiles in the decorin null collagen fibers (*asterisks*) with evidence of lateral fusion (*A, arrowheads*). Bar: 90 nm.

populations of fibrils of different diameter but rather was due to a single population of fibrils with markedly irregular profiles along their axes (261). These studies establish a functional role for decorin in maintaining the structural integrity of the cutis. While decorin may affect other functions (see below), this genetic evidence unequivocally demonstrates that the tensile strength of skin is under the control of so-called modifiers of collagen fibrillogenesis and that decorin is an important member of these modifier proteins.

Interactions with TGF-β and the Control of Cell Proliferation and Corneal Transparency

Increased TGF-β production is a hallmark of a variety of fibrotic states, including cirrhosis, pulmonary fibrosis, and glomerular sclerosis. One of the most important characteristics of decorin is a high affinity for TGF-β (262). TGF-β1, -β2, and -β3 isoforms all bind to the decorin core protein with similar efficiency

(262), allowing decorin to function as a reservoir for these growth factors in the extracellular milieu. This line of research derives from the observation that Chinese hamster ovary (CHO) cells, which do not constitutively synthesize decorin, are growth suppressed when decorin is ectopically expressed (263). Investigators later showed that this growth inhibition was caused by a blocking of TGF-β-activity by decorin (264). Because CHO cells require TGF-β for their growth, it was concluded that decorin-induced growth inhibition is secondary to the blockage of this growth factor. Administration of recombinant decorin or decorin purified from bovine tissues prevents glomerular sclerosis in a rat model of glomerulonephritis (265). The lesions caused by this kind of glomerular injury can be ameliorated by gene therapy with a decorin-expressing vector directly transfected into skeletal muscle (266).

Although this evidence implicates decorin in blocking the action of TGF-β, controversy remains as to whether decorin is a universal TGF-β inhibitor. For example, in quiescent fibroblasts that produce up to 40 times more decorin than cells in the logarithmic phase of growth, nanomolar amounts of TGF-β are still fully active, that is, they can induce the endogenous expression of biglycan (226). Intuitively, the amount of decorin in the medium would have been sufficient to block the exogenous TGF-β. One possibility is that decorin binds avidly to fibrillar collagen, thereby preventing further interaction with TGF-β. Moreover, in the presence of 10,000-fold molar excess of exogenous decorin, TGF-β effects could still be detected in human monocytic cells (267). The addition of decorin to osteoblastic cells enhances the binding of radiolabeled TGF-β to its receptors (268).

Therefore, it appears that in certain cellular systems, decorin blocks the activity of TGF-β, whereas in others its binding augments the bioactivity of the cytokine. How can we reconcile these seemingly conflicting observations? One possibility is that the decorin/TGF-β complexes may still be capable of interacting with at least one signal-transducing pathway of TGF-β signaling. That is, under certain conditions, these complexes may activate rather than repress the cytokine activity, a phenomenon that would be cell specific.

An emerging function of decorin is its ability to inhibit cellular proliferation. When a full-length cDNA driven by the potent cytomegalovirus promoter is introduced into colon carcinoma cells (which do not synthesize decorin), the cells become quiescent, form small colonies in soft agar, and do not generate tumors in immunocompromised hosts (269). All of these effects are independent of TGF-β. Interestingly, a number of clones are arrested in the G_1 phase of the cell cycle, and their growth suppression can be restored by treatment with decorin antisense oligodeoxynucleotides. The decorin-induced growth arrest is associated with an induction of p21, a potent inhibitor of cyclin-dependent kinase activity (270). Ectopic expression of decorin proteoglycan or protein core, a mutated form lacking any glycosaminoglycan side chains,

induces growth suppression in neoplastic cells of various histogenetic origins (271). Even when recombinant decorin is added to tumor cells, all react by slowing their proliferative status and by inducing p21 (271). Thus it appears that decorin is an important inhibitor of growth that can act directly on a signal-transduction pathway that leads to activation of cyclin-dependent kinase inhibitors and ultimately to arrest in G_1. Decorin causes rapid phosphorylation of the EGF receptor and a concurrent activation of the mitogen-activated protein (MAP) kinase signal pathway (271a). This leads to a protracted induction of endogenous p21 and ultimate cell cycle arrest. Moreover, recombinant decorin causes a rapid increase in intracellular Ca^{2+} levels in A431 cells (271b). The effects of decorin persist in the absence of extracellular Ca^{2+} but are blocked by AG1478, an EGF-specific tyrosine kinase inhibitor and by downregulation of the EGF receptor. These results indicate a novel action of decorin on the EGF receptor, which results in mobilization of intracellular Ca^{2+} and activation of a signal-transducing pathway that culminates in growth suppression by blocking the cell cycle machinery. The discovery that two tyrosine kinase orphan receptors, the discoidin domain receptors DDR1 and DDR2, are the receptors for fibrillar collagen (Types I, III, and V) opens a novel perspective in extracellular matrix research (271c, 271d). Both fibrillar collagen and decorin converge on similar tyrosine kinase receptors, and their activation may allow a specific cross talk betwen cells and the extracellular matrix.

The ability of secreted decorin to induce growth suppression gives further support to the concept that abnormal production of this proteoglycan around invading carcinomas represents a specialized biological response of the host designed to counterbalance the invading tumor cells (176, 272, 273).

Decorin, as well as other SLRP members, may also modulate the remodeling of the extracellular matrix since decorin can induce collagenase (274). When vascular endothelial cells, which usually synthesize neither type I collagen nor decorin, initiate the formation of tubes or cords, they begin to synthesize both molecules (275). In contrast, when endothelial cells are wounded in vitro, biglycan is induced at the edge of the migrating endothelial cells, a process that is apparently mediated by release of endogenous FGF-2 (276). Thus the same theme emerges, with decorin and biglycan, though structurally related, serving diverse functions when cells are migrating to form new blood vessels. The high-affinity interaction of decorin and biglycan with important mediators of the inflammatory process, such as the C1q component of the C1 complex (277, 278), and the affinity of the other SLRP members for a variety of extracellular matrix constituents suggest that these molecules play a primary role in repair processes.

