UNIT 10.1 Overview of Extracellular Matrix

Robert P. Mecham, Ph.D.
Dept. of Cell Biology & Physiology, Washington University School of Medicine,
St. Louis, MO  63110.

INTRODUCTION

In all multicellular organisms, development is influenced by the interactions between cells and their extracellular matrix (ECM). Information contained in the ECM provides the cell with temporal and positional clues, such as where it is, where it should be going, how old it is (in terms of cellular differentiation), and in some instances, when it is time for it to die (through apoptosis). It is not surprising, then, that there has been a great deal of interest in defining the extracellular signals as well as the cell surface receptors that interact with these molecules and interpret the information. Now more than ever, understanding cell biology requires understanding the ECM.

Studying the ECM, however, is not for the faint of heart! In most instances, the functional form of matrix macromolecules is a large, sparingly soluble aggregate that cannot easily be solubilized or dissociated into component units. Even when dissociated matrix components are obtained, the biological properties of the constituent chains often differ from the intact form. To complicate matters, most ECM macromolecules participate in supramolecular assemblies where their biological properties are modified by the molecules with which they interact.

These unusual physical properties create serious problems for matrix characterization using a standard "wet chemistry" approach. They also create some, though not many, unique advantages. For example, the multimeric, cross-linked nature of ECM imparts an element of stability that is not found in other proteins. This is most obvious if one takes a historical look at techniques used for matrix purification (Partridge, 1962; Piez, 1997). In the early days of matrix biology, "connective tissue" was purified using extraction protocols that relied on the ability of the target matrix component to withstand relatively harsh conditions: acid solutions were used for purifying collagen, chaotropic agents for mucopolysaccharides (now called proteoglycans), and, the harshest of them all, boiling sodium hydroxide for purifying elastin. It is quite remarkable that so much of what we know about these three matrix classes resulted from experiments using products purified in this way. Although purification strategies are now a bit more sophisticated, modifications of these basic protocols are still used today. The use of molecular biology and mouse genetics has quickened the pace of matrix characterization and opened the door to functional studies of complex matrices that were unthinkable several years ago.

One of the most important properties of ECM is its functional diversity (Kleinman, 1993). Some components are designed to be rigid, others elastic; some wet, others sticky. All have modular designs that impart diverse roles, yet allow for highly specialized functions. The formation of a basement membrane, for example, requires the assembly of ECM molecules that have significant tensile strength (collagen), can act as charged molecular sieves (proteoglycans), and facilitate cell attachment (laminin). These molecules are woven together through processes that involve self-assembly and interactions with molecules that are specifically designed to serve as molecular bridges or linkers (nidogen/entactin; Yurchenco, 1994).

It is not possible to ascertain the functional properties of a complex matrix such as basement membrane without studying its individual components. At the same time, however, it is also clear that the functional complexity of the assembled basement membrane is greater than the sum of its component parts. To comprehend this greater sum requires shifting one's view away from a reductionist biochemical approach to one focused on cell and developmental biology. Here the
cell becomes the reagent, interpreting informational signals contained in the ECM and adjusting its physiology accordingly. The researcher's task is to understand the readout.

The sections below contain an overview of the major classes of ECM. Molecules have been selected to illustrate specific functional or structural properties that are common to a matrix class or to ECM macromolecules generally. Where possible, recent reviews with references to more detailed literature are cited. Although somewhat dated, the text *Cell Biology of Extracellular Matrix* (Hay, 1991) provides an excellent overview of ECM biology. More detailed reviews can be found in various volumes of the *Biology of Extracellular Matrix* series, published by Academic Press.

**COLLAGENS**

**Structure of Collagens**

Collagen is the most ubiquitous ECM protein and is designed to provide structure and resiliency to tissues. It is defined by the presence of a triple-helical domain containing peptide chains with repeating Gly-Xaa-Yaa triplets, and by the presence of hydroxyproline and hydroxylysine (Kuhn, 1987; Prockop and Kivirikko, 1995). To date, nineteen distinct genetic collagen types have been identified. The characteristic molecular form of collagen is a triple helix made up of three polypeptides, called α chains, that coil into a right-handed triple helix. Collagens exist either as homotrimers composed of three identical α chains (α1)3 or as heterotrimers consisting of two (α1α2)3 or three (α1α2α3) α chains.

The nomenclature for the collagen superfamily consists of an indication of their genetic type (a Roman numeral that generally denotes the chronological order in which the collagens were characterized) together with the α-chain composition. Type I collagen, for example, is a heterotrimer of two α1 chains and one α2 chain, and is indicated as (α1[II])2α2(I). Type II collagen is a homotrimer of three α1 chains and is written (α1[II])3. Other collagens consist of three different α chains and (using type IX as an example) are written in the form α1(IX)α2(IX)α3(IX). It is important to note that each α chain within a collagen type is a distinct gene product; that is, an α1 chain in one collagen type is not the same protein as the α1 chain in any other collagen type. It is critical, therefore, to indicate the collagen type when referring to a particular α chain (e.g., the α1 chain of type I collagen).

