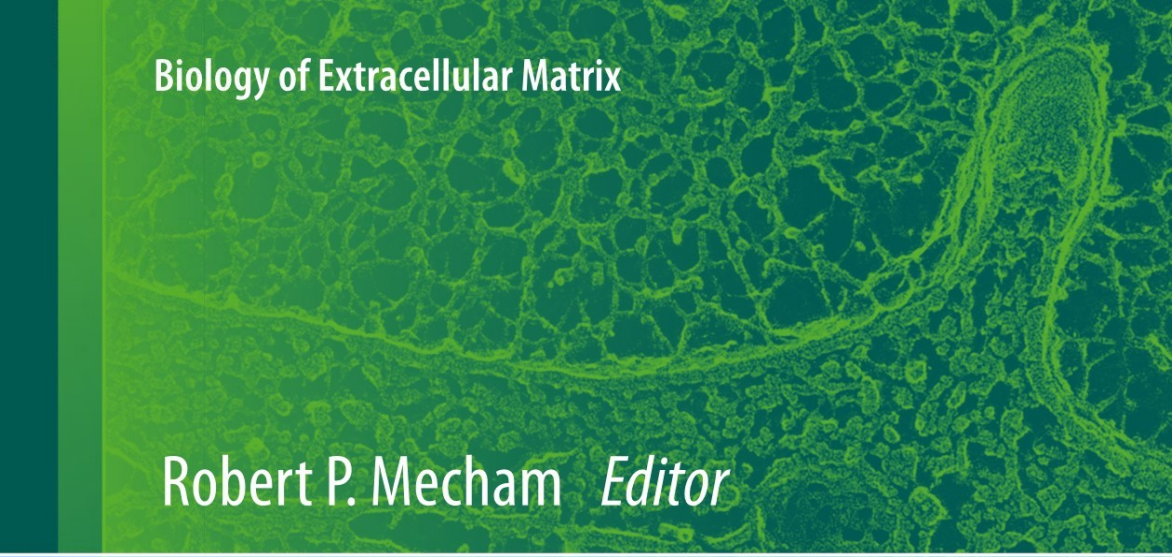

Biology of Extracellular Matrix

Robert P. Mecham *Editor*

The Extracellular Matrix: an Overview

Springer

Biology of Extracellular Matrix

For further volumes:
www.springer.com/series/8422

Robert P. Mecham
Editor

The Extracellular Matrix: an Overview

Springer

Editor
Professor Robert P. Mecham
Washington University School of Medicine
Department of Cell Biology & Physiology
660 South Euclid Ave., Campus Box 8228
St. Louis, MO 63110
USA
bmecham@wustl.edu

ISSN 0887-3224 e-ISSN 2191-1959
ISBN 978-3-642-16554-2 e-ISBN 978-3-642-16555-9
DOI 10.1007/978-3-642-16555-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011920814

Cover design: deblik, Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Series

The first volume of the *Biology of Extracellular Matrix* series was published in 1986 and was titled "Regulation of Matrix Accumulation." Twelve volumes in the series were published over a period of 12 years and each volume provided timely reviews on current topics of ECM biology. With the contraction of the publishing industry in the late 1990s, Academic Press, the former Series publisher, was purchased by Elsevier and they decided to discontinue most of their monograph series, including the *Biology of Extracellular Matrix*. I was able to retain the rights to the series title and was pleased when Springer agreed to resume publication. The volume "The Extracellular Matrix: An Overview," Robert P. Mecham (Editor), is the first under the new publisher. It should also be noted that the series is being published in collaboration with the American Society for Matrix Biology.

The Study of Extracellular Matrix Biology

Over the years, our understanding of extracellular matrix (ECM) function has evolved from the early concept of a static "connective tissue" that ties everything together to one of a dynamic biomaterial that provides strength and elasticity, interacts with cell-surface receptors, and controls the availability of growth factors. There is now no question that ECM is an important part of cell biology, and to understand cellular differentiation, tissue development, and tissue remodeling requires an in-depth consideration of the ECM components that are produced by the cell. As we look back through the relatively short history of ECM biology, we find that the field was first dominated by biochemistry (mostly chemistry!) where investigators were trying to isolate and identify the individual ECM components. The proteins that were identified were, indeed, unique in their structure, composition, and function and were unlike other proteins in living cells. The ECM is designed to function as homo- and heteropolymers that are generally insoluble in their mature state. They also have relatively long half-lives compared with other proteins in the body. Some contain unique cross-links, some have high amounts of

sulfated polysaccharides, some are designed to be "sticky" in terms of interacting with cells, and others form complex adhesion surfaces and diffusion barriers between different cell layers. In all cases, however, each class of ECM molecule is designed to interact with another to produce the unique physical and signaling properties that support tissue structure, growth, and function.

This brings us to the field today, where the tools of cell and molecular biology together with the power of model organism genetics allow us to focus on the functional complexities of the ECM biopolymer. Constructing a complex, mechanically appropriate matrix requires the cell to know the instructions for assembly, to have knowledge of the available building materials, and be able to interpret information about the stresses that the final material will have to endure. In this regard, it is clear that cells are adept at reading the instructive signals from the microenvironment, and changing the mix of matrix proteins needs to be added at any particular instance. While there is still a need to use biochemistry to characterize the individual ECM components, to fully understand the ECM requires a fundamental knowledge of cell biology. We need to understand, for example, the cellular mechanisms that lead to coordinated expression, both temporally and spatially, of complex sets of genes that encode ECM proteins as well as the enzymes responsible for their secretion and assembly. Building a functional collagen fiber, for example, involves activating and regulating genes for collagen alpha chains, hydroxylating enzymes, proteases to process propeptide regions, lysyl oxidases for cross-linking, and other chaperones and assembly proteins. Similar complexities are involved in the processing and assembly of most ECM networks, including basement membranes, elastic fibers, and large proteoglycan matrices. Virtually, all fields of animal and plant biology are concerned with questions of extracellular matrix in some manner. It is my hope that this series will prove helpful to all those seeking an introduction to EMC biology as well as experienced ECM investigators who are interested in greater insight into ECM function. In the preface to volume one of this series over 2 decades ago, I pointed out that the series cannot thrive without a large measure of enthusiasm and active participation from the ECM community. I welcome your suggestions of topics for future volumes and look forward to your feedback as we explore the extracellular matrix.