Biglycan has also been directly implicated in regulating hemopoiesis. For example, biglycan was identified as a potent factor stimulating monocytic activity from thymic myoid cells (279) and as one of several products that have

affinity for precursors of B lymphocytes (280). Notably, while biglycan is incapable of stimulating the growth of myeloid cells, it markedly enhances the cloning of IL-7-responsive precursors (280).

An important function of members of the SLRP gene family is related to the maintenance of corneal transparency. In cornea, the diameter of the collagen fibrils and the interfibrillary gaps must be kept constant to prevent corneal opacity (256). Lateral growth of collagen segments isolated from bovine (214) or avian (281) corneas can be retarded by addition of decorin or lumican extracted from their respective corneal stromas. Moreover, treatment of the developing avian cornea with β-D-xyloside, an inhibitor of glycosaminoglycan attachment (which presumably affects only dermatan/chondroitin sulfate glycanation but not the addition of keratan sulfate) does not alter collagen fibril diameter (282).

The developmental changes of lumican, primarily those occurring in the first two weeks of corneal development, are fundamental to the maintenance of corneal transparency (283). In fact, the fully sulfated species of lumican is not abundant before day 15 of avian corneal development. Before that time, there is a progressive accumulation of lumican substituted with polylactosamine, suggesting that its subsequent sulfation could play a role during the acquisition of corneal transparency (283). This notion is also supported by the finding that in macular corneal dystrophy, the polylactosamine form of lumican is the primary species (284). A recent study using synchrotron X-ray diffraction has demonstrated a unique 4.6-C periodicity in macular corneal dystrophy (285). After observing digestion of proteoglycans with various polysaccharide lyases, researchers concluded that dermatan sulfate or keratan sulfate proteoglycan hybrids would be also present in this disease.

PERSPECTIVES

Our appreciation of the molecular design of the matrix proteoglycans is leading to a better understanding of their cellular functions. The application of modern gene technology in this field has been slower than in others, but initial successes are now being reported. Current work aims at understanding the mechanisms that govern proteoglycan diversity, especially in regard to tissue-specific constraints that favor the expression of specialized variants. Contemporary areas of active research also focus on clarifying the mechanisms that regulate the generation of proteoglycan isoforms—containing a specific set of glycosaminoglycans or a proteolytically processed species—and of biologically active growth factor/proteoglycan complexes.

Some of the major tasks that lie ahead relate to the delineation of the primary mechanisms that trigger a specific cellular response to various proteoglycans,

and the determinants of cellular responsiveness to them. Ultimately, proteoglycan-dependent induction of transcription factors and their modulation will be an area of vibrant research. How do proteoglyans induce so many pleiotropic responses? How do diverse proteoglycans activate specific signal-transducing pathways? What is the molecular cross-talk between a given proteoglycan family and the transcriptional machinery? How do cells regulate their complex biosynthetic pathways?

In part, some of these questions will be answered by the analysis of mutant animals, whereas others will be answered by resolving the various biochemical steps of proteoglycan assembly and metabolism. The correlation of structure and function may soon be possible for several actively investigated members of the matrix proteoglycans by using modern genetic approaches that involve site-directed mutagenesis, transfection into appropriate recipients, and analysis of their functional consequences. One can anticipate continued rapid progress and await with anticipation the further advances in this exciting area of research.

ACKNOWLEDGMENTS

I am grateful to many colleagues for helpful discussion and apologize to those whose work was not referenced because of space limitations. I thank J Esko for his friendship and generosity; W Halfter, G Cole, D Heinegård, J Gallagher, and U Rauch for sharing data before publication; CC Clark for critical reading of this review; and I Eichstetter for help with the illustrations. The original work was supported by National Institutes of Health grants 5RO1 CA39481-14 and 5RO1 CA47282-08.

Literature Cited

1. Gallagher JT. 1989. *Curr. Opin. Cell Biol.* 1201–18
2. Kjellén L, Lindahl U. 1991. *Annu. Rev. Biochem.* 60:443–75
3. Esko JD. 1991. *Curr. Opin. Cell Biol.* 3: 805–16
4. Bernfield M, Kokenyesi R, Kato M, Hinkes MT, Spring J, et al. 1992. *Annu. Rev. Cell Biol.* 8:365–93
5. Couchman JR, Woods A. 1993. In *Cell Surface and Extracellular Glycoconjugates*, ed. DD Roberts, RP Mecham, pp. 33–81. San Diego: Academic
6. Humphries DE, Stevens RL. 1992. In *Heparin and Related Polysaccharides*, ed. DA Lane, I Bjork, U Lindahl, pp. 59–67. New York: Plenum
7. Elenius K, Jalkanen M. 1994. *J. Cell Sci.* 107:2975–82
8. Roughley PJ, Lee ER. 1994. *Micro. Res. Tech.* 28:385–97
9. Iozzo RV, Murdoch AD. 1996. *FASEB J.* 10:598–614
10. Iozzo RV. 1997. *Crit. Rev. Biochem. Mol. Biol.* 32:141–74
11. Rosenberg RD, Shworak NW, Liu J, Schwartz JJ, Zhang L. 1997. *J. Clin. Invest.* 99:2062–70