**Synthesis of Collagens**

Collagen α chains are synthesized on membrane-bound ribosomes (ER) as large precursors, called pre-pro-α chains. In addition to the signal peptide (the "pre" part of the name) required for transport into the ER, each collagen precursor has extension peptides (the "pro" part) on both its N- and C-terminal ends (Fig. 10.1.1). Each pro-α chain combines with two others in the lumen of the ER to form the triple-helical molecule. The extension peptides are required for correct triple helix formation and remain with the triple-helical unit throughout the secretory pathway.

In the triple helix, the side chain of every third α-chain residue is directed towards the center of the helix, shifted by 30° from the preceding central residue of the same chain (Brodsky and Ramshaw, 1997). Steric constraints dictate that the center of the helix be occupied only by glycine residues; side chains of any other amino acid would perturb the triple-helical conformation.

Hydroxylation of proline residues in the Yaa position occurs as a post-translational modification in the lumen of the ER. The side-chain hydroxyl group of hydroxyproline stabilizes the helix through the formation of intermolecular hydrogen bonds. In fact, hydroxylation of ~100 prolyl residues is essential for the three pro-α chains of fibrillar collagens to form a triple helix that is stable at body temperatures. Hydroxylation of α chains is catalyzed by prolyl 4-hydroxylase, a tetrameric
enzyme consisting of two $\alpha$ and two $\beta$ subunits ($\alpha_2\beta_2$; Kivirikko and Myllyharju, 1998). Interestingly, the $\beta$ subunit is protein disulfide isomerase, an ER protein that catalyzes thiol-disulfide interchange during protein folding (Koivu et al., 1987). The hydroxylation reaction catalyzed by prolyl 4-hydroxylase requires Fe2+, 2-oxoglutarate, O2, and ascorbate. Conditions that prevent proline hydroxylation (such as nutritional deficiency of iron or of vitamin C) affect helix formation or stability. In scurvy, a human disease caused by a dietary deficiency of vitamin C, the nonhydroxylated pro-$\alpha$ chains are unstable and the skin and blood vessels become extremely fragile.

A second post-translational modification of procollagen that is crucial to its function is the hydroxylation of lysine residues. This reaction, which also occurs in the ER, is catalyzed by the enzyme lysyl hydroxylase. The active enzyme is a homodimer and, like prolyl hydroxylase, requires Fe2+, 2-oxoglutarate, O2, and ascorbate. Hydroxylysine residues have two important functions: their hydroxy groups act as attachment sites for carbohydrate units, and they are essential for the stability of the intermolecular collagen cross-links that occur in the extracellular space after secretion. The glycosylation of hydroxylysine is unusual, consisting of a single galactose residue or a glucosyl-galactosyl disaccharide attached to the hydroxyl group. The amount of carbohydrate added to procollagen varies greatly among different types of collagen, and its function is unknown.

**Assembly of Collagens**

After secretion into the extracellular space, the extension peptides of procollagen are removed by specific proteolytic enzymes. Both the N and C proteinases are members of the zinc metallopeptidase family and contain domains that suggest the ability to interact with cells and other matrix components (Kessler et al., 1996; Colige et al., 1997). Removal of the extension peptides converts the procollagen molecules to collagen (once called tropocollagen). Triple-helical collagen units then come together in the extracellular space to form the much larger collagen fibrils. The process of fibril formation is driven, in part, by the tendency of the collagen molecules to self-assemble. The fibrils form close to the cell surface, however, and it seems likely that the cell regulates the sites and rates of fibril assembly. The nonfibrillar collagens (see below) undergo only limited proteolytic processing prior to assembly. Here it is important to distinguish between collagen and gelatin. As stated above, collagen is the triple-helical form of the protein and can exist as single triple-helical units or triple-helical units polymerized into fibrils. Gelatin is denatured collagen. The individual $\alpha$ chains are no longer in a triple helix but can nevertheless polymerize into a random gel under appropriate conditions of temperature and ionic strength.

Collagen fibrils are greatly strengthened by covalent cross-links within and between the constituent collagen molecules. The types of covalent bonds involved are unique to connective tissue and are formed through deamination of certain lysine and hydroxylysine residues to yield highly reactive aldehyde groups. The aldehydes then undergo classical condensation reactions to form covalent bonds with each other or with other lysine or hydroxylysine residues. The extent and type of cross-linking varies from tissue to tissue, depending on tissue requirements. For example, collagen is highly cross-linked in tendons, where tensile strength is crucial. Lysyl oxidase, the enzyme that catalyzes cross-link formation, requires copper and molecular oxygen. If cross-linking is inhibited, collagenous tissues become fragile, and structures such as skin, tendons, and blood vessels tend to tear.