St. Louis MO, USA Robert P. Mecham

Preface to the Volume

The objective of this overview volume in the "Biology of Extracellular Matrix" series is to update and build upon topics discussed in previous volumes in this series as well as in classic ECM review texts, such as Betty Hay's *Cell Biology of Extracellular Matrix*. The first chapter by Jürgen Engel and Matthias Chiquet is the ideal introduction to ECM biology. It provides an overview of ECM structure and function in a creative and insightful interpretation of the ECM as a complex "machine." The authors outline the basic features of the major classes of ECM components and describe how their multidomain structure allows multiple functions to be combined in one, often large molecule that is engineered to undergo multimeric assembly and extended multimolecular networks. They show how ECM components together with their cell surface receptors can be viewed as intricate nanodevices that allow cells to physically organize their 3D environment, as well as to sense and respond to various types of mechanical stress. They also make the point that metazoan evolution would not have been possible without the concomitant expansion of ECM complexity. Using examples of phylogenetically "old" versus "young" ECM protein families, they review the evidence that today's incredible diversity of ECM components arose from the recombination of preexisting protein modules by exon shuffling during evolution.

The second chapter focuses on fibronectin and other glycoproteins that mediate cell adhesion through interactions with integrins. Jielin Xu and Deane Mosher use an in-depth analysis of fibronectin as a prototype to illustrate how the domain organization of adhesive glycoproteins is structured to bridge interactions between cells (integrin binding domains) and other components of the ECM, including collagen, heparan, and fibrin. They also discuss fibronectin assembly and the importance of integrins and the cellular cytoskeleton in this process. Other glycoproteins discussed in this chapter include vitronectin, the laminins, thrombospondins, tenascins, entactins, nephronectin, and fibrinogen. A short section on integrin signaling is also included.

Of all of the ECM proteins, few are as old as collagen. In early parazoa (like sponges), cells are embedded in an ECM consisting mainly of fibrillar collagens not

unlike those of higher animals. In Chap. 3, David Birk and Peter Brückner bring us up-to-date on collagen types and collagen fibril assembly. There are 28 different types of collagen in vertebrates (many more in invertebrates) that assemble into a variety of supramolecular structures including fibrils, microfibrils, and network-like structures. This chapter begins with a general discussion of collagen molecules and their supramolecular structure, assembly, and function within extracellular matrices. One of the more interesting aspects of collagen biology as outlined in this chapter is the description of mechanistic principles involved in the assembly of collagen-containing suprastructures. This includes the characterization of tissue-specific collagen fibrillogenesis, which serves to generate the diversity in extracellular matrix structures and functions required for individual tissue function.

As multicellular organisms evolved and grew more complex, there arose a need during development to separate polarized epithelial cells from underlying mesenchymal cells. This separation process, i.e., gastrulation, would not be possible without the appearance of the basement membrane – a unique ECM structure that combines the structural rigidity and unique basket-weave-forming properties of collagen type IV with cell-adhesive proteins (e.g., laminins) and charged proteoglycans (e.g., perlecan and agrin). The chapter on basement membranes by Jeffrey Miner summarizes our current knowledge about the basement membrane components and their receptors on cells. Basement membrane assembly is also discussed along with a number of human genetic diseases caused by mutations that affect basement membrane components.

The discussion of proteoglycans is separated into two chapters. The first, Chap. 5, authored by Thomas Wight, Bryan Toole, and Vincent Hascall, focuses on hyaluronan and the large aggregating proteoglycans. This family includes aggrecan, versican, neurocan, and brevican. These proteoglycans form macromolecular complexes with hyaluronan and contribute to the structural and mechanical stability of different tissues. Considerable evidence suggests that the large hydrodynamic space occupied by glycosaminoglycan chains influences tissue turgidity and viscoelasticity. In addition, recent data point to a prominent role for these ECM structures in direct cell signaling as well as an ability to bind and sequester growth factors and morphogens that are important for cell movement and differentiation. The chapter also contains a description of new functions mapped to the proteoglycan core protein.

The small leucine-rich proteoglycans (SLRPs) are discussed in Chap. 6 by Renato Iozzo, Silvia Goldoni, Agnes Berendsen, and Marian Young. SLRPs serve as tissue organizers by orienting and ordering various collagenous matrices during ontogeny, wound repair, and cancer. They also interact with a number of surface receptors and growth factors thereby regulating cell behavior. The focus of this chapter is on novel conceptual and functional advances in our understanding of SLRP biology with special emphasis on genetic diseases, cancer growth, fibrosis, osteoporosis, and other biological processes where these proteoglycans play a central role.

One of the newest ECM structures to be described and characterized, but among the oldest ECM structures in evolution, is the microfibril. The core elements of

these 10–15 nm filaments are the fibrillins – large cysteine-rich proteins that can be found as far back in evolution as the placozoans and, perhaps, parazoans. First described as components of elastic fibers, microfibrils are now known to be important regulators of growth factor signaling through their ability to bind and sequester growth factors, particularly TGF-β family members. In Chap. 7, Dirk Hubmacher and Dieter Reinhardt provide an overview of the structure, assembly, and functions of fibrillins and microfibrils as well as the pathobiology associated with genetic aberrations in the microfibril system.

Vertebrate evolution would not have been as successful as it was without elastin. As the name implies, elastin imparts elasticity to tissues, particularly large blood vessels and the lung. Without elastic vessels, it would not be possible to evolve an efficient closed, pulsatile circulatory system that supports efficient distal perfusion and body growth. Similarly, the mechanical function of the vertebrate lung would not be possible without elastin. Beth Kozel, Robert Mecham, and Joel Rosenbloom discuss this unique, highly cross-linked protein in Chap. 8. Emphasis is given to how the protein works as an elastomer and why damage to elastic fibers is so detrimental to tissue integrity and overall longevity. Diseases linked to mutations in the elastin gene are discussed, as are animal models of these diseases.

Collagen and elastin function is a polymer where individual chains are cross-linked one to another via modified lysine residues. The enzyme responsible for initiating the cross-linking reaction is one or more members of the lysyl oxidase family. These copper-requiring enzymes catalyze the oxidative removal of lysine epsilon-amino groups to form a reactive aldehyde, the cross-link precursor. There are five known members of this amine oxidase family (lysyl oxidase and 4 lysyl oxidase-like enzymes), and in Chap. 9, Herbert Kagan and Faina Ryfkin provide a detailed analysis of the amino oxidase mechanism of lysyl oxidase and bring us up-to-date on the known functions of the individual family members. They also review evidence showing that LOX can function both as an anti-oncogenic agent as well as an enhancer of malignancy in selected cancerous conditions.

Fibulins are a family of proteins that share a common architectural signature, namely a series of epidermal growth factor (EGF)-like modules followed by a carboxy terminal fibulin-type module. Over the last few years, the biological role of the fibulins has become clearer as new members of the family were identified and knockout mice provided insight into fibulin function. In Chap. 10, Marion Cooley and Scott Argraves review the current understanding of structure–function relationships for the fibulins, particularly with regards to elastogenesis. They also discuss the role that fibulins play in diseases such as cancer, cardiovascular disease, and eye disease.