12. Noonan DM, Fulle A, Valente P, Cai S, Horigan E, et al. 1991. *J. Biol. Chem.* 266:22939–47
13. Tsen G, Halfter W, Kröger S, Cole GJ. 1995. *J. Biol. Chem.* 270:3392–99
14. Wu R-R, Couchman JR. 1997. *J. Cell Biol.* 136:433–44
15. Noonan DM, Hassell JR. 1993. *Kidney Int.* 43:53–60
16. Murdoch AD, Iozzo RV. 1993. *Virchows Arch. A* 423:237–42
17. Timpl R. 1993. *Experientia* 49:417–28
18. Iozzo RV. 1994. *Matrix Biol.* 14:203–8
19. Iozzo RV, Cohen IR, Grässel S, Murdoch AD. 1994. *Biochem. J.* 302:625–39
20. Kokenyesi R, Silbert JE. 1995. *Biochem. Biophys. Res. Commun.* 211:262–67
21. Groffen AJA, Buskens CAF, Tryggvason K, Veerkamp JH, Monnens LAH, et al. 1996. *Eur. J. Biochem.* 241:827–34
22. Dolan M, Horchar T, Rigatti B, Hassell JR. 1997. *J. Biol. Chem.* 272:4316–22
23. Costell M, Mann K, Yamada Y, Timpl R. 1997. *Eur. J. Biochem.* 243:115–21
24. Bork P, Patthy L. 1995. *Protein Sci.* 4: 1421–25
25. Costell M, Sasaki T, Mann K, Yamada Y, Timpl R. 1996. *FEBS Lett.* 396:127–31
26. Battaglia C, Mayer U, Aumailley M, Timpl R. 1992. *Eur. J. Biochem.* 208: 359–66
27. Reinhardt D, Mann K, Nischt R, Fox JW, Chu M-L, et al. 1993. *J. Biol. Chem.* 268:10881–87
28. Timpl R, Brown JC. 1994. *Matrix Biol.* 14:275–81
29. Schulze B, Mann K, Battistutta R, Wiedemann H, Timpl R. 1995. *Eur. J. Biochem.* 231:551–56
30. Schulze B, Sasaki T, Costell M, Mann K, Timpl R. 1996. *Matrix Biol.* 15:349–57
31. Chakravarti S, Horchar T, Jefferson B, Laurie GW, Hassell JR. 1995. *J. Biol. Chem.* 270:404–9
32. Kallunki P, Tryggvason K. 1992. *J. Cell Biol.* 116:559–71
33. Murdoch AD, Dodge GR, Cohen I, Tuan RS, Iozzo RV. 1992. *J. Biol. Chem.* 267: 8544–57
34. Yurchenco PD, Cheng Y-S, Ruben GC. 1987. *J. Biol. Chem.* 262:17668–76
35. Hunter DD, Porter BE, Bulock JW, Adams SP, Merlie JP, et al. 1989. *Cell* 59:905–13
36. Iozzo RV, Kovalszky I, Hacobian N, Schick PK, Ellingson JS, et al. 1990. *J. Biol. Chem.* 265:19980–89
37. Danielson KG, Martinez-Hernandez A, Hassell JR, Iozzo RV. 1992. *Matrix* 11: 22–35
38. SundarRaj N, Fite D, Ledbetter S, Chakravarti L, Hassell JR. 1995. *J. Cell Sci.* 108:2601–4
39. Couchman JR, Kapoor R, Sthanam M, Wu R-R. 1996. *J. Biol. Chem.* 271:9595–602
40. Iozzo RV. 1989. *J. Biol. Chem.* 264: 2690–99
41. Iozzo RV, Hassell JR. 1989. *Arch. Biochem. Biophys.* 269:239–49
42. Denzer AJ, Gesemann M, Schumacher B, Ruegg MA. 1995. *J. Cell Biol.* 131: 1547–60
43. Hagen SG, Michael AF, Butkowski RJ. 1993. *J. Biol. Chem.* 268:7261–69
44. Denzer AJ, Brandenberger R, Gesemann M, Chiquet M, Ruegg MA. 1997. *J. Cell Biol.* 137:671–83
45. Tsen G, Napier A, Halfter W, Cole GJ. 1995. *J. Biol. Chem.* 270:15934–37
46. Denzer AJ, Gesemann M, Ruegg MA. 1996. *Semin. Neurosci.* 8:357–66
47. O'Toole JJ, Deyst KA, Bowe MA, Nastuk MA, McKechnie BA, et al. 1996. *Proc. Natl. Acad. Sci. USA* 93:7369–74
48. Iozzo RV, Clark CC. 1986. *J. Biol. Chem.* 261:6658–69
49. Iozzo RV, Clark CC. 1987. *Histochemistry* 88:23–29
50. Cohen IR, Grässel S, Murdoch AD, Iozzo RV. 1993. *Proc. Natl. Acad. Sci. USA* 90:10404–8
51. Rupp F, Özçelik T, Linial M, Peterson K, Francke U, et al. 1992. *J. Neurosci.* 12:3535–44
52. Dodge GR, Kovalszky I, Chu M-L, Hassell JR, McBride OW, et al. 1991. *Genomics* 10:673–80
53. Wintle RF, Kisilevsky R, Noonan D, Duncan AMV. 1990. *Cytogenet. Cell Genet.* 54:60–61
54. Iozzo RV, Pillarisetti J, Sharma B, Murdoch AD, Danielson KG, et al. 1997. *J. Biol. Chem.* 272:5219–28
55. Mitchell PJ, Tjian R. 1989. *Science* 245: 371–78
56. Piédagnel R, Prié D, Cassingéna R, Ronco PM, Lelong B. 1994. *J. Biol. Chem.* 269:17469–76
57. McBain JA, Pettit GR, Mueller GC. 1990. *Cell Growth Differ.* 1:281–91
58. Grässel S, Cohen IR, Murdoch AD, Eichstetter I, Iozzo RV. 1995. *Mol. Cell. Biochem.* 145:61–68
59. Orkin SH. 1995. *J. Biol. Chem.* 270: 4955–58
60. Murdoch AD, Liu B, Schwarting R, Tuan RS, Iozzo RV. 1994. *J. Histochem. Cytochem.* 42:239–49