**Collagen Classification**

The polymeric structures formed by members of the collagen family vary depending on collagen type (Prockop and Kivirikko, 1995). The structures formed result, in large part, from the nontriple-helical "modules" found within many of the nonfibrillar collagens (Brown and Timpl, 1995). Based on structural similarities, the collagen superfamily can be divided into the following classes.
Fibril-forming collagens: types I, II, III, V, and XI. These collagens (Kuhn, 1987; Kadler, 1994) all share a long triple-helical segment with a continuous Gly-Xaa-Yaa repeat over its entire length. They assemble into cross-striated fibers upon cleavage of N and C propeptides, with the individual units adopting a one-quarter stagger relative to their neighbors in the fibril. Types II and XI collagen undergo alternative splicing, and hybrid molecules containing both types V and XI collagen have been identified in some tissues.

Network-forming collagens: types IV, VIII, and X. α1 chains in the type IV collagen family (Hulmes, 1992; Kuhn, 1994; Yurchenco, 1994) contain a large collagenous domain that is frequently interrupted by short noncollagenous sequences (i.e., something other than Gly-Xaa-Yaa). Noncollagenous domains are also found at the N and C termini of the chain, with the C-terminal domain being the larger of the two. Monomers associate at the C termini to form dimers and at the N termini to form tetramers. The triple-helical domains intertwine to form supercoiled structures, resulting in a net-like structure. Type VIII collagen is found in Descemet's membrane in the eye and forms a stack of hexagonal lattices. A similar structure is formed by type X collagen synthesized by hypertrophic chondrocytes in the deep-calcifying zone of cartilage.

Fibril-associated collagens with interrupted triple helices (FACIT): types IX, XII, XIV, XVI, and XIX. These collagens (Mayne and Brewton, 1993; Olsen et al., 1995) are characterized by short triple-helical segments interrupted by short noncollagenous domains. They attach to the surface of fibril-forming collagens and do not form fibrils themselves. Type IX collagen is found on the surface of type II collagen, to which it is covalently bound. An unusual property of this collagen is the presence of a glycosaminoglycan (GAG) chain attached to a noncollagenous domain of the α2(IX) chain. Types XII and XIV collagen show structural similarities to type IX, including an attached GAG side chain. Types XVI and XIX have not been fully characterized but show similarities in structure to other members of the family.

Beaded filaments and anchoring fibrils: types VI and VII. Among the collagens of this family (Burgeson, 1993; Timpl and Chu, 1994), type VI collagen is characterized by α chains containing large N- and C-terminal globular domains separated by a small triple-helical segment. Alternative splicing produces variants of the α2(VI) and α3(VI) chains. Type VI collagen forms small beaded filaments in the ECM. Type VII collagen forms anchoring fibrils that link basement membranes to anchoring plaques of type IV collagen and laminin in the underlying ECM. Type VII collagen contains the longest triple helix of any known collagen, with only small interruptions throughout. The NC1 domain of type VII collagen binds to collagen types I and IV, fibronectin, and laminin 5.

Collagens with a transmembrane domain: types XIII and XVII. Types XIII and XVII collagen (Li et al., 1996) are unique in having a transmembrane domain with its N terminus predicted to be in the cytoplasm. Type XIII collagen undergoes extensive alternative splicing. Type XVII collagen is found primarily in the hemidesmosomes of the skin and is one of the antigens that produces the autoimmune disease bullous pemphigoid.

Other nonfibrillar collagens: types XV and XVIII. Types XV and XVIII collagen (Rehn and Pihlajaniemi, 1994) have large N- and C-terminal globular domains and a highly interrupted triple helix. Their large number of potential N- and O-linked glycosylation sites suggests that both types have the potential to be highly glycosylated.

ELASTIN AND MICROFIBRILLAR PROTEINS

Elastin

During evolution, with the advent of the closed circulatory system, came the requirement for blood vessels to accommodate the pulsatile blood flow of the heart. Vessels made mostly of collagen were too stiff, so in its place, we see the emergence of a matrix protein that has the properties of elastic recoil. This protein, elastin, is the predominant protein component of the elastic fiber that is
of particular importance to the structural integrity and function of tissues in which reversible extensibility or deformability are crucial, such as the major arterial vessels, lungs, and skin.

In contrast to the genetic diversity evident in the collagen gene family, elastin is encoded by only one gene. Like collagen, elastin maturation in the ECM involves the assembly of a soluble precursor molecule (tropoelastin) into a highly cross-linked polymer. This assembly process is more complex than that for collagen, however, because the ability to self-assemble does not appear to be an intrinsic property of tropoelastin. Instead, elastin assembly requires helper proteins to align the multiple cross-linking sites on elastin monomers (Mecham and Davis, 1994).

