

Instructional lecture

Current concepts of extracellular matrix

MARVIN L. TANZER

Department of Cell Biology and Anatomy, University of Arizona Health Sciences Center, PO Box 86535, Tucson, AZ 85754-6535, USA

I dedicate this lecture to the memory of Professor Katsuyuki Fujii, MD. Early in his career Professor Fujii studied as a postdoctoral research fellow in my laboratory. He was one of my most outstanding students and has been acclaimed as a leader by the international orthopedic community.

Extracellular matrix origin

The extracellular matrix (ECM) provides the scaffolding, support, and strength to tissues and organs. The ECM appeared during evolution with the advent of multicellular invertebrate animals.¹ The contemporary animal species that represent those early invertebrates contain the same macromolecular components found in all ECM, namely collagens, proteoglycans, and glycoproteins. New matrix domains appeared in multicellular animals that had not been present in primordial single-cell organisms, and the new domains then became conserved in the multicellular organisms.² Moreover, ancestral protein domains became genetically rearranged in various matrix proteins, yielding new molecules, including some whose functions remain to be discovered.³

It is commonly believed that the ECM provides a framework for cell adhesion and tissue development. We now know that cells have well-developed communication pathways between cell surfaces and the ECM.⁴ The view from the cell surface outward to the ECM has two perspectives: one toward the basement membrane and the other toward the interstitial matrix. Cells adherent to a basement membrane are necessarily polarized, containing basal and apical regions; and they show a precise orientation within specific tissues and organs. These cells rely primarily on integrins for recognition of basement membrane components; the integrins, as

transmembrane receptors, communicate signals that affect cell shape and motility.⁵

ECM properties

General principles of ECM organization are well known. Organized fibrous polymers, including collagens, elastins, and resilins, are embedded within an amorphous mixture of nonfibrous components as the fundamental theme. The need for inextensibility coupled with great strength mandates the presence of fibrous macromolecules, primarily those that aggregate into near crystalline arrays such as the collagens.⁶ Tissue resilience becomes predominant when elastic fibers such as elastin and resilin are strategically located in tissues. The amorphous substances in the ECM are predominantly proteoglycans, which can impart turgor.^{7,8} Adhesive glycoproteins, such as laminin, fibronectin, and entactin (nidogen), are enriched in basement membranes^{9–13} and are specifically recognized by discrete integrins.

The relative proportion of fibrous and nonfibrous components partially dictate the overall physical properties of a particular ECM, as does the diameter and weave of the fibers and the organizational arrangement of nonfibrous substances. We recognize ECM proteins as large, multifunctional molecules. They contain many domains, often repeated and occurring in tandem with each other. The domains yield various three-dimensional structures, often imparting several functions to one macromolecule.² Such functions include recognition by integrins and other cell receptors, ability to form coiled coils, predisposition to oligomerize, specific recognition by metalloproteinases,^{14,15} and ability to form distinct intramatrix interactions.

Collagen family

The most abundant ECM components tend to be large and have complex, unique structures. For example, collagenous proteins are defined by their presence of gly-X-Y repetitive sequences that give rise to a characteristic triple-helix conformation. That conformation might occur just once in a protein, may be present many times in tandem with noncollagen domains, or may comprise the entire protein. The family of collagenous proteins now includes representatives of each of these types. Collagenous family members are highly diverse,¹⁶ including the classical fiber-forming collagens, amorphous collagens such as the ones that form basement membranes, specialized collagens that organize into compact unique structures, and transmembrane collagens.¹⁷

Because the collagens are often crucial for tissue structure and function, defects in their normal expression may produce lethal and nonlethal phenotypes.^{16,18} Mutations of collagenous proteins are responsible for many clinical syndromes, affecting skeletal and nonskeletal tissues. The syndromes range from lethal to nonlethal, the latter ranging from mild clinical entities to debilitating disorders such as diverse forms of osteogenesis imperfecta. Analogous mutations of collagenous proteins are seen in domestic and laboratory animals, helping to establish the role of a specific collagen in physiologic homeostasis.