In the final chapter of the volume (Chap. 10), David Roberts and Lester Lau provide an extensive review of a class of extracellular matrix components referred to as “matricellular proteins.” These proteins, in general, share a complex modular structure that enables them to interact with specific components of the matrix while engaging specific cell surface receptors through which they control cell behavior. Matricellular proteins, including the thrombospondins, some thrombospondin-repeat superfamily members, tenascins, SPARC, CCN proteins, and SIBLING

proteins, are increasingly recognized to play important roles in inherited disorders, responses to injury and stress, and the pathogenesis of several chronic diseases of aging.

What Is Not Included and Plans for the Future

Trying to review the entirety of extracellular matrix in one volume is an impossible task. For this reason, I have chosen to focus this first volume on the major molecules that make up the ECM. Subsequent volumes that are either in production or in the planning stages include ECM turnover, glycoprotein biology, integrins and receptors for ECM, and volumes devoted to topics such as ECM in development and the role of ECM in specific diseases. It is hoped that this "overview" volume will be used as a basis of reference as we explore ECM function more deeply in subsequent publications.

St. Louis MO, USA Robert P. Mecham

Contents

1 An Overview of Extracellular Matrix Structure and Function 1
Jürgen Engel and Matthias Chiquet

2 Fibronectin and Other Adhesive Glycoproteins 41
Jielin Xu and Deane Mosher

3 Collagens, Suprastructures, and Collagen Fibril Assembly 77
David E. Birk and Peter Brückner

4 Basement Membranes 117
Jeffrey H. Miner

5 Hyaluronan and the Aggregating Proteoglycans 147
Thomas N. Wight, Bryan P. Toole, and Vincent C. Hascall

6 Small Leucine-Rich Proteoglycans 197
Renato V. Iozzo, Silvia Goldoni, Agnes Berendsen,
and Marian F. Young

7 Microfibrils and Fibrillin 233
Dirk Hubmacher and Dieter P. Reinhardt

8 Elastin 267
Beth A. Kozel, Robert P. Mecham, and Joel Rosenbloom

9 Lysyl Oxidase and Lysyl Oxidase-Like Enzymes 303
Herbert M. Kagan and Faina Ryvkin

10 The Fibulins 337
Marion A. Cooley and W. Scott Argraves

11 Matricellular Proteins 369
David D. Roberts and Lester F. Lau

Index 415

Contributors

W. Scott Argraves Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, 173 Ashley Avenue, Charleston, SC 29425, USA, argraves@musc.edu

Agnes Berendsen Craniofacial and Skeletal Diseases Branch, NIDCR, National Institutes of Health, Bethesda, MD, USA

David E. Birk Department of Pathology and Cell Biology, University of South Florida, College of Medicine, 12901 Bruce B. Downs Boulevard, MDC11, Tampa, FL 33612-4799, USA, dbirk@health.usf.edu

Peter Brückner Institute for Physiological Chemistry and Pathobiochemistry, University Hospital of Münster, Waldeyerstr. 15, 48149 Münster, Germany, peter.bruckner@uni-muenster.de

Matthias Chiquet Department of Orthodontics and Dentofacial Orthopedics, University of Bern Freiburgstrasse 7, CH-3010, Bern, Switzerland, matthias. chiquet@zmk.unibe.ch

Marion A. Cooley Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, 173 Ashley Avenue, Charleston, SC 29425, USA

Jürgen Engel Department of Biophysical Chemistry, Biocenter, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland, Juergen.engel@ unibas.ch

Silvia Goldoni Department of Pathology, Anatomy and Cell Biology and the Cancer Cell Biology and Signaling Program, Kimmel Cancer Center, Thomas Jefferson University, Room 249 JAH, 1020 Locust Street, Philadelphia, PA 19107, USA

Vincent C. Hascall Department of Biomedical Engineering, Cleveland Clinic Foundation, Cleveland, OH, USA

Dirk Hubmacher Department of Anatomy and Cell Biology, Faculty of Medicine, McGill University Montreal, Montreal, QC, Canada H3A 2B2

Renato V. Iozzo Department of Pathology, Anatomy and Cell Biology and the Cancer Cell Biology and Signaling Program, Kimmel Cancer Center, Thomas Jefferson University, Room 249 JAH, 1020 Locust Street, Philadelphia, PA 19107, USA, iozzo@mail.jci.tju.edu

Herbert M. Kagan Department of Biochemistry, Boston University School of Medicine, 715 Albany Street, Boston, MA 02118, USA, kagan@biochem.bumc.bu.edu

Beth A. Kozel Department of Pediatrics, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, USA

Robert P. Mecham Department of Cell Biology and Physiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, USA, bmecham@wustl.edu

Jeffrey H. Miner Renal Division, Washington University School of Medicine, Box 8126, 660 S. Euclid Avenue, St. Louis, MO 63110, USA, minerj@wustl.edu

Deane Mosher Department of Biomolecular Chemistry, University of Wisconsin, 4285 MSC, 1300 University Avenue, Madison, WI 53706, USA, dfm1@medicine.wisc.edu

Dieter P. Reinhardt Department of Anatomy and Cell Biology, Faculty of Medicine, McGill University Montreal, Montreal, QC, Canada H3A 2B2, dieter.reinhardt@mcgill.ca

Joel Rosenbloom Thomas Jefferson University, Philadelphia, PA, USA

Faina Ryvkin Emmanuel College, 400 The Fenway, Boston, MA 02115, USA, ryvkin@emmanuel.edu

Bryan P. Toole Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, Charleston, SC, USA

Thomas N. Wight The Hope Heart Matrix Biology Program, Benaroya Research Institute at Virginia Mason, 1201 Ninth Avenue, Seattle, WA, USA, twight@benaroyaresearch.org

Jielin Xu Department of Biomolecular Chemistry, University of Wisconsin, 4285 MSC, 1300 University Avenue, Madison, WI 53706, USA

Marian Young Craniofacial and Skeletal Diseases Branch, NIDCR, National Institutes of Health, Bethesda, MD, USA

Chapter 1
An Overview of Extracellular Matrix Structure and Function

Jürgen Engel and Matthias Chiquet

Abstract Extracellular matrix (ECM) not only provides a stable framework for maintaining the shape of multicellular organisms under physical load and during locomotion, but it is also essential for their morphogenesis and regenerative capacity. In this introductory chapter, we describe the basic features of the major classes of ECM components, namely, collagens, glycoproteins, and proteoglycans. We emphasize their multidomain structure that allows multiple functions to be combined in one, often large molecule. Of the many types of protein modules found in ECM components, some are devoted to multimeric assembly, and hence, for their crucial ability to form extended multimolecular networks or matrices. We argue that ECM components together with integrin receptors on the cell surface can be viewed as intricate nanodevices that allow cells to physically organize their 3D environment, as well as to sense and respond to various types of mechanical stress. In addition, ECM functions as part of a cell-controlled machinery to store and activate growth factors during development. We also make the point that metazoan evolution would not have been possible without the concomitant expansion of ECM complexity. Using examples of phylogenetically "old" versus "young" ECM protein families, we review the evidence that today's incredible diversity of ECM components arose from the recombination of preexisting protein modules by exon shuffling during evolution.