Two functional domains repeat along the tropoelastin molecule (Fig. 10.1.2). One domain, related to cross-link formation, is an $\alpha$ helix containing alanine and lysine. The other, related to extensibility, is enriched in glycine, valine, and proline. The hydrophobic amino acids in this domain are arranged in repeating sequences that form a succession of $\beta$ turns. The stacked $\beta$ turns form a $\beta$ spiral with a hydrophobic core. Stretching the elastin polymer exposes the hydrophobic core to water. Recoil occurs when the leaves of the $\beta$ spiral contract to shield the hydrophobic amino acids from the aqueous microenvironment. Mature, cross-linked elastin is extremely hydrophobic and insoluble under most conditions (including when boiled in sodium hydroxide; Partridge, 1962). Its unusual physical properties make insoluble elastin one of the most stable proteins in the body—lasting the lifetime of the organism. Two polyfunctional cross-links, desmosine and isodesmosine, are unique to elastin and can be used as specific markers for this protein.

**Fibrillin**

Microfibrils were first identified as components of elastic fibers. They are found in greatest abundance in elastic tissues or in the ciliary zonules of the eye, although their distribution is widespread. Fibrillin-1 and -2 play key roles in microfibrillar architecture. These 350-kDa glycoproteins are highly homologous (Fig. 10.1.2), with modular structures consisting of repeating calcium-binding epidermal growth factor (EGF)-like domains interspersed between 8-cysteine domains similar to those found in the latent transforming growth factor-$\beta$ (TGF-$\beta$)-binding protein family (Lee et al., 1991). Tandemly arranged EGF domains form a structural motif found frequently in ECM macromolecules (e.g., laminin, fibulin, latent TGF-$\beta$-binding protein, nidogen). When stacked together, these tightly folded, disulfide-bonded loop structures form a rigid, rod-like arrangement stabilized by interdomain calcium binding and hydrophobic interactions (Downing et al., 1996). The precise function of microfibrils is unclear, although their association with developing elastic fibers suggests a role in elastin assembly. Both fibrillin-1 and fibrillin-2 interact with the $\alpha v 3$ integrin through an Arg-Gly-Asp (RGD) sequence (see Adhesive Glycoproteins).

**ADHESIVE GLYCOPROTEINS**

**Introduction**

Most, if not all, ECM macromolecules interact with binding proteins on the surface of cells. In many instances, this is through a unique sequence motif that is accessible as part of the protein’s folded functional structure, or cryptic and exposed only when the protein undergoes a conformational change induced by binding to another protein or as the result of degradation or denaturation. One such “recognition motif” is the well-known RGD sequence that is recognized by several members of the integrin family.

**Fibronectin**

A great deal of biochemical work has led to a model of the fibronectin molecule in which the protein’s binding functions and its structure are clearly correlated (Hynes, 1990). The molecule is
secreted as a dimer consisting of two similar subunits joined together at the C terminus by disulfide bonds (Fig. 10.1.3). Each chain has a molecular weight of ~220 to 250 kDa and is subdivided into a series of tightly folded domains. Each domain is responsible for one of fibronectin's binding functions. In plasma, fibronectin exists as a soluble dimer, but in the ECM it is found as an insoluble multimer.

Amino acid sequence analysis of fibronectin shows that the molecule is made up mostly of three repeating motifs, referred to as types I, II, and III repeats. These repeats are organized into functional domains that contain binding sites for ECM proteins and cell surface receptors (see Fig. 10.1.3). For example, there are two fibrin-binding domains consisting of multiple type I repeats on each subunit of the protein. Type I repeats are also found in the collagen-binding domain, and the first five type I repeats play an important role in matrix assembly. The N-terminal domain of fibronectin also mediates fibronectin's binding to gram-positive bacteria through type I modules. The type I module contains ~45 amino acids with four cysteines forming two disulfide bonds (Potts and Campbell, 1994). This module has also been found in a number of other proteins. In addition to type I repeats, the collagen-binding domain contains the only type II repeats found in fibronectin. Like type I repeats, these motifs contain two disulfide bonds, but they are larger than type I motifs.

The predominant structural feature of fibronectin consists of type III repeats, accounting for more than 60% of the sequence. No disulfide bonds are present in this structure, although two of the repeats contain a free cysteine. The cell-binding RGD sequence is located in the tenth type III repeat. This sequence is recognized by many members of the integrin family, including α5β1, αvβ3, αvβ5, αvβ6, αIIbβ3, and α8β1. Other cell-binding regions include the C-terminal heparin-binding domain and the type III-connecting segment (IIIICS), including the CS1 region. The type III consensus sequence is frequently found in other proteins.

Only one gene for fibronectin has been identified, but mRNAs for fibronectin have been shown to give rise to multiple versions of the protein through variable patterns of RNA splicing during gene transcription. Alternative splicing occurs predominately at three sites, termed extra type III domain A (EDA or EIIIA), extra type III domain B (EDB or EIIIB), and the connecting segment between the fourteenth and fifteenth type III repeats (IIIICS or V). Splicing within the IIIICS segment produces five variants, such that twenty different fibronectin subunits can result from splicing within the three segments.