Biosynthesis of collagens necessarily must be complex to accommodate formation of a unique triple-helix conformation as well as including post-translational modifications such as proline and lysine hydroxylation. Proline hydroxylation yields primarily 4-hydroxyproline, which is responsible for stabilizing the molecular triple helix. Lysine hydroxylation, in contrast, is needed to stabilize the intermolecular crosslinks that develop between collagen molecules within collagen polymers.¹⁹ Such crosslinks help retain collagen polymer organization, enabling tensile forces to be transmitted without causing disruption of tissues and organs.

Collagenous proteins are synthesized as precursor forms, the procollagens, which are larger than the corresponding mature collagen molecules.^{20,21} Proper conversion of procollagen to collagen is essential, as mutations in that conversion process may also lead to defective individuals who clinically resemble those with a specific collagen mutation.

Many central details of collagen biosynthesis are known.²² A nascent collagenous chain is synthesized, as a procollagen precursor, on ribosomes bound to the cytoplasmic surface of the endoplasmic reticulum (ER). The growing nascent chain enters the cisternal compartment of the ER and, upon chain completion, is

released therein. Posttranslational hydroxylation and glycosylation occurs at this stage. Then, three nascent chains must recognize each other at their carboxyl termini to form the proper molecular composition (there are many genetically different chains), and the chains must align in proper register prior to zipping up into a triple helix conformation. The helical procollagen is transported from the ER to the Golgi region and is then secreted from the cell where the procollagen amino- and carboxyl-terminal domains are cleaved by specific proteases. The residual collagen molecule undergoes oxidation of specific lysines and hydroxylysines into aldehydes, catalyzed by lysyl oxidase, setting the scene for intermolecular crosslinking. The modified collagen self-assembles into fibrils or, in the case of nonfibrillar collagens, into random types of polymer. These macromolecular structures progressively develop crosslinks, gaining strength and imparting tensile properties to tissues and organs.¹⁹

Importantly, there are a large number of noncollagen mutations that affect skeletal patterning and differentiation, resulting in a great variety of body shape disorders that alter the axial skeleton, the facial skeleton, and the limbs.^{16,18} The mutated genes include those for growth factors and their receptors, transcription factors, homeoboxes, structural proteins such as cartilage oligomeric matrix protein (COMP), proteoglycans, hormone receptors, and vitamin receptors. Given such a wide variety of noncollagen genes it is not surprising that noncollagen mutations are more numerous than collagen mutations. Certain transcriptional factors play critical roles in early skeletal development.²³ They include Sox9, L-Sox5, Sox6, and Runx2, which are essential for cartilage formation; Sox9, Runx2, and Osterix are crucial for bone development.

Proteoglycans

Mature articular cartilage owes its resilience to the presence of aggrecan, a prototypical proteoglycan. Aggrecan's role was dramatically illustrated by Lewis Thomas' accidental results, when he caused erect rabbit ears to collapse by intravenously injecting the protease papain. Thomas astutely recognized that his dramatic result would provide insight into proteoglycan properties; he demonstrated that the affected rabbit ear cartilage lost its characteristic metachromatic staining, a hallmark of tissue proteoglycans. Amazingly, the rabbits recovered within a week, their ears regaining erect stature with restoration of the cartilage metachromatic staining. These observations were made prior to the molecular characterization of aggrecan but are sufficient to illustrate its essential role in creating cartilage turgor.

Aggrecan's critical role in skeletal development is illustrated by the chicken mutation nanomelia; the mutant animals fail to survive. The affected animals are dwarfs with insignificant amounts of cartilage; consequently, the bones fail to develop because they rely on a preformed cartilage template for development. Genetic analysis shows that a truncated aggrecan core protein is present in the mutant animals; the protein is about two-thirds normal length due to a premature stop codon. The missing portion of the core protein appears essential for proper routing of the core protein from the ER cisternal compartment into the Golgi region, ordinarily followed by secretion from the chondrocyte. Nanomelic core protein remains in the chondrocyte ER lumen without moving forward, and so nanomelic cartilage ECM is devoid of aggrecan.