J. Engel
Department of Biophysical Chemistry, Biocenter, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland
e-mail: Juergen.engel@unibas.ch

M. Chiquet (✉)
Department of Orthodontics and Dentofacial Orthopedics, University of Bern Freiburgstrasse 7, CH-3010, Bern, Switzerland
e-mail: matthias.chiquet@zmk.unibe.ch

R.P. Mecham (ed.), *The Extracellular Matrix: an Overview*, Biology of Extracellular Matrix, 1
DOI 10.1007/978-3-642-16555-9_1,

1.1 Introduction: No Metazoans Without Extracellular Matrix

> Were the various types of cells to lose their stickiness for one another and for the supporting extracellular white fibers, reticuli, etc., our bodies would at once disintegrate and flow off into the ground in a mixed stream of ectodermal, muscle, mesenchyme, endothelial, liver, pancreatic, and many other types of cells. (W. H. Lewis, 1922)

Extracellular matrix (ECM) is the term for the organic matter that is found between most cells in plants (Dhugga 2001) and animals (Hay 1991). ECM glues together the cells and organs of multicellular organisms. It is also essential for their morphogenesis, and later provides a stable framework for tissues that is required to maintain shape under gravity and other physical loads (Alberts et al. 2002). Without ECM, multicellular life could not exist, plants would not grow tall, and animals could neither swim, nor walk or fly. Given the importance of ECM for the integrity of complex organisms, it is not surprising that the evolution of its components is closely linked to the phylogeny of plants and animals themselves (Exposito et al. 2002; Hynes and Zhao 2000; Muller et al. 2004; Adair and Mecham 1990). Life on earth started about 4 billion years ago, and after at least 3 billion years of single cell life forms, multicellularity presumably evolved gradually from colony-forming unicellular organisms (Furusawa and Kaneko 2002; Muller et al. 2004). It is difficult to date the dawn of true multicellular organisms. An important step in evolution was certainly the emergence within the same cell colony of various types of specialized, mortal somatic cells next to the potentially immortal stem and germ line cells (Sanchez Alvarado and Kang 2005). Eventually, this division of labor resulted in higher order organisms whose isolated parts could no longer survive on their own. Multicellular individuals with their own identity evolved that were born, propagated, and died. In their body plan, some of the first multicellular species might have resembled today's most primitive green algae, e.g., Volvox globator (Hallmann 2003). Volvox is a small sphere of translucent ECM with identical, ciliated somatic cells on its surface, and with centrally located germ cells. The ECM of Volvox is well structured but very divergent from that of today's plants and animals, and the same might have been the case for the first multicellular species during evolution.

The plant and animal kingdoms have been separated about a billion years ago, and this event is likely to coincide with the birth of two types of "modern" ECM. The sessile algae, fungi, and plants acquired a comparatively uniform ECM in their cell walls (Dhugga 2001) that consists primarily of various long-chain polysaccharides (e.g., cellulose in plants) with few, albeit important, proteins (Brownlee 2002). This contrasts with the more sophisticated and diverse ECM of animals, which is the theme of this book. About 450 million years ago, in an evolutionarily short time called the "Cambrian explosion" (Couso 2008; Cummings 2006), essentially all of the animal phyla appeared on earth that live today, together with many more that have long vanished again and for which only fossil records are available. Evolution of metazoan animals, for which motility is a way of life, rapidly generated an

incredible diversity of body plans. This was possible because entire gene families became devoted to intercellular adhesion and communication, as well as to the construction of very intricate ECM networks with diverse structure and function (Hynes and Zhao 2000; Whittaker et al. 2006). Obviously, the engagement of many genes allows for more complex regulation of ECM structure in metazoans than that found in modern plants. An elaborate gene network controls assembly and remodeling, as well as physical organization and functional properties of ECM in the various animal tissues. Although all metazoan ECM has highly preserved structural features and is made from the same classes of molecules (Exposito et al. 2002; Whittaker et al. 2006), namely, collagens, glycoproteins, and proteoglycans (see Sect. 1.2), it can have the material properties of a soft gel (e.g., in the vitreous body of our eyes), a polymer fiber rope (as in tendons and ligaments), or a rock-hard composite (e.g., in our bones) (Alberts et al. 2002). In terms of functions, animal ECM covers an entire spectrum, from maintaining body shape, to sustaining large mechanical stresses during motions, to acting as an instructive environment for the adhesion, growth, and differentiation of cells and organs (Adams and Watt 1993). By providing a mechanically stable yet permanently reconstructing framework, metazoan ECM is indispensable for both embryonic development and tissue remodeling in the adult.

1.2 Building Blocks of Extracellular Matrix

1.2.1 Basic Features of Extracellular Matrix Proteins

The ECM is formed by a large variety of proteins with different structures and functions but some common features are apparent. Many proteins of the ECM are very large. To their size contributes an often extensive glycosylation, which is on average 35 weight%, and in the case of proteoglycans the covalent attachment of glycosaminoglycan (GAG) chains. Molar masses of 100–1,000 kDa are frequent and even larger proteins are known. In general, ECM proteins are highly asymmetric in shape.

All ECM proteins are multidomain proteins, in which different or equal domains are arranged in a specific domain organization. Domains are defined as homologous units. The homology follows from amino acid sequence comparisons. In many cases, structures of domains are known at atomic resolution, which provides a more sensitive detection of structural homology. Individual domains may have distinct functions even after fragmentation from the intact protein. Even homologous domains may have sufficiently large sequential and structural differences to show rather different functions. The combination of different domains leads to a multifunctionality of essentially all ECM proteins. Commonly, several domains in a protein act in a concerted fashion. It was also observed that domains interact with each other in a given multidomain protein and form new functional entities in this way. The multifunctionality and the expanded shapes provide the potential for

lateral interactions, favoring the formation of fibers and other supramolecular assemblies of ECM proteins.

ECM proteins are normally grouped as glycoproteins, proteoglycans, and collagens. Proteoglycans contain long, charged glycosaminoglycan chains covalently attached to serines or threonines of the core protein. Some GAG chains are also found unconnected to a protein, e.g., hyaluronan. Collagens are defined as glycoproteins or proteoglycans with one or more collagenous domains. The latter consist of segments with a repeating $(GXY)_n$ sequence. Three chains with such sequences combine to a collagen triple helix.

1.2.2 Domains in Extracellular Matrix Proteins

As mentioned earlier, domains are defined as homologous protein units. Homology means that the domains have a common precursor. This is a yes or no decision and the often used phrase of a percentage of homology is therefore meaningless. It should be replaced by percentage of identity.