Subunits of plasma fibronectin produced by adult hepatocytes contain neither EDA nor EDB segments, and one subunit lacks the entire IIIICS domain. Cultured fibroblasts, however, typically produce a form of fibronectin, referred to as cellular fibronectin that contains the EDA and/or EDB segments. Fibronectins expressed in fetal and tumor tissues contain a greater percentage of EDA and EDB segments than those expressed in normal adult tissues. The biological functions of fibronectin isoforms are only poorly understood, despite having been studied extensively. Differences in solubility have been demonstrated, but it has been difficult to detect functional differences between plasma and cellular fibronectin in their ability to promote cell adhesion and spreading.

**Vitronectin**

Vitronectin (also called serum spreading factor, S-protein, and epibolin) is a multifunctional protein found in plasma and ECM. It is synthesized as a single chain that undergoes N-glycosylation, tyrosine sulfation, and phosphorylation prior to secretion. In plasma, vitronectin circulates in two forms: a single chain of ~75 kDa and a proteolytically cleaved, two-chain form that dissociates into 65- and 10-kDa fragments upon reduction. It is present in fibrillar form in the ECM of a variety of tissues, where it sometimes colocalizes with fibronectin and elastic fibers. While little vitronectin immunoreactivity is detectable in most normal tissues, increased deposition has been observed in areas of tissue injury and necrosis. Tissue vitronectin was believed to be plasma derived, but recent studies indicate that extrahepatic cells have the biosynthetic potential
to produce vitronectin and that its synthesis can be regulated under inflammatory conditions
(Seiffert, 1997).

The cell attachment activity of vitronectin results from an RGD sequence that is recognized by a
wide variety of integrins. Most of the cell adhesive activity of serum used for tissue culture can be
attributed to vitronectin.

**Laminin and Basement Membranes**

Like fibronectin, the laminins are modular proteins with domains that interact with both cells and
ECM (Ekblom and Timpl, 1996). They constitute a family of basement membrane glycoproteins
that affect cell proliferation, migration, and differentiation. Eleven different laminins have been
identified, each containing an \( \alpha \), \( \beta \), and \( \gamma \) chain (Fig. 10.1.4). Electron microscopy has revealed
that all laminins have a cross-like shape with three short arms and one rod-like long arm, a shape
well suited for mediating interactions between sites on cells and components of the ECM (Beck et
al., 1990; Maurer and Engel, 1996). The rod-like regions separating the globular units of the short
arms are made up of repeating EGF-like domains. The long arm is formed by all three component
chains folding into an \( \alpha \)-helical coiled-coil structure, and is the only domain composed of multiple
chains. It is terminated by a large globular domain composed of five homologous subdomains
formed by the C-terminal region of the \( \alpha \) chain.

Along with type IV collagen, laminins are a major structural element of the basal lamina (Timpl,
1996). The molecular architecture of these matrices results from specific binding interactions
among the various components. The structural skeleton is formed by type IV collagen chains that
assemble into a covalently stabilized polygonal network. Laminin self-assembles through terminal
domain interactions to form a second polymer network. Nidogen (Mayer and Timpl, 1994) binds
laminin near its center and interacts with type IV collagen, bridging the two. A large heparan
sulfate proteoglycan (HS-PG), perlecan, binds laminin and type IV collagen through its GAG
chains and forms dimers and oligomers through a core-protein interaction. Perlecan is important
for charge-dependent molecular sieving, one of the critical functions of basement membrane.
Other components that are sometimes found associated with basement membranes but may not
be intrinsic components include fibronectin, type V collagen, fibulin, osteonectin (also known as
BM-40 or SPARC), and chondroitin sulfate proteoglycans.

Cells attach to laminin through specific interaction sites created by its multidomain structure. For
example, sites for receptor-mediated cell attachment and promotion of neurite outgrowth reside in
the terminal region of the long arm. A second cell-attachment site and a cell-signaling site with
mitogenic action are localized in the short arms. Cell binding to laminin occurs via a variety of
receptors, including non-integrins (Mecham and Hinek, 1996) and integrins (Aumailley et al.,
1996). The \( \alpha 1 \) family includes most of the laminin-binding integrins (\( \alpha 1 \beta 1, \alpha 2 \beta 1, \alpha 3 \beta 1, \alpha 7 \beta 1,
\alpha 9 \beta 1 \)). Other integrins that bind laminin include \( \alpha v \beta 3 \) and \( \alpha 6 \beta 4 \). Basement membrane can also
have an indirect effect on cells by binding and sequestering growth and differentiation factors,
such as fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), and TGF-\( \beta \).