Proteoglycans^{7,8} are implicated in other biological processes, serving as repositories for growth factors, assisting in modulating developmental processes, and playing recognition roles at cell surfaces. Molecular cloning has revealed families of proteoglycans; for example, aggrecan, brevican, versican, and neurocan are members of the lectican family, binding hyaluronan at their amino termini, whereas their C-termini have a selectin domain. The lecticans are decorated predominantly with chondroitin sulfate glycosaminoglycan (GAG) chains, with lesser numbers of keratan sulfate chains. Versican has been implicated in stimulating proliferation of fibroblasts and chondrocytes, apparently via its epidermal growth factor (EGF)-like motifs.

Numerous GAG-binding proteins have been described in the literature. The interaction between GAGs and proteins can have profound physiological effects on hemostasis, lipid transport and absorption, cell growth and migration, and development. Binding to GAGs can result in immobilization of proteins at their sites of production and in the matrix (potentially for future mobilization), regulation of enzyme activity, binding of ligands to their receptors, and protection of proteins against degradation. In some cases, the interaction has been shown to depend on a minor but highly specific sequence of modified sugars in the GAG chain. Most of the GAG-binding proteins that have been described interact with heparin/heparan sulfate or hyaluronic acid; relatively few are known to interact, with comparable avidity and affinity, with chondroitin sulfate or keratan sulfate.

One proteoglycan family has a protein core composed of leucine-rich repeats; members are decorin, biglycan, fibromodulin, and keratocan. They help organize collagen fibrillar networks and are decorated with chondroitin/dermatan sulfate or keratan sulfate. Decorin has recently been shown to participate in cellular signal transduction via the EGF receptor with down-

stream signaling through the mitogen-activated protein kinase (MAPK) pathway.

Proteoglycan GAG chains play vital roles in tissues by interacting with specific domains of other proteins. This phenomenon has been best characterized for the heparan sulfate proteoglycans, in particular the anticoagulant heparin, which specifically binds to circulating antithrombin, thereby inhibiting blood clotting. Heparin is used clinically to prevent blood clotting rapidly in patients who have a life-threatening thrombosis. This useful property of heparin has been pinpointed to a specific structure in heparin GAG chains, a unique pentasaccharide that binds to antithrombin with high affinity.

Cells also elaborate a diverse group of membrane proteoglycans that typically have type I orientations and either single membrane-spanning domains or a GPI anchor. Membrane proteoglycans tend to contain mostly HS (the glypicans), but many are hybrid structures containing both HS and CS (the syndecans and betaglycan). A few membrane proteoglycans contain exclusively CS (CD44 and NG2).

These cell membrane-associated proteoglycans bind to growth factors, forming ternary structures containing the proteoglycan plus the growth factor and its receptor. Such complexes modulate the signaling response of cells once they bind a growth factor. Syndecans are known to create such complexes with fibroblast growth factor, hepatocyte growth factor, platelet-derived growth factor, heparin-binding epidermal growth factor, and vascular endothelial growth factor. Downstream cellular activities such as adhesion, migration, proliferation, and differentiation are altered, demonstrating the importance of such ternary cell surface complexes.

Basement membranes

Virtually all organs and tissues in the adult animal have an organized architecture containing cells adherent to an underlying basement membrane. Ordinarily, the adherent cells do not move; and with few exceptions, the basement membrane is an impermeable barrier to cell migration. Leukocytes breach that barrier when they are recruited from the capillary for body defense, for example, against bacteria present in tissues. During pregnancy, the trophoblast invades the subendothelial basement membrane of the uterus, but this becomes reversible with expulsion of the placenta after birth. Metastatic cancer cells enter the circulation by breaching the basement membrane at the primary cancer site and breach it again upon entering the parenchyma of a distant tissue. This behavior is reminiscent of migrating cells during embryonic development, a stage when base-

ment membranes are not impermeable to cell penetration, in contrast to adult tissues.

Basement membranes are predominantly composed of: (1) type IV collagen, a nonfibrous polymer; (2) laminin, an adhesive glycoprotein; (3) fibronectin, another adhesive glycoprotein; (4) entactin (nidogen); (5) and perlecan, a proteoglycan.^{9–13,24,25} Small amounts of other proteoglycans and glycoproteins are also found in basement membranes. Ultrastructural and microscopic observations do not reveal obvious architectural differences, comparing basement membranes of various organs and tissues; however, immunostaining does reflect the genetic heterogeneity of both the collagen and laminin components. There are five known genes encoding for type IV collagen, providing molecular diversity in collagen chain composition. Laminin has a cruciform structure organized about three genetically separate chains, and each chain has isogenic forms; 12 different laminin molecules are generated from multiple genetically distinct chains. Both type IV collagen and laminin molecular isoforms are synthesized by cells in a tissue-specific manner, as reflected by distinct immunostaining patterns.