61. Iozzo RV, Danielson KG. 1998. *Prog. Nucleic Acids Res. Mol. Biol.* In press
62. Couchman JR, Ljubimov AV, Sthanam M, Horchar T, Hassell JR. 1995. *J. Histochem. Cytochem.* 43:955–63
63. Ritzenthaler JD, Goldstein RH, Fine A, Lichtler A, Rowe DW, et al. 1991. *Biochem. J.* 280:157–62
64. Rossi P, Karsenty G, Roberts AB, Roche NS, Sporn MB, et al. 1988. *Cell* 52:405–14
65. Marigo V, Volpin D, Vitale G, Bressan GM. 1994. *Biochem. Biophys. Res. Commun.* 199:1049–56
66. Riccio A, Pedone PV, Lund LR, Olesen T, Olsen HS, et al. 1992. *Mol. Cell. Biol.* 12:1846–55
67. Courtois SJ, Lafontaine DA, Lemaigre FP, Durviaux SM, Rousseau GG. 1990. *Nucleic Acids Res.* 18:57–64
68. Dodge GR, Kovalszky I, Hassell JR, Iozzo RV. 1990. *J. Biol. Chem.* 265: 18023–29
69. Morris JE, Gaza G, Potter SW. 1994. *Cell. Dev. Biol.* 30A:120–28
70. Chakravarti S, Hassell JR, Phillips SL. 1993. *Dev. Dyn.* 197:107–14
71. Ko CW, Bhandari B, Yee J, Terhune WC, Maldonado R, et al. 1996. *Mol. Cell. Biochem.* 162:65–73
72. Kasinath BS, Grellier P, Ghosh-Choudhury G, Abboud SL. 1996. *J. Cell. Physiol.* 167:131–36
73. Kanwar YS, Rosenzweig LJ, Linker A, Jakubowski ML. 1983. *Proc. Natl. Acad. Sci. USA* 80:2272–75
74. Templeton DM, Fan M-Y. 1996. *Metabolism* 45:1136–46
75. Wahab NA, Harper K, Mason RM. 1996. *Biochem. J.* 316:985–92
76. Kanwar YS, Liu ZZ, Kumar A, Usman MI, Wada J, et al. 1996. *J. Clin. Invest.* 98:2478–88
77. Kanwar YS, Liu ZZ, Wallner EI. 1997. *Lab. Invest.* 76:671–81
78. Dziadek M, Fujiwara S, Paulsson M, Timpl R. 1985. *EMBO J.* 4:905–12
79. Carson DD, Tang J-P, Julian J. 1993. *Dev. Biol.* 155:97–106
80. Smith SE, French MM, Julian J, Paria BC, Key SK, et al. 1997. *Dev. Biol.* 184: 38–47
81. Handler M, Yurchenco PD, Iozzo RV. 1997. *Dev. Dyn.* 210:130–45
82. Friedl A, Chang Z, Tierney A, Rapraeger AC. 1997. *Am. J. Pathol.* 150:1443–55
83. Nugent MA, Karnovsky MJ, Edelman ER. 1993. *Circ. Res.* 73:1051–60
84. Forsten KE, Courant NA, Nugent MA. 1997. *J. Cell. Physiol.* 172:209–20
85. Lyon M, Rushton G, Gallagher JT. 1997. *J. Biol. Chem.* 272:18000–6
86. Stringer SW, Gallagher JT. 1997. *J. Biol. Chem.* 272:20508–14
87. Clowes AW, Karnovsky MJ. 1977. *Nature* 265:625–26
88. Benitz WE, Kelley RT, Anderson CM, Lorant DE, Bernfield M. 1990. *Am. J. Respir. Cell Mol. Biol.* 2:13–24
89. Weiser MCM, Belknap JK, Grieshaber SS, Kinsella MG, Majack RA. 1996. *Matrix Biol.* 15:331–40
90. Joseph SJ, Ford MD, Barth C, Portbury S, Bartlett PF, et al. 1996. *Development* 122:3443–52
91. Weiser MCM, Grieshaber NA, Schwartz PE, Majack RA. 1997. *Mol. Biol. Cell* 8:999–1011
92. Whitelock J, Mitchell S, Underwood PA. 1997. *Cell Biol. Int.* 21:181–89
93. Yayon A, Klagsbrun M, Esko JD, Leder P, Ornitz DM. 1991. *Cell* 64:841–48
94. Rapraeger AC, Krufka A, Olwin BB. 1991. *Science* 22:1705–8
95. Roberts R, Gallagher J, Spooncer E, Allen TD, Bloomfield F, et al. 1988. *Nature* 332:376–78
96. Turnbull JE, Fernig DG, Ke Y, Wilkinson MC, Gallagher JT. 1992. *J. Biol. Chem.* 267:10337–41
97. Ornitz DM, Herr AB, Nilsson M, Westman J, Svahn C, et al. 1995. *Science* 268:432–36
98. Aviezer D, Hecht D, Safran M, Eisinger M, David G, et al. 1994. *Cell* 79:1005–13
99. Folkman J, Klagsbrun M, Sasse J, Wadzinski M, Ingber D, et al. 1988. *Am. J. Pathol.* 130:393–400
100. Vigny M, Ollier-Hartmann MP, Lavigne M, Fayein N, Jeanny JC, et al. 1988. *J. Cell. Physiol.* 137:321–28
101. Ohtani H, Nakamura S, Watanabe Y, Mizoi T, Saku T, et al. 1993. *Lab. Invest.* 68:520–27
102. Timar J, Ladanyi A, Lapis K, Moczar M. 1992. *Am. J. Pathol.* 141:467–74
103. Cohen IR, Murdoch AD, Naso MF, Marchetti D, Berd D, et al. 1994. *Cancer Res.* 54:5771–74
104. Marchetti D, Menter D, Jin L, Nakajima M, Nicolson GL. 1993. *Int. J. Cancer* 55:692–99
105. Imamura T, Tokita Y, Mitsui Y. 1991. *Cell Struct. Funct.* 16:225–30
106. Aviezer D, Iozzo RV, Noonan DM, Yayon A. 1997. *Mol. Cell. Biol.* 17:1938–46
107. Mathiak M, Yenisey C, Grant DS, Sharma B, Iozzo RV. 1997. *Cancer Res.* 57:2130–36