The importance of laminin to cell differentiation and migration has been demonstrated in
developmental studies. Isoforms of laminin assembled from different chains are focally and
transiently expressed and may serve distinct functions at early stages of development even
before being deposited as components of basement membranes. Laminin is present at the two-
cell stage in the mouse embryo, making it one of the first ECM proteins detected during
embryogenesis.
MATRICELLULAR PROTEINS

Introduction

The term "matricellular" has been applied to a group of extracellular proteins that function by binding to matrix proteins and to cell surface receptors, but do not contribute to the structural integrity of the ECM (Bornstein, 1995). Proposed members of this group include the thrombospondins, members of the tenascin protein family, SPARC/osteonectin (Lane and Sage, 1994), and osteopontin. These proteins are frequently called "antiadhesive proteins" because of their ability to induce rounding and partial detachment of some cells in vitro (Sage and Bornstein, 1991). Their ability to interact with many different matrix proteins and cell surface receptors may explain their complex range of biological functions.

Thrombospondin

The thrombospondin (TSP) family consists of five secreted glycoproteins (Adams et al., 1995). TSP-1 and TSP-2 have identical domain structures and are secreted as disulfide-bonded homotrimers (Fig. 10.1.5). TSP-3, TSP-4, and TSP-5/COMP (cartilage oligomeric matrix protein) are pentamers whose expression is more limited than that of TSP-1 and TSP-2. TSP-1 binds HS-PGs, various integrins, the integrin-associated protein, and CD36. It also binds plasminogen, fibrinogen, fibronectin, urokinase, and TGF-β (which it can also activate). TSP-1 exhibits variable effects on cell adhesion and cell proliferation (Bornstein, 1995). For example, TSP-1 promotes proliferation of vascular smooth muscle cells, yet inhibits proliferation of endothelial cells. It supports attachment and spreading of skeletal myoblasts but expresses antiadhesive activity toward endothelial cells. Thrombospondin is the most abundant protein of platelet alpha granules and is released when platelets are activated.

Tenascin

The tenascins constitute a gene family consisting of four members: tenascins-C, -R, -X, and -Y (Erickson, 1993; Chiquet-Ehrismann, 1995). Tenascin-C (early names include GMEM, cytotoxin, J1, hexabrachion, and neuronectin) was the first form discovered and exists as a hexamer of disulfide-bonded subunits. Each subunit consists of a cysteine-rich N-terminal domain involved in oligomerization, EGF-like repeats, fibronectin type III-like repeats, and a fibrinogen-like globular domain (Fig. 10.1.5). The number of fibronectin type III-like repeats varies as a result of alternative splicing. Like TSP, tenascin-C has diverse biological effects when applied to cells. Both stimulation and inhibition of cellular proliferation have been observed in response to tenascin-C. In terms of cell adhesion, some cells do attach to tenascin, but weakly. In most instances, tenascin does not allow cell adhesion and can even inhibit cell attachment to other matrix proteins such as fibronectin and laminin. The finding that tenascin-C contains defined cell attachment sites suggests that the overall antiadhesive properties of the glycoprotein are effected by separate domains that override the attachment domains.

PROTEOGLYCANS

Introduction

The proteoglycans (once called acid mucopolysaccharides) constitute a number of genetically unrelated families of multidomain proteins that have covalently attached GAG chains. To date, more than 25 distinct gene products have been identified that carry at least one GAG chain (Iozzo and Murdoch, 1996). Like other matrix components discussed in this review, proteoglycans exist as structural variants, further increasing their functional and structural diversity.

For historical reasons, proteoglycans are named based on the type of attached GAG chain(s): (1) chondroitin sulfate and dermatan sulfate, consisting of a repeating disaccharide of galactosamine
and either glucuronic acid or iduronic acid; (2) heparin and heparan sulfate, consisting of a repeating disaccharide of glucosamine and either glucuronic acid or iduronic acid; and (3) keratan sulfate, consisting of a repeating disaccharide of glucosamine and galactose. Hyaluronate is also a repeating disaccharide but is not sulfated and not bound to a core protein. GAG chains are usually attached through O-glycosidic linkages to serine residues in the proteoglycan core protein. A characteristic feature of GAG chains is that at physiological pH they contain one to three negative charges per disaccharide due to carboxylate and sulfate groups.

Knowledge of the structure and function of proteoglycans increased dramatically when molecular biology was used to study the core proteins (Hassell et al., 1993). The heterogeneity of this family of matrix proteins also became evident with the finding that there are no structural domains common to all proteoglycans. There are, however, distinguishing characteristics that allow them to be grouped into four broad categories.

Large Proteoglycans that Form Aggregates by Interaction with Hyaluronan

These proteoglycans interact with strands of hyaluronate to form a very-high-molecular-weight aggregate. A structural trait shared by these proteoglycans is the presence of three functional domains: a globular hyaluronan-binding domain at the N terminus, a central extended region that carries most of the GAG chains, and a modular C-terminal domain containing two EGF repeats, a C-type lectin domain, and a complement-regulatory-protein-like motif (Iozzo and Murdoch, 1996).