It is often difficult to define a homologous group. Comparing sequences, the range of 25% sequence identity and below is a twilight zone, in which it is difficult to decide on the existence of homology (Doolittle 1992). A comparison of 3D-structures is a more sensitive way to detect homology. Definitions of domains are done with some ambiguity. In particular, differently defined domains may have a distant common origin and may belong to the same homology group. It is mentioned in section 1.4.1 that the creation of completely new folds was a rare event in evolution. Consequently, the number of homologous groups without a common origin should be small.

In spite of the mentioned ambiguities, the domain concept has a large practical value for grouping and comparing different ECM proteins. Peer Bork and Amos Bairoch, the inventors of the SwissProt database, pioneered the domain concept (reviewed in Bork et al. 1996). Today large lists of domains can be found in databases like PROSITE, SMART, CDART also called CDSEARCH, PFAM, and others. These databases are continuously updated. In the last issue of the SMART database, about 850 domains are listed, of which 250 are found in extracellular proteins. Numbers of domains are even larger in PFAM and CDSEARCH. The databases can be used to display the domain organization of a protein of known sequence (Adams and Engel 2007). It is also possible to list all proteins, which have similar organizations or which contain a given set of domains. In addition, the databases guide to the three-dimensional structures of domains, in case such structures were solved by crystallography or by NMR-spectroscopy. A list of important domains frequently found in ECM proteins is given in Table 1.1.

Most domains are of globular shape and have a defined size. For many of the domains, the three-dimensional structure at atomic resolution was elucidated by X-ray crystallography or NMR-spectroscopy. Domain structures are reviewed

Table 1.1 Domains occuring in ECM Proteins

Letter code	Full name	Size (aa)	ECM proteins with homologues	Frequent function
CA	Cadherin domain	110–130	E-cadherin, N-cadherin, desmoglein, many other adhesion proteins	Homo association
C4	Collagen IV carboxy terminal domain	Trimer (3×110)	Collagen IV	Hexamer formation
CUB	CUB	About 110	BMP-1, Tolloid, neuropilin, many complement components	
EGF	Epidermal growth factor domain	About 50	Agrin, BMP-1, CASP, CMP, TSP-1 to -5, tenascins, many others	One of the most abundant domains with many functions
EF	EF-hand domain	12 Flanked by α-helices	BM-40/SPARC and very many cytosolic proteins	Binds Ca^{2+} and other divalent ions
KU	Kunitz inhibitor domain	About 60	α3 chain collagen VI, α1 chain collagen VII and many protease inhibitors	No protease inhibition in the collagens
F1	Fibronectin type 1 domain	About 40	Fibronectin and many coagulation factors	
F2	Fibronectin type 2 domain	About 60	Fibronectin and many coagulation factors	Involved in collagen binding
F3	Fibronectin type 3 domain	About 90	Fibronectin, tenascin, α1-chain of collagen I	RGD-loop binds to several integrins
FBG	Fibrinogen C-terminal domain	About 225	Tenascin, ficolin, angiopoitin, fibrinogen	
Ig	IG-like domain	70–100	NCAM, FGFR most abundant in IgGs and MHC	
LE	Laminin-type EGF-like domain	About 60	Laminin α-, β-, γ-chain, agrin, perlecan, unc-52, netrin	Specialized domains bind entactin/ nidogen
LN	Laminin N-terminal domain		Laminin α-, β-, γ-chain, netrin	Involved in basement membrane assembly
LG	Laminin G-like domain	About 190	Laminin α-chain, agrin, neurexin, slit protein	In some cases binding to integrin α6β1
TB	TGF-beta binding domain		TGF-beta binding protein, fibrillin-1,-2,-3, follistatin	

(*continued*)

Table 1.1 (continued)

Letter code	Full name	Size (aa)	ECM proteins with homologues	Frequent function
TSP1	Thrombospondin type 1 domain	About 55	Thrombospondin-1,-2,-3,-4 ADAMTS, properdin	
TSP2	Synonymous to EGF			
TSP3	Thrombospondin type 3 domain	Composed of many EF-hand like repeats	Thrombospondins 1–5	Binding of Ca-ions
TSPC	C-terminal L-lectin-like thrombospondin domain		Highly conserved domain of all thrombospondins	
VWFA	Von Willebrand factor A domain, I-domain of integrin	About 190	Present in 22 ECM proteins including VWF, collagens VI, VII, XII, XIV, matrilin, integrin α-chain	Binding of specific sites of some collagens
VWFC	Von Willebrand factor C domain	About 70	VWF, thrombospondins-1 and -2, chordin	
VWFD	Von Willebrand factor D domain		VWF, BMP-binding regulator protein	

by Bork et al. (1996) and Hohenester and Engel (2002). Detailed information is contained in the RCSB Protein Data Bank.

Many types of folds can be distinguished and the three-dimensional structures reveal interesting details. For example, the VWFA domain has six α-helices connected by parallel β-strands. The C-terminal and N-terminal are very close, which directs the repeat of several domains in a polypeptide chain. The collagen-binding site was localized by cocrystallization of an inhibiting Fab–antibody fragment (Romijn et al. 2001). The F3 domain is a β-sheath protein of an immunoglobulin-like fold. Its N- and C-termini are at opposite ends of the molecule making it suitable for a linear arrangement of repeating domains. The RGD site of the cell-binding domain of fibronectin, which binds to integrin α5β1, is located in a flexible surface exposed loop. In general, binding sites are located at the surface of domains. Erroneous assignments of binding sites were frequently corrected by the elucidation of 3D-structures.

Three-dimensional structures were also obtained for fragments containing several globular domains. As an example, the structure of a fibronectin fragment FN1-FN2-FN2 revealed noncovalent interactions between the first and third domain (Pickford et al. 2001) demonstrating that the traditional "pearls on a string" representation of multidomain proteins is woefully inadequate. Interactions between domains are important for many recognition processes and may even force the polypeptide to fold back on itself.

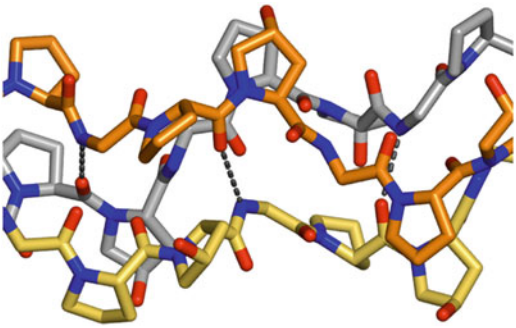

Fig. 1.1 Structure of the collagen triple helix as demonstrated for the model peptide (Gly-Pro-Hyp). Each polypeptide chain forms a left-handed polyproline-II-type structure. These are not stable on their own but three chains (colored *orange*, *yellow*, and *gray*) wind to a right-handed superhelix. The entire structure is stabilized by hydrogen bonds between the NH groups of glycines and the CO groups of hydroxyproline in a neighboring chain. N and O are marked in *blue* and *red*, respectively. An important stabilization originates from the frozen N–C bonds in the imino acid residues, favoring the polyproline-II-helix and reducing the stability of the unfolded chains. In addition the OH of hydroxyproline has a stabilizing function, probably by inductive effects of its dipole on ring conformation. Note that all side chains [in case of $(Gly\text{-}Pro\text{-}Hyp)_n$ the proline rings] are pointing out of the triple helix and can therefore interact with neighboring molecules in collagen fibrils and other assembly forms. For details on structure and stabilization see Bächinger and Engel (2005)

In addition to the globular domains discussed so far, multidomain proteins also contain structures of a different type with an elongated rod-like shape. Examples are the collagen triple helix (Fig. 1.1) and the α-helical coiled-coil structure (Fig. 1.2).