Reconstitution studies show that type IV collagen polymerizes to form a scaffold that is recognized by nidogen, laminin, and perlecan. The latter molecules bind to the scaffold, creating an interlocking network that becomes tough and insoluble. It serves as a barrier to large solutes and is impenetrable to most cells in adult tissues, save for the exceptions noted above.

Laminins

The prototypical laminin molecule was discovered in the mouse EHS sarcoma more than 20 years ago and was quickly shown to be abundant in basement membranes. The cruciform laminin molecule contains an alpha, beta, and gamma chain; to date there are five alpha chains, three beta chains, and three gamma chains, each encoded by individual genes. Twelve distinct laminin molecular isoforms have been characterized. They share certain domains, such as binding sites for nidogen and for integrins, which are transmembrane receptors that enable cells to recognize and respond to the ECM.

Individual laminins^{9,10,12} modulate certain cell functions. These functions include differentiation of epithelial cells, neurite outgrowth, anchoring of epidermal basal cells, and anchoring of leukocytes that are exiting from the microvasculature.

Laminin isoforms are produced by a wide variety of cells, primarily in a tissue-specific manner, as reflected by immunostaining patterns and confirmed by biochemical analyses of cell cultures that produce laminin.

Thus, epithelial cells, muscle (smooth, skeletal, cardiac) cells, nerve cells, endothelial cells, marrow cells, and neuroretinal cells all produce laminin. The secreted laminin becomes deposited with nidogen, perlecan, and type IV collagen in basement membranes.

Fibronectins

Fibronectins^{11,13,24} are multifunctional glycoproteins that play a central role in cell adhesion. They are heterodimeric molecules that contain three types of repeating domain, termed I, II, and III. More than half of the molecule consists of type III repeats. Fibronectin has binding domains that recognize fibrin, heparin, collagen, integrins, and other cell surface moieties.

Fibronectin molecular isoforms arise via alternative splicing of a single gene. Protein products are disulfide-linked heterodimers of 235–270 kDa that are decorated with *N*- and *O*-linked carbohydrates. Fibronectin is abundant in the blood and in most tissues, where it assembles into fibrils that enhance cell adhesion.

Integrins

A family of heterodimeric receptor molecules, integrins^{4,5} enable cells to adhere to the ECM. Integrins provide mechanical continuity between the interior and exterior of the cell. Integrins contain an alpha subunit and a beta subunit joined by disulfide bonds; currently there are 18 alpha and 3 beta subunits, which are organized into 24 distinct integrins. Ligand recognition is related to integrin subunit composition, but there is considerable ligand overlap between many integrins. Collagens, laminins, fibronectins, and leukocyte-specific surface moieties are ligands.

Interior linkage of integrins to the actin cytoskeleton is mediated via a repertoire of linker proteins including talin, α -actinin, vinculin, and paxillin. Integrin engagement by an ECM ligand triggers a signaling cascade that affects cell survival, cell motility, cell proliferation, and gene transcription. Cell migration necessarily requires that integrins at the leading edge of a motile cell be transiently attached to the ECM, whereas those at the trailing edge release their contacts as the cell moves. This attachment–detachment cycle has been clearly demonstrated in cell cultures in which intracellular actin has been vitally stained with green fluorescent protein (GFP) molecules. Similarly, single cell coalescence into multicellular cords and ribbons has been tracked in cell culture, demonstrating another level of complexity, that of cell–cell recognition.

Integrin molecular motion has been deduced by X-ray diffraction analyses of integrin-bound ligands. Con-

formational changes in the integrin structure occur in cyclic fashion, enabling binding and release of ligand, thereby accounting for the attachment–detachment phases of cell motility. Thus, integrin molecular movement is transformed into cellular motion. This conceptual framework provides insight into the phenomenon of leukocyte diapedesis, that is, the departure of leukocytes from the microvasculature into tissue parenchyma. A leukocyte that has attached to the basement membrane underlying the vascular endothelium is using its integrins for primary contact with the ECM; the leukocyte then elaborates matrix metalloproteases to provide a small puncture for cell entry into the underlying parenchyma. Directed leukocyte motion then occurs via the attachment–detachment mechanism of its surface integrins.