The largest member of this family is versican (Zimmermann and Ruoslahti, 1989), a major proteoglycan in blood vessels that is also expressed in nonvascular tissues. Aggrecan, the large aggregating proteoglycan of cartilage, has a smaller core protein than versican but contains nearly 3-fold more GAG chains (Fig. 10.1.6). The high charge density of aggrecan results in each monomer occupying a large hydrodynamic volume. Aggrecan's GAG chains result in a high density of fixed charge in cartilage, producing an osmotic swelling pressure that is balanced by tension in the collagenous network. The reversible redistribution of proteoglycan-bound water under loading gives cartilage the ability to absorb compressive loads (Wight et al., 1991). Two other members of this family include neurocan (Rauch et al., 1992) and brevican (Yamada et al., 1994), both found in brain tissues.

Basement Membrane Proteoglycans

HS-PGs appear to be ubiquitous components of all basement membranes. Perlecan is the largest basement membrane proteoglycan, with a modular core protein of 467 kDa (Fig. 10.1.6; Iozzo et al., 1994). It provides the basement membrane with a negative charge that is important to its sieving properties. The heparan sulfate chains of perlecan also bind growth factors and cytokines and sequester them into the basement membrane, where they may function as a reserve to be released during tissue repair. The interaction of heparan sulfate with the FGFs has been extensively studied (Aviezer et al., 1994). Perlecan interacts with other components of the basement membrane, particularly laminin and nidogen. The multidomain structure of perlecan core protein is reminiscent of other ECM proteins, and includes EGF repeats and repeats of structures found in the low-density lipoprotein receptor, laminin chains, and neural cell adhesion molecule.

Agrin was originally isolated from torpedo ray electric organ and was found to induce acetylcholine receptor aggregation. It is secreted by motor neurons and deposited in the synaptic cleft basement membrane. Agrin may also play a role in the sequestration of growth factors in the basement membrane. Like perlecan, agrin is a multidomain protein with regions of EGF and laminin G-domain homology. Agrin is found predominantly in the brain, but has also been localized to smooth and cardiac muscle.
HS-PGs on the cell surface influence several important biological functions, including cell adhesion; the sequestration of heparin-binding ligands on the plasma membrane; and the promotion of dimerization/oligomerization of bound ligands, which enhances activation of primary signaling receptors.

Cell-associated HS-PGs have been divided into two major families, syndecan-like integral membrane HS-PGs (SLIPs) and glypican-related integral membrane HS-PGs (GRIPs; David, 1993; Carey, 1997). The SLIPs are transmembrane HS-PGs with a conserved intracellular domain that likely interacts with cytoskeletal and regulatory proteins. The GRIPs are linked to the cell surface by glycosyl phosphatidyl inositol in the outer leaflet of the lipid bilayer.

The syndecans, the principal form of cell-surface HS-PG, are synthesized by many cells. Syndecans bind a variety of extracellular ligands via their covalently attached heparan sulfate chains and are thought to play important roles in cell-matrix and cell-cell adhesion, migration, and proliferation. To date, four homologous syndecan core proteins have been cloned from vertebrate cells. All syndecans are type I transmembrane proteins, with an N-terminal signal peptide, an ectodomain that contains several consensus sequences for GAG attachment, a single hydrophobic transmembrane domain, and a short C-terminal cytoplasmic domain. The majority of GAG chains added to syndecan core proteins are of the heparan sulfate type, although syndecan-1 and syndecan-4 have chondroitin sulfate chains attached as well. Syndecans act as cell surface receptors for a number of matrix molecules, thereby mediating cell attachment and tissue organization. They influence the interactions of basic FGF and other growth factors with their receptors on cells and are responsible for the maintenance of a nonthrombogenic surface on endothelial cells.

Small Leucine-Rich Proteoglycans

Small leucine-rich proteoglycans (SLRPs) comprise a class of secreted proteoglycans that include five structurally related members: decorin, biglycan, fibromodulin, lumican, and epiphycan (see Fig. 10.1.7). Each has a leucine-rich core protein that assumes an arch-shaped structure with a concave surface capable of interacting with various other proteins. The N-terminal region contains one (decorin) or two (biglycan and epiphycan) GAG chains that can be either dermatan or chondroitin sulfate. Instead of GAG chains, fibromodulin and lumican have tyrosine sulfate in the N terminus, which provides an analogous negatively charged domain. These two SLRPs also contain N-linked keratan sulfate chains in their central domain.

SLRPs interact with numerous ECM proteins (e.g., fibronectin, TSP, fibrillin, microfibril-associated glycoprotein) and act to orient and order collagen fibers during development and tissue remodeling. Interactions with matrix proteins occur through the leucine-rich core which, in the case of type I collagen, influences collagen fibrillogenesis by binding to the surface of the collagen fibril at the d-band with the highly charged GAG chain extending out to regulate interfibrillar distances. Like other proteoglycans, SLRPs bind to growth factors (e.g., TGF-β) and thereby likely influence cellular differentiation and matrix synthesis. Decorin has recently been shown to directly regulate cell growth by activating the EGF receptor (Moscatello et al., 1998).