The collagen triple helix is formed by three polypeptide chains with the repeating sequence $(Gly\text{-}X\text{-}Y)_n$, in which proline occurs frequently in the X-position and 4(R)hydroxyproline in the Y-position. Each chain forms a left-handed polyproline type II helix and the three helices intertwine to form a right-handed super helix. The translation per residue is about 0.29 nm. The triple helical structure is stabilized by hydrogen bonds between Gly on one chain and Pro in the X-position of a neighboring chain and by the sterical constrains of the proline rings. For the formation of a regular triple helix, it is essential that Gly residues should repeat in every third position. Only these residues fit into the center of the triple helix and any side chain larger than H would destabilize it (Fig. 1.1). 4-(R)hydroxyproline in the Y-position causes a strong additional stabilization, which originates from the inductive effect of the OH-group on ring puckering (Vitagliano et al. 2001). Inductive effects are well known in organic chemistry and cause electron withdrawals in the ring structure. Importantly, the chains in the collagen triple helix are staggered by one residue. This gives rise to two stereoisomers A-B-C and A-C-B in which the B chain is either staggered by one or by two residues against the A chain in the circular arrangement of the three chains. Many collagens have two or three different chains and in these cases, the surface of the triple helix is very different in the

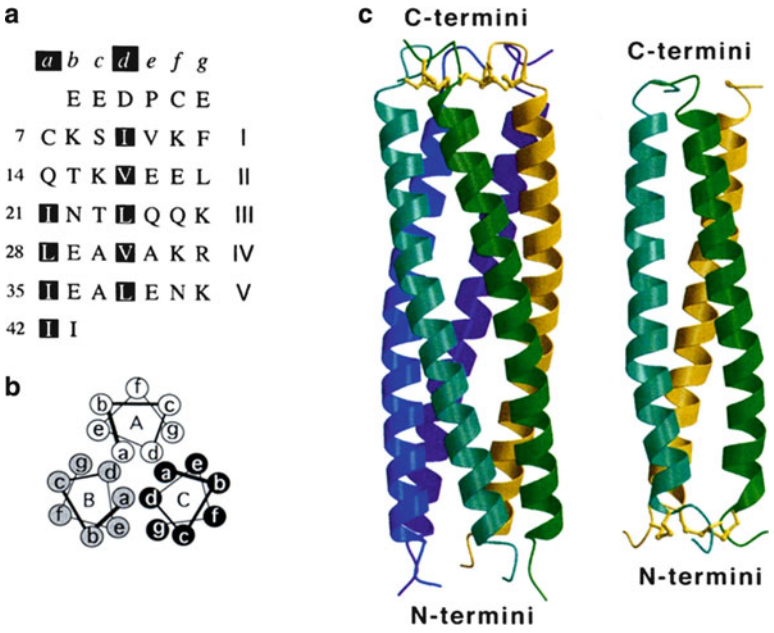

Fig. 1.2 Schematic representation of α-helical coiled coil structures. In (**a**) the sequence of the coiled-coil domain of matrilin-1 (also called cartilage matrix protein) is broken up in five successive heptad repeats *abcdefg* in which residues with hydrophobic side chains occur predominantly in the positions *a* and *d*. This sequence region also processes a high potential for α-helix formation. The α-helix has 3.6 residues per turn, which implies that residues in the positions *a* and *d* get very close to each other thus forming a hydrophobic boundary that runs almost parallel to the helix axis. This is shown in (**b**) in which single heptad repeats in α-helical conformation are shown in cross section; the three-stranded structure of matrilin-1 is depicted. The thickness of the lines connecting the residues indicates the position perpendicular to the paper plane. Thick lines are nearer to the viewer than thin lines. In all coiled-coil structures several α-helices are assembled and stabilized by interactions between the hydrophobic boundaries of residues in *a* and *d* positions. Electrostatic interaction between residues in *e* and *g* positions may also play a role. Depending on special features of the sequence two-stranded, three- and five-stranded coiled-coils are possible. In (**c**), the side views of the five-stranded coiled-coil domain of thrombospondin-5 (also called cartilage oligomeric matrix protein or COMP) and the three-stranded structure of matrilin-1 (cartilage matrix protein) are shown as ribbon diagrams. These structures are five to six heptads long but other coiled-coil domains can be shorter (e.g., in tenascins) or much longer (e.g., in laminins). Original data for matrilin-1, Dames et al. (1998) and for thrombospondin-5, Malashkevich et al. (1996)

stereoisomers. For most collagens, it is not known in which isomeric form they exist and whether two forms may coexist (Bächinger and Engel 2005).

Like the collagen helix, the α-helical coiled-coil structure is extended and its translation per residue is about 0.15 nm. It exists with different numbers of polypeptide chains. The most common coiled-coil structures in ECM contain three chains, but five-stranded structures are also known. For unknown reasons, the many cytosolic coiled-coil structures are only two-stranded. The coiled-coil structures are

stabilized by hydrophobic interactions between repeated hydrophobic side chains. The most common repeat is the heptad repeat, *abcdefg*, in which residues in positions *a* and *d* carry hydrophobic side chains (Parry et al. 2008). For α-helices, these residues are located at a surface boundary, which runs almost parallel to the helix axis. The hydrophobic boundaries combine and form a three- or five-stranded structure (Fig. 1.2). Some deviations are found from the regular heptad repeat, which also lead to suitable interaction boundaries. In contrast to the collagen triple helix, the chains in the α-helical coiled-coil structure are not staggered.