108. Klein G, Conzelmann S, Beck S, Timpl R, Müller CA. 1995. *Matrix Biol.* 14: 457–65
109. Gauer S, Schulzelohoff E, Schleicher E, Sterzel RB. 1996. *Eur. J. Cell Biol.* 70:233–42
110. Frieser M, Hallmann R, Johansson S, Vestweber D, Goodman SL, et al. 1996. *Eur. J. Immunol.* 26:3127–36
111. Whitelock JM, Murdoch AD, Iozzo RV, Underwood PA. 1996. *J. Biol. Chem.* 271:10079–86
112. Saksela O, Rifkin DB. 1990. *J. Cell Biol.* 110:767–75
113. Kanwar YS, Linker A, Farquhar MG. 1980. *J. Cell Biol.* 86:688–93
114. van den Born J, van den Heuvel LPWJ, Bakker MAH, Veerkamp JH, Assmann KJM, et al. 1992. *Kidney Int.* 41:115–23
115. van den Born J, van den Heuvel LPWJ, Bakker MAH, Veerkamp JH, Assmann KJM, et al. 1993. *Kidney Int.* 43:454–63
116. Oda O, Shinzato T, Ohbayashi K, Takai I, Kunimatsu M, et al. 1996. *Clin. Chim. Acta* 255:119–32
117. Ruegg MA. 1996. *Curr. Opin. Neurobiol.* 6:97–103
118. Halfter W, Schurer B, Yip J, Yip L, Tsen G, et al. 1997. *J. Comp. Neurol.* 381:1–17
119. Godfrey EW, Dietz ME, Morstad AL, Wallskog PA, Yorde DE. 1988. *J. Cell Biol.* 106:1263–72
120. Barber AJ, Lieth E. 1997. *Dev. Dyn.* 208:62–74
121. Gautam M, Noakes PG, Moscoso L, Rupp F, Scheller RH, et al. 1996. *Cell* 85:525–35
122. DeChiara TM, Bowen DC, Valenzuela DM, Simmons MV, Poueymirou WT, et al. 1996. *Cell* 85:501–12
123. Glass DJ, Bowen DC, Stitt TN, Radziejewski C, Bruno J, et al. 1996. *Cell* 85:513–23
124. Slater CR. 1996. *Nature* 381:478–79
125. McCarthy KJ, Abrahamson DR, Bynum KR, St. John PL, Couchman JR. 1994. *J. Histochem. Cytochem.* 42:473–84
126. McCarthy KJ, Couchman JR. 1990. *J. Histochem. Cytochem.* 38:1479–86
127. Thomas GJ, Shewring L, McCarthy KJ, Couchman JR, Mason RM, et al. 1995. *Kidney Int.* 48:1278–89
128. Yamaguchi Y. 1996. *Perspect. Dev. Neurobiol.* 3:307–17
129. LeBaron RG. 1996. *Perspect. Dev. Neurobiol.* 3:261–71
130. Zimmermann DR, Ruoslahti E. 1989. *EMBO J.* 8:2975–81
131. Shinomura T, Nishida Y, Ito K, Kimata K. 1993. *J. Biol. Chem.* 268:14461–69
132. Neame PJ, Barry FP. 1993. *Experientia* 49:393–402
133. Kohda D, Morton CJ, Parkar AA, Hatanaka H, Inagaki FM, et al. 1996. *Cell* 86:767–75
134. LeBaron RG, Zimmermann DR, Ruoslahti E. 1992. *J. Biol. Chem.* 267:10003–10
135. Dours-Zimmermann MT, Zimmermann DR. 1994. *J. Biol. Chem.* 269:32992–98
136. Naso MF, Zimmermann DR, Iozzo RV. 1994. *J. Biol. Chem.* 269:32999–3008
137. Zako M, Shinomura T, Kimata K. 1997. *J. Biol. Chem.* 272:9325–31
138. Kishore U, Eggleton P, Reid KBM. 1997. *Matrix Biol.* 15:583–92
139. Ujita M, Shinomura T, Ito K, Kitagawa Y, Kimata K. 1994. *J. Biol. Chem.* 269:27603–9
140. Aspberg A, Binkert C, Ruoslahti E. 1995. *Proc. Natl. Acad. Sci. USA* 92: 10590–94
141. Doege KJ, Sasaki M, Kimura T, Yamada Y. 1991. *J. Biol. Chem.* 266:894–902
142. Watanabe H, Gao L, Sugiyama S, Doege K, Kimata K, et al. 1995. *Biochem. J.* 308:433–40
143. Doege KJ, Coulter SN, Meek LM, Maslen K, Wood JG. 1997. *J. Biol. Chem.* 272:13974–79
144. Fülöp C, Walcz E, Valyon M, Glant TT. 1993. *J. Biol. Chem.* 268:17377–83
145. Drickamer K. 1993. *Curr. Opin. Struct. Biol.* 3:393–400
146. Rauch U, Karthikeyan L, Maurel P, Margolis RU, Margolis RK. 1992. *J. Biol. Chem.* 267:19536–47
147. Retzler C, Wiedemann H, Kulbe G, Rauch U. 1996. *J. Biol. Chem.* 271: 17107–13
148. Yamada H, Watanabe K, Shimonaka M, Yamaguchi Y. 1994. *J. Biol. Chem.* 269:10119–26
149. Yamada H, Watanabe K, Shimonaka M, Yamasaki M, Yamaguchi Y. 1995. *Biochem. Biophys. Res. Commun.* 216: 957–63
150. Rauch U, Meyer H, Brakebusch C, Seidenbecher C, Gundelfinger ED, et al. 1997. *Genomics* 44:15–21
151. Seidenbecher CI, Richter K, Rauch U, Fässler R, Garner CC, et al. 1995. *J. Biol. Chem.* 270:27206–12
152. Iozzo RV, Naso MF, Cannizzaro LA, Wasmuth JJ, McPherson JD. 1992. *Genomics* 14:845–51
153. Naso MF, Morgan JL, Burchberg AM, Siracusa LD, Iozzo RV. 1995. *Genomics* 29:297–300