CONCLUSIONS

The furious pace of advances in the molecular biology of ECM has greatly expanded the knowledge of individual matrix components. The structure of many matrix macromolecules, for example, was determined from cloned cDNAs or genes long before complete protein information was available. With this increased knowledge as background, there is a growing realization that the information contained in the ECM is not a monosyllabic message encoded by individual molecules, but a complex and intricate arrangement dictated by the combinatorial organization of
the supramolecular structure. As the focus of biological research changes from the letters to the message, understanding how cells read and interpret this information will undoubtedly reveal more about the letters in the code.

LITERATURE CITED


From Current Protocols in Cell Biology Online
Copyright © 2003 John Wiley & Sons, Inc. All rights reserved.
Figure 10.1.1 Functional domains of the type I procollagen molecule. Following cleavage of the propeptide domains in the extracellular space, collagen units assemble in a quarter-stagger arrangement to form a fibril.
Figure 10.1.2 Domain map of tropoelastin and the fibrillins. Tropoelastin is secreted as a peptide of ~70 kDa and undergoes extensive covalent cross-linking during incorporation into the elastic fiber. Fibrillin-1 and fibrillin-2 each have a molecular weight of ~350 kDa and are the major structural elements of 10- to 12-nm-diameter microfibrils. Abbreviations: K-A, alanine-rich cross-linking domain; K-P, proline-rich cross-linking domain; RGD, Arg-Gly-Asp; CC, Cys-Cys sequences; CCC, Cys-Cys-Cys sequences; EGF, epidermal growth factor.
Figure 10.1.3 Domain map of fibronectin. The subunits of fibronectin vary in size between ~235 and 270 kDa. Alternative splicing occurs at three positions: EDA, EDB, and IIICS. Binding sites for other molecules and cells are indicated. Abbreviations: EDA, extra type III domain A; EDB, extra type III domain B; IIICS, connecting segment between the fourteenth and fifteenth type III repeats; RGD, Arg-Gly-Asp; CS-PG, chondroitin sulfate proteoglycan; aa, amino acids.
Figure 10.1.4 Domain map of laminin chains. Three polypeptide chains (α, β, and γ) form the laminin cross. The chain composition of known laminin types is shown in the insert. Abbreviation: EGF, epidermal growth factor.
Figure 10.1.5 Domain map of thrombospondin-1 and tenascin-C. The functional form of thrombospondin is a homotrimer of ~420 kDa. It interacts with numerous matrix proteins and modulates cell attachment by interacting with various cell-surface receptors. Tenascin-C monomers form a hexameric structure joined at their N termini by disulfide bonds. Alternative splicing leads to subunits of differing molecular weights. Abbreviations: HS-PG, heparan sulfate proteoglycan; CS-PG, chondroitin sulfate proteoglycan; TGF-$\beta$, transforming growth factor-$\beta$; RGD, Arg-Gly-Asp; EGF, epidermal growth factor.
Figure 10.1.6 Domain map of two representative large proteoglycans. Aggrecan is the core protein of the aggregating proteoglycan found in cartilaginous tissues. The molecular weight of the aggrecan core protein is 210 to 250 kDa. There are 100 to 150 keratan sulfate chains and many more chondroitin sulfate chains that contribute to the 2500-kDa molecular weight of the mature proteoglycan. The glycosaminoglycans are attached to repetitive sequences in the middle two-thirds of the molecule, including several types of repeats containing Ser-Gly, the linkage site for chondroitin sulfate. Perlecan is the largest of the basement membrane proteoglycans and has two or three attached heparan sulfate side chains. Removal of heparan sulfate side chains by heparatinase produces a core protein of 400 to 450 kDa on SDS-PAGE. Abbreviations: HA, hyaluronate; KS, keratan sulfate; CS, chondroitin sulfate; HS, heparan sulfate; LDL, low-density lipoprotein; EGF, epidermal growth factor.
Figure 10.1.7 Domain map of representative members of the small leucine-rich proteoglycans. Decorin contains a single chondroitin or dermatan sulfate chain attached near the N terminus. The core protein is ~38 kDa. Decorin is heterogeneous with respect to glycosaminoglycan (GAG) chain size, such that the secreted proteoglycan shows a range of molecular weights centered between 100 and 250 kDa. The core protein of biglycan is similar in size to that of decorin, except biglycan contains two chondroitin or dermatan sulfate chains. The GAG chains are also heterogeneous in size, resulting in a broad band on SDS-PAGE centered anywhere from 200 to 350 kDa. Removal of GAG chains with chondroitin ABC-lyase results in a 45-kDa band. Fibromodulin has a core protein size of 42 kDa. Four of the five potential N-glycosylation sites in the leucine-rich region of the molecule are substituted with keratan sulfate chains. Five to seven closely spaced tyrosine sulfate residues are found in the N-terminal domain. 

Abbreviations: CS, chondroitin sulfate; DS, dermatan sulfate.