The collagen triple helix and the α-helical coiled-coil structure occur as domains in many ECM proteins. In contrast to the globular domains, their size varies over a broad range. Collagen triple helices may be only 14 nm long in minicollagens of jellyfish and 800 nm long in annelid cuticle collagens (Engel 1997). The coiled-coil structures are also found in different lengths. In matrilins and thrombospondins, only three to five heptad repeats form short rods of 3–5 nm, whereas the long arm of laminin is about 75 nm long. In laminin, short sequence regions of a different type (Engel 1992) frequently interrupt the repeats. Interruptions of the regular GXY-repeats are also found in many collagens. They may serve to introduce flexible kinks. For collagen IV, such flexible kinks were demonstrated by electron microscopy (Hofmann et al. 1984). Uninterrupted collagen triple helices and coiled-coil structures possess a high stiffness, which give them the appearance of rods with only gradual bending in electron micrographs. The diameter of the collagen triple helix is about 1.2 nm, whereas the coiled-coil structure is somewhat thicker (see Figs. 1.1 and 1.2).

Membrane spanning domains occur only in a limited number of ECM proteins. They have been predicted from N-terminal hydrophobic sequences in collagen XIII, XVII, XXIII, and XXV. These collagens are classified as cell surface receptors. With their collagenous ectodomains, they are involved in cell attachment. Similarly, membrane spanning domains of the α- and β-subunits of integrins link cells to ECM components (Hynes 2002). The integrins are major receptors for many ECM proteins and will be introduced below (Sect. 1.3.3). Their ectodomains have a multidomain organization similar to that of ECM proteins and they are intimately connected to the ECM.

1.2.3 Do Homologous Domains Have Related Functions?

It is the hope of computational biologists that the function of proteins can be predicted from their sequence. A frequently discussed concept is the assignment of hypothetical functions to multidomain proteins with the assumption that homologous domains will have similar function (Friedberg et al. 2006). As a general principle, this assumption does not hold and only for rather basic functions like membrane spanning and oligomerization may the concept be successful.

A large number of experimental investigations show that homologous domains adopted a variety of functions. They should be looked at as related folds in which

specific regions have been adopted for different functions. One of the many examples is the Kunitz inhibitor domain, which is found in the γ-chains of collagen VI. The active Kunitz inhibitor forms a strong complex with trypsin, α-chymotrypsin, or related proteases and its specificity depends on the structure of its active site. An extensive search for a related function of the Kunitz domain in collagen VI was not successful (Mayer et al. 1994). In addition, the three-dimensional structure of the collagen VI domain was solved, and it was realized that an inhibitor site is missing in the Kunitz domain of collagen VI and that this domain may be involved in a different function (Kohfeldt et al. 1996). This function is still ill defined, but the domain participates in the linkage of several collagen VI molecules during assembly of beaded filaments. The Ig domains in perlecan provide a second example. Only one of the Ig binds to nidogen/entactin, whereas the other Ig domains have no binding function and are classified as spacer elements (Kvansakul et al. 2001).

The issue is further complicated by the experimental observation that many domains exhibit their functions only when properly glycosylated or after another type of posttranslational modification. A frequently occurring modification is limited proteolytic cleavage by one of the many matrix proteinases. Again these changes do not depend on the type the domain alone but on its specific structure, tissue environment, and other factors.

Only a few and in part trivial functional predictions are possible on the basis of the amino acid sequence alone. It is possible to predict binding sites for bivalent cations in the EGFCa or EF-hand domains. It is also possible to predict the oligomerization potential of coiled-coil domains and collagen triple helices. More detailed predictions are possible with the help of the three-dimensional structure by which potential interaction sites can be explored. The most valid approach is still experimental investigation. Isolated domains of interest may be recombinantly expressed for functional studies. For ECM-domains, expression in mammalian cells is preferred over expression in *Escherichia coli* because of the need for proper disulfide linkage and glycosylation. Experiments teach us that single domains often show only incomplete functions. Elucidation of the full function requires a concerted action of many different domains, which are arranged in a machine-like, exactly defined spatial arrangement (Engel 2007).

1.2.4 Domain Organization

The polypeptide chains of ECM proteins often consist of a large number of individual domains. The linear representation of domains is called domain organization. The domain organization of the typical adhesive ECM glycoprotein fibronectin is shown in Fig. 1.3.

As mentioned earlier, the domain organization of any protein of known sequence may be obtained from databases like CDART, SMART, and SPAM. Clearly, only domains will be identified that are already entered in the database. Some of the sequence regions for which no domains are displayed may contain novel domains.

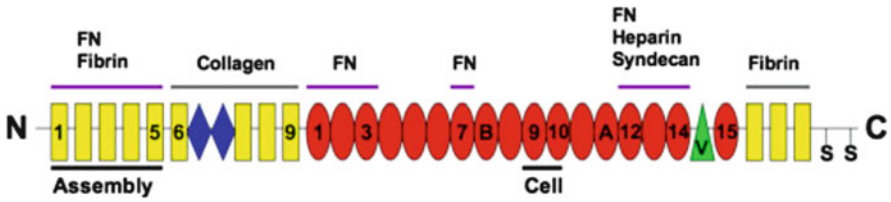

Fig. 1.3 Domain organization and functional regions of fibronectin. Internally homologous domains are shown in a linear arrangement for one of the two subunits of fibronectin. A second chain (not shown) is connected by disulfide bonds (S–S) at the C-terminus. Rectangles stand for F1-, chevrons for F2-, ellipsoids for F3-domains, and a triangle for a variable domain. Regions with binding potential for other matrix components are indicated on top of the structure and other functional regions below the structure. The cell-binding F3-domain 10 contains an exposed Arg-Gly-Asp (RGD)-site to which the fibronectin-specific cellular receptor integrin α5β1 binds. For more information see: Hynes (1985); Mao and Schwarzbauer (2005); Potts and Campbell (1996)

The curators of the databases ask researchers for suggestions for potentially new domains.

In single molecule electron micrographs, many ECM proteins show an extended shape (Fig. 1.4). The observed structures can often be matched with the domain organization (see Fig. 1.5). By electron microscopy in combination with structural studies of individual domains, a representation of the total structure of a complex protein may be obtained (Engel 1994).

It should be noted that under the experimental conditions employed for electron microscopy, noncovalent interdomain interactions might be disrupted. Interactions between domains within the same protein have been frequently demonstrated and interactions between distant domains are also possible. Such interactions may lead to large changes in the global conformation of the protein. An important example is fibronectin, which exists in a condensed and extended form (Markovic et al. 1983). Only the latter is able to polymerize into fibronectin fibrils (Mao and Schwarzbauer 2005).

The domain organization of many ECM proteins may also vary due to the existence of splice variants. The variants contain novel domains and frequently a repertoire of proteins with slightly different domain organization is found. The expression of the splice variants is highly regulated by splicing factors and is frequently tissue specific. The functions of splice variants often differ significantly from those of the parent protein. An example is agrin for which only a specific splice variant interacts with the acetylcholine receptor (Gesemann et al. 1995).

1.2.5 Multimerization of Several Polypeptide Chains

The complexity of ECM proteins is further increased by the fact that several identical or different polypeptide chains associate into large oligomers. Examples of such proteins are shown in Fig. 1.5.