Similar considerations apply to cancer cell metastasis. Here, the invading cells depart from the primary cancer site, migrate to a vascular channel, puncture the endothelial basement membrane, enter the circulation, and then depart from the vascular channel in a manner corresponding to the leukocyte pathway. It is in this way that distant metastases develop in various organs.

Matrix metalloproteinases

Leukocyte diapedesis and cancer cell metastasis require these migrating cells to penetrate basement membranes. The cells can do so because they are armed with a repertoire of proteolytic enzymes, the matrix metalloproteinases (MMPs).^{14,15} Currently, more than 20 mammalian MMPs have been isolated and characterized. Although each enzyme has unique specificity, in combination the group can degrade all ECM proteins.

Most of the MMPs are secreted and released as inactive proenzymes that have a catalytic domain containing a zinc-bound active site consensus sequence. Some MMPs have additional molecular features that enhance their biological activities, such as adhesion to cell surfaces via their tandem fibronectin repeat sequences. Some members of the MMP family have transmembrane domains, ensuring their sustained proximity to the cell surface.

Metalloproteinases are produced by a wide variety of cells, normal and abnormal. Ordinarily, cells at rest exhibit low levels of MMPs, but when cytokines and growth factors stimulate cells during infection or tissue remodeling MMP levels rapidly increase. MMP expression is controlled in multiple ways, including transcription, proteolytic activation, tissue localization, and inhibitor binding. Activation ordinarily occurs via other proteases removing the prodomain, but autocatalytic activation may occur as well.

Secreted MMPs seem to localize transiently at the cell surface via interactions with cell surface adhesion receptors and proteoglycans. Localization may be needed for proper ECM degradation and may also protect MMPs from inhibitors, especially the tissue inhibitors of metalloproteinases (TIMPs). MMPs seem responsible for augmenting the release of bound growth factors from the ECM and may activate latent growth factors such as transforming growth factor β and α (TGF β , TGF α), insulin-like growth factor (IGF), and heparin-binding epidermal growth factor (HB-EGF). Finally, MMPs are known to cleave cell surface receptors that recognize growth factors, cytokines, and chemokines. In sum, MMPs are potent regulators of normal and abnormal cell physiology.

Future prospects

The explosion in biological information during the past 25 years has illustrated the great complexity of nature and has provided a rich tapestry that is just unfolding for the scientific community. Unlike an earlier era when great advances were made in deciphering fundamental laws of physics, biology seems limited by its very nature to rely on generalizations that often contain exceptions. Even such a key point as the “central dogma” of information flow from nucleic acids to proteins allows that RNA or DNA may be the repository of information, depending on a particular species and its genome. For example, some viral genomes are encoded in RNA, not DNA.

The great success in understanding the building blocks of cells and the ECM emerged when a reductionist strategy was used to identify, isolate, and characterize molecular components. The relation of such components to each other and to the genome is now under investigation using the strategy of systems biology.^{26–29} Here, the intent is to decipher the relations between components of pathways and between pathways. Thus, genomics, proteomics, glycomics, metabolomics, and signalomics are all legitimate “omics” under intense study; basically, such systems analyses will illustrate how molecular components interact to create dynamic functional activities. Recent examples are yeast systems analysis and systems analysis of early embryonic development. Some insight into the pathogenesis of osteoarthritis has been made by using similar systems analysis methods; detailed relationships were illustrated, comparing anabolic and catabolic pathways in normal and osteoarthritic articular cartilage.³⁰

Potentially, such a systems biology strategy can help decipher the pathogenesis of complex diseases and provide insight into tailoring medications that interfere

with critical pathways, thereby stopping disease progression. Ideally, early interception of disease states before significant irreversible damage occurs may become possible by developing tools that identify individuals at risk for specific diseases. Initially, genomic analyses can be implemented; later, as other tools such as proteomics and glycomics come on stream, they will provide important ancillary information. In this way, individual profiles can be made at one's request, similar to contemporary genetic testing to detect hereditary diseases. Progressive developments in technology will make this idea into reality in the near future, opening a new dimension in disease prevention.