154. Doege KJ, Garrison K, Coulter SN, Yamada Y. 1994. *J. Biol. Chem.* 269: 29232–40
155. Walcz E, Deák F, Erhardt P, Coulter SN, Fülöp C, et al. 1994. *Genomics* 22:364–71
156. Valhmu WB, Palmer GD, Rivers PA, Ebara S, Cheng J-F, et al. 1995. *Biochem. J.* 309:535–42
157. Korenberg JR, Chen XN, Doege K, Grover J, Roughley PJ. 1993. *Genomics* 16:546–48
158. Watanabe H, Kimata K, Line S, Strong D, Gao L-y, et al. 1994. *Nat. Genet.* 7:154–57
159. Rauch U, Grimpe B, Kulbe G, Arnold-Ammer I, Beier DR, et al. 1995. *Genomics* 28:405–10
160. Ushkaryov YA, Südhof TC. 1993. *Proc. Natl. Acad. Sci. USA* 90:6410–14
161. Boyd CD, Pierce RA, Schwarzbauer JE, Doege K, Sandell LJ. 1993. *Matrix* 13:457–69
162. Grover J, Roughley PJ. 1993. *Biochem. J.* 291:361–67
163. Kähäri V-M, Larjava H, Uitto J. 1991. *J. Biol. Chem.* 266:10608–15
164. Landolt RM, Vaughan L, Winterhalter KH, Zimmermann DR. 1995. *Development* 121:2303–12
165. Perris R, Perissinotto D, Pettway Z, Bronner-Fraser M, Mörgelin M, et al. 1996. *FASEB J.* 10:293–301
166. Bode-Lesniewska B, Dours-Zimmermann MT, Odermatt BF, Briner J, Heitz PU, et al. 1996. *J. Histochem. Cytochem.* 44:303–12
167. Wight TN, Kinsella MG, Qwarnström EA. 1992. *Curr. Opin. Cell Biol.* 4:793–801
168. Halpert I, Sires UI, Roby JD, Potter-Perigo S, Wight TN, et al. 1996. *Proc. Natl. Acad. Sci. USA* 93:9748–53
169. Camaioni A, Salustri A, Yanagishita M, Hascall VC. 1996. *Arch. Biochem. Biophys.* 325:190–98
170. du Cros DL, LeBaron RG, Couchman JR. 1995. *J. Invest. Dermatol.* 105:426–31
171. Schönherr E, Järveläinen HT, Sandell LJ, Wight TN. 1991. *J. Biol. Chem.* 266: 17640–47
172. Häkkinen L, Westermarck J, Kähäri V-M, Larjava H. 1996. *J. Dent. Res.* 75:1767–78
173. Schönherr E, Kinsella MG, Wight TN. 1997. *Arch. Biochem. Biophys.* 339: 353–61
174. Tiedemann K, Malmström A, Westergren-Thorsson G. 1996. *Matrix Biol.* 15:469–78
175. Adany R, Iozzo RV. 1990. *Biochem. Biophys. Res. Commun.* 171:1402–13
176. Iozzo RV. 1995. *Lab. Invest.* 73:157–60
177. Isogai Z, Shinomura T, Yamakawa N, Takeuchi J, Tsuji T, et al. 1996. *Cancer Res.* 56:3902–8
178. Nara Y, Kato Y, Torii Y, Tsuji Y, Nakagaki S, et al. 1997. *Histochem. J.* 29:21–30
179. Yamagata M, Kimata K. 1994. *J. Cell Sci.* 107:2581–90
180. Rohde M, Warthoe P, Gjetting T, Lukas J, Bartek J, et al. 1996. *Oncogene* 12: 2393–401
181. Zimmermann DR, Dours-Zimmermann MT, Schubert M, Bruckner-Tuderman L. 1994. *J. Cell Biol.* 124:817–25
182. Li H, Schwartz NB, Vertel BM. 1993. *J. Biol. Chem.* 268:23504–11
183. Primorac D, Stover ML, Clark SH, Rowe DW. 1994. *Matrix Biol.* 14:297–305
184. Grumet M, Flaccus A, Margolis RU. 1993. *J. Cell Biol.* 120:815–24
185. Friedlander DR, Milev P, Karthikeyan L, Margolis RK, Margolis RU, et al. 1994. *J. Cell Biol.* 125:669–80
186. Grumet M, Milev P, Sakurai T, Karthikeyan L, Bourdon M, et al. 1994. *J. Biol. Chem.* 269:12142–46
187. Milev P, Maurel P, Häring M, Margolis RK, Margolis RU. 1996. *J. Biol. Chem.* 271:15716–23
188. Engel M, Maurel P, Margolis RU, Margolis RK. 1996. *J. Comp. Neurology* 366:34–43
189. Meyer-Puttlitz B, Milev P, Junker E, Zimmer I, Margolis RU, et al. 1995. *J. Neurochem.* 65:2327–37
190. Watanabe E, Aono S, Matsui F, Yamada Y, Naruse I, et al. 1995. *Eur. J. Neurosci.* 7:547–54
191. Margolis RU, Margolis RK. 1994. *Methods Enzymol.* 245:105–26
192. Oldberg C, Antonsson P, Lindblom K, Heinegård D. 1989. *EMBO J.* 8:2601–4
193. Blochberger TC, Vergnes J-P, Hempel J, Hassell JR. 1992. *J. Biol. Chem.* 267:347–52
194. Corpuz LM, Funderburgh JL, Funderburgh ML, Bottomley GS, Prakash S, et al. 1996. *J. Biol. Chem.* 271:9759–63
195. Bengtsson E, Neame PJ, Heinegård D, Sommarin Y. 1995. *J. Biol. Chem.* 270:25639–44
196. Deere M, Johnson J, Garza S, Harrison WR, Yoon S-J, et al. 1996. *Genomics* 38:399–404
197. Johnson J, Rosenberg L, Choi HU, Garza S, Höök M, et al. 1997. *J. Biol. Chem.* 272:18709–17