References

- Har-el R, Tanzer ML. Extracellular matrix. 3. Evolution of the extracellular matrix in invertebrates. *FASEB J* 1993;7:1115–23.
- Hohenester E, Engel J. Domain structure and organisation in extracellular matrix proteins. *Matrix Biol* 2002;21:115–28.
- Thornton JM. From genome to function. *Science* 2001;292:2095–7.
- Hynes RO, Zhao Q. The evolution of cell adhesion. *J Cell Biol* 2000;150:F89–96.
- Mould AP, Humphries MJ. Regulation of integrin function through conformational complexity: not simply a knee-jerk reaction? *Curr Opin Cell Biol* 2004;16:544–51.
- Tanzer ML. Collagens and elastin: structure and interactions. *Curr Opin Cell Biol* 1989;1:968–73.
- Iozzo RV. Matrix proteoglycans: from molecular design to cellular function. *Annu Rev Biochem* 1998;67:609–52.
- Lander AD, Selleck SB. The elusive functions of proteoglycans: in vivo veritas. *J Cell Biol* 2000;148:227–32.
- Aumailley M, Smyth N. The role of laminins in basement membrane function. *J Anat* 1998;193:1–21.
- Colognato H, Yurchenco PD. Form and function: the laminin family of heterotrimers. *Dev Dyn* 2000;218:213–34.
- Mao Y, Schwarzbauer JE. Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biol* 2005;24:389–99.
- Sasaki T, Fassler R, Hohenester E. Laminin: the crux of basement membrane assembly. *J Cell Biol* 2004;164:959–63.
- Wierzbicka-Patynowski I, Schwarzbauer JE. The ins and outs of fibronectin matrix assembly. *J Cell Sci* 2003;116:3269–76.
- Brinckerhoff CE, Matrisian LM. Matrix metalloproteinases: a tail of a frog that became a prince. *Nat Rev Mol Cell Biol* 2002;3:207–14.
- Mott JD, Werb Z. Regulation of matrix biology by matrix metalloproteinases. *Curr Opin Cell Biol* 2004;16:558–64.
- Myllyharju J, Kivirikko KI. Collagens and collagen-related diseases. *Ann Med* 2001;33:7–21.
- Franzke CW, Bruckner P, Bruckner-Tuderman L. Collagenous transmembrane proteins: recent insights into biology and pathology. *J Biol Chem* 2005;280:4005–8.
- Zelzer E, Olsen BR. The genetic basis for skeletal diseases. *Nature* 2003;423:343–8.
- Tanzer ML, Waite JH. Collagen cross-linking. *Coll Relat Res* 1982;2:177–80.
- Canty EG, Kadler KE. Procollagen trafficking, processing and fibrillogenesis. *J Cell Sci* 2005;118:1341–53.
- Kadler K. Matrix loading: assembly of extracellular matrix collagen fibrils during embryogenesis. *Birth Defects Res C Embryo Today* 2004;72:1–11.
- Myllyharju J, Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* 2004;20:33–43.
- Tuan RS. Biology of developmental and regenerative skeletogenesis. *Clin Orthop Relat Res* 2004:S105–17.
- Hynes RO. The dynamic dialogue between cells and matrices: implications of fibronectin's elasticity. *Proc Natl Acad Sci USA* 1999;96:2588–90.
- Patarroyo M, Tryggvason K, Virtanen I. Laminin isoforms in tumor invasion, angiogenesis and metastasis. *Semin Cancer Biol* 2002;12:197–207.
- Davidson EH, McClay DR, Hood L. Regulatory gene networks and the properties of the developmental process. *Proc Natl Acad Sci USA* 2003;100:1475–80.
- Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. *Science* 2004;306:640–3.
- Kitano H. Systems biology: a brief overview. *Science* 2002;295:1662–4.
- Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res* 2004;3:179–96.
- Aigner T, Bartnik E, Sohler F, Zimmer R. Functional genomics of osteoarthritis: on the way to evaluate disease hypotheses. *Clin Orthop Relat Res* 2004:S138–43.