198. Krusius T, Ruoslahti E. 1986. *Proc. Natl. Acad. Sci. USA* 83:7683–87
199. Fisher LW, Termine JD, Young MF. 1989. *J. Biol. Chem.* 264:4571–76
200. Scholzen T, Solursh M, Suzuki S, Reiter R, Morgan JL, et al. 1994. *J. Biol. Chem.* 269:28270–81
201. Sawhney RS, Hering TM, Sandell LJ. 1991. *J. Biol. Chem.* 266:9231–40
202. Oldberg C, Antonsson P, Moses J, Fransson L-C. 1996. *FEBS Lett.* 386:29–32
203. Hocking AM, McQuillan DJ. 1996. *Glycobiology* 6:717
204. Bianco P, Riminucci M, Fisher LW. 1993. In *Dermatan Sulphate Proteoglycans*, ed. JE Scott, pp. 193–205. London: Portland
205. Yeo TK, Torok MA, Kraus HL, Evans SA, Zhou Y, et al. 1995. *J. Vasc. Res.* 32:175–82
206. Plaas AHK. 1992. *Trends Glycosci. Glycotech.* 4:445–55
207. Funderburgh JL, Funderburgh ML, Brown SJ, Vergnes J-P, Hassell JR, et al. 1993. *J. Biol. Chem.* 268:11874–80
208. Neame PJ, Choi HU, Rosenberg LC. 1989. *J. Biol. Chem.* 264:8653–61
209. Mann DM, Yamaguchi Y, Bourdon MA, Ruoslahti E. 1990. *J. Biol. Chem.* 265:5317–23
210. Blaschke UK, Hedbom E, Bruckner P. 1996. *J. Biol. Chem.* 271:30347–53
211. Plaas AHK, Neame PJ, Nivens CM, Reiss L. 1990. *J. Biol. Chem.* 265: 20634–40
212. Plaas AHK, Wong-Palms S. 1993. *J. Biol. Chem.* 268:26634–44
213. Scott PG, Winterbottom N, Dodd CM, Edwards E, Pearson CH. 1986. *Biochem. Biophys. Res. Commun.* 138:1348–54
214. Rada JA, Cornuet PK, Hassell JR. 1993. *Exp. Eye Res.* 56:635–48
215. McBride OW, Fisher LW, Young MF. 1990. *Genomics* 6:219–25
216. Danielson KG, Fazzio A, Cohen I, Cannizzaro LA, Eichstetter I, et al. 1993. *Genomics* 15:146–60
217. Chakravarti S, Stallings RL, SundarRaj N, Cornuet PK, Hassell JR. 1995. *Genomics* 27:481–88
218. Traupe H, van den Ouweland AMW, van Oost BA, Vogel W, Vetter U, et al. 1992. *Genomics* 13:481–83
219. Grover J, Chen X-N, Korenberg JR, Recklies AD, Roughley PJ. 1996. *Genomics* 38:109–17
220. Fisher LW, Heegaard A-M, Vetter U, Vogel W, Just W, et al. 1991. *J. Biol. Chem.* 266:14371–77
221. Wegrowski Y, Pillarisetti J, Danielson KG, Suzuki S, Iozzo RV. 1995. *Genomics* 30:8–17
222. Antonsson P, Heinegård D, Oldberg C. 1993. *Biochim. Biophys. Acta* 1174: 204–6
223. Grover J, Chen X-N, Korenberg JR, Roughley PJ. 1995. *J. Biol. Chem.* 270: 21942–49
224. Ujita M, Shinomura T, Kimata K. 1995. *Gene* 158:237–40
225. Santra M, Danielson KG, Iozzo RV. 1994. *J. Biol. Chem.* 269:579–87
226. Mauviel A, Santra M, Chen YQ, Uitto J, Iozzo RV. 1995. *J. Biol. Chem.* 270:11692–700
227. Mauviel A, Korang K, Santra M, Tewari D, Uitto J, et al. 1996. *J. Biol. Chem.* 271:24824–29
228. Heino J, Kähäri V, Mauviel A, Krusius T. 1988. *Biochem. J.* 252:309–12
229. Wegrowski Y, Paltot V, Gillery P, Kalis B, Randoux A, et al. 1995. *Biochem. J.* 307:673–78
230. Iozzo RV, Cohen I. 1993. *Experientia* 49:447–55
231. Klemsz MJ, McKercher SR, Celada A, Van Beveren C, Maki RA. 1990. *Cell* 61:113–24
232. Ungefroren H, Krull NB. 1996. *J. Biol. Chem.* 271:15787–95
233. Meyer DH, Krull N, Dreher KL, Gressner AM. 1992. *Hepatology* 16:204–16
234. Ungefroren H, Cikós T, Krull NB, Kalthoff H. 1997. *Biochem. Biophys. Res. Commun.* 235:413–17
235. Nelimarkka L, Kainulainen V, Schönherr E, Moisander S, Jortikka M, et al. 1997. *J. Biol. Chem.* 272:12730–37
236. Westergren-Thorsson G, Antonsson P, Malmström A, Heinegård D, Oldberg C. 1991. *Matrix* 11:177–83
237. Romaris M, Heredia A, Molist A, Bassols A. 1991. *Biochim. Biophys. Acta* 1093:229–33
238. Vogel KG, Hernandez DJ. 1992. *Eur. J. Cell Biol.* 59:304–13
239. Roughley PJ, Melching LI, Recklies AD. 1994. *Matrix Biol.* 14:51–59
240. Kähäri VM, Hakkinen L, Westermarck J, Larjava H. 1995. *J. Invest. Dermatol.* 104:503–8
241. Geerkens C, Vetter U, Just W, Fedarko NS, Fisher LW, et al. 1995. *Hum. Genet.* 96:44–52
242. Kadler KE, Holmes DF, Trotter JA, Chapman JA. 1996. *Biochem. J.* 316:1–11
243. Toole BP, Lowther DA. 1968. *Biochem. J.* 109:857–66
244. Toole BP. 1969. *Nature* 222:872–73

245. Scott JE, Orford CR. 1981. *Biochem. J.* 197:213–16
246. Vogel KG, Paulsson M, Heinegård D. 1984. *Biochem. J.* 223:587–97
247. Hedbom E, Heinegård D. 1993. *J. Biol. Chem.* 268:27307–12
248. Nurminskaya MV, Birk DE. 1996. *Biochem. J.* 317:785–89
249. Vogel KG, Koob TJ, Fisher LW. 1987. *Biochem. Biophys. Res. Commun.* 148: 658–63
250. Vogel KG, Trotter JA. 1987. *Collagen Rel. Res.* 7:105–14
251. Chandrasekhar S, Kleinman HK, Hassell JR, Martin GR, Termine JD, et al. 1984. *Collagen Rel. Res.* 4:323–38
252. Birk DE, Nurminskaya MV, Zycband EI. 1995. *Dev. Dyn.* 202:229–43
253. Weber IT, Harrison RW, Iozzo RV. 1996. *J. Biol. Chem.* 271:31767–70
254. Scott JE. 1996. *Biochemistry* 35:8795–99
255. Kobe B, Deisenhofer J. 1993. *Nature* 366:751–56
256. Scott JE. 1995. *J. Anat.* 187:259–69
257. Yu L, Cummings C, Sheehan JK, Kadler KE, Holmes DF, et al. 1993. In *Dermatan Sulphate Proteoglycans*, ed. JE Scott, pp. 183–92. London: Portland
258. Spiro RC, Countaway JL, Gaarde WA, Garcia JA, Leisten J, et al. 1994. *Mol. Biol. Cell* 55:A303
259. Svensson L, Heinegård D, Oldberg C. 1995. *J. Biol. Chem.* 270:20712–16
260. Schönherr E, Hausser H, Beavan L, Kresse H. 1995. *J. Biol. Chem.* 270: 8877–83
261. Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, et al. 1997. *J. Cell Biol.* 136:729–43
262. Hildebrand A, Romaris M, Rasmussen LM, Heinegård D, Twardzik DR, et al. 1994. *Biochem. J.* 302:527–34
263. Yamaguchi Y, Ruoslahti E. 1988. *Nature* 336:244–46
264. Yamaguchi Y, Mann DM, Ruoslahti E. 1990. *Nature* 346:281–84
265. Border WA, Noble NA, Yamamoto T, Harper JR, Yamaguchi Y, et al. 1992. *Nature* 360:361–64
266. Isaka Y, Brees DK, Ikegaya K, Kaneda Y, Imai E, et al. 1996. *Nat. Med.* 2:418–23
267. Kresse H, Hausser H, Schönherr E, Bittner K. 1994. *Eur. J. Clin. Chem. Clin. Biochem.* 32:259–64
268. Takeuchi Y, Kodama Y, Matsumoto T. 1994. *J. Biol. Chem.* 269:32634–38
269. Santra M, Skorski T, Calabretta B, Lattime EC, Iozzo RV. 1995. *Proc. Natl. Acad. Sci. USA* 92:7016–20
270. De Luca A, Santra M, Baldi A, Giordano A, Iozzo RV. 1996. *J. Biol. Chem.* 271:18961–65
271. Santra M, Mann DM, Mercer EW, Skorski T, Calabretta B, et al. 1997. *J. Clin. Invest.* 100:149–57
271a. Moscatello DK, Santra M, Mann DM, McQuillan DJ, Wong AJ, et al. 1998. *J. Clin. Invest.* 101:406–12
271b. Patel S, Santra M, McQuillan DJ, Iozzo RV, Thomas AP. 1998. *J. Biol. Chem.* 273:3121–24
271c. Vogel W, Gish GD, Alves F, Pawson T. 1997. *Mol. Cell.* 1:13–23
271d. Shrivastava A, Radziejewski C, Campbell E, Kovac L, McGlynn M, et al. 1997. *Mol. Cell.* 1:25–34
272. Adany R, Heimer R, Caterson B, Sorrell JM, Iozzo RV. 1990. *J. Biol. Chem.* 265:11389–96
273. Adany R, Iozzo RV. 1991. *Biochem. J.* 276:301–6
274. Huttenlocher A, Werb Z, Tremble P, Huhtala P, Rosenberg L, et al. 1996. *Matrix Biol.* 15:239–50
275. Järveläinen HT, Iruela-Arispe ML, Kinsella MG, Sandell LJ, Sage EH, et al. 1992. *Exp. Cell Res.* 203:395–401
276. Kinsella MG, Tsoi CK, Järveläinen HT, Wight TN. 1997. *J. Biol. Chem.* 272: 318–25
277. Krumdieck R, Höök M, Rosenberg LC, Volanakis JE. 1992. *J. Immunol.* 149:3695–701
278. Hocking AM, Strugnell RA, Ramamurthy P, McQuillan DJ. 1996. *J. Biol. Chem.* 271:19571–77
279. Kamo I, Kikuchi A, Nonaka I, Yamada E, Kondo J. 1993. *Biochem. Biophys. Res. Commun.* 195:1119–26
280. Oritani K, Kincade PW. 1996. *J. Cell Biol.* 134:771–81
281. Birk DE, Hahn RA, Linsenmayer C, Zycband EI. 1996. *Matrix Biol.* 15:111–18
282. Hahn RA, Birk DE. 1992. *Development* 115:383–93
283. Cornuet PK, Blochberger TC, Hassell JR. 1994. *Invest. Ophthalmol. Vis. Sci.* 35:870–77
284. Nakazawa K, Hassell JR, Hascall VC, Lohmander LS, Newsome DA, et al. 1984. *J. Biol. Chem.* 259:13751–57
285. Quantock AJ, Klintworth GK, Schanzlin DJ, Capel MS, Lenz ME, et al. 1996. *Biophys. J.* 70:1966–72