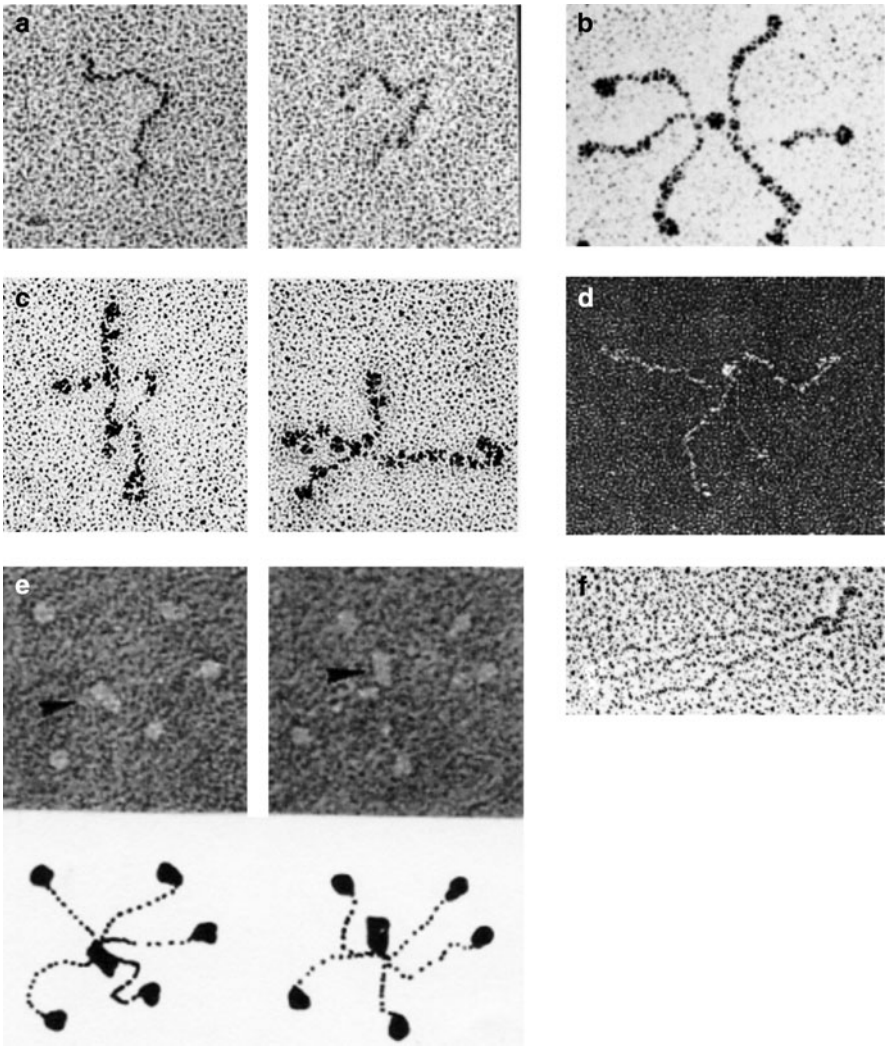

Fig. 1.4 Electron microscopic images of (**a**) fibronectin, (**b**) tenascin-C, (**c**) laminin-111, (**d**) collagen XII, (**e**) thrombospondin-5 (COMP), and (**f**) perlecan. Images were taken after rotary shadowing (**a**–**c**, **f**) or negative staining (**d**, **e**). For fibronectin (**a**), the two subunits of about 70 nm in length are connected with an average angle of 70°. Bending of the chains indicates a flexibility of the structural organization. Individual domains (see Fig. 1.3) cannot be distinguished at the resolution of the electron microscope. The six arms of tenascin-C (**b**) are about 75–90 nm long, depending on species and splice variant. Three arms are linked by a short coiled-coil domain, which is linked face to face to a second coiled-coil domain connecting the three other arms. This assembly structure appears as a dense cluster of platinum crystallites formed during rotary shadowing. Also the C-terminal FBG-domains can be recognized as globules, whereas other domains cannot be distinguished. Laminin-111 (**c**) has the shape of a Greek cross with three short arms of 35 nm and a long arm of 77 nm in length. Several of the domains in laminin (see Fig. 1.5) can be distinguished and the entire structure has a high degree of flexibility. Collagen XII

Association is mediated by oligomerization domains, of which the collagen triple helix and the α-helical coiled-coil domains are the most common. All collagen helix-containing proteins are trimers and all noncollagenous domains of these proteins also exist in three copies. By convention, these proteins are normally called collagens, although their content of triple helix may be small and other domains may predominate. For example, collagen XII (Fig. 1.5) consists mainly of VWFA and F3 domains and its triple helix is small (Koch et al. 1992). FACIT collagens contain several short triple helices. To facilitate proper chain association and to avoid slippage of chains, these and other collagens contain short coiled-coil regions (McAlinden et al. 2003) or other globular domains with a strong potential for trimerization (Boudko et al. 2009). Coiled-coil domains form trimers (see laminin in Fig. 1.5) and in some cases (thrombospondin 5, cartilage oligomeric matrix protein) pentamers. Tenascins are hexamers in which two three-stranded coiled-coil structures are arranged in an antiparallel way (Fig. 1.5). Several other globular domains also have a potential for oligomerization. An example is the POZ/BTB-domain in Mac-2 binding protein (Muller et al. 1999). This domain exists only as a dimer and is unstable in monomeric form. Domains with this property are called obligatory oligomerization domains. An obligatory trimer is the C-terminal NC1 domain of collagen IV (C4 in Table 2.1), which leads to a strong and specific antiparallel dimerization of this collagen (Khoshnoodi et al. 2008; Than et al. 2002). Fibronectin is normally found as a disulfide linked dimer and the disulfide-containing region at its C-terminus may also be classified as a dimerization domain.

1.2.6 Posttranslational Modifications

Domains of ECM proteins are extensively modified by a large number of chemical modifications. Some of these like the hydroxylation of proline to hydroxyproline occur in the interior of the cells; others like the proteolytic removal of the N-terminal and C-terminal domains of interstitial collagens occur at the cell surface. Many different limited proteolysis steps are mediated by matrix proteases

Fig. 1.4 (continued) (**d**) is a homotrimer. The domain organization of the subunits is shown in Fig. 1.5. In the electron micrograph, the noncollagenous domains appear as three flexible arms 90 nm in length. They are joined in the about 50 nm long collagen triple helix that is seen as a thin strand pointing to 5 pm in the picture. In thrombospondin-5 (COMP) (**e**), five about 40 nm long strands are connected by the pentameric coiled-coil domain (see Fig. 1.2). This domain appears as a rectangular body, which is marked by an *arrowhead* in the electron micrograph. Perlecan (**f**) is a low density proteoglycan whose three glycosaminoglycan (GAG)-chains are only faintly seen in the electron micrograph. The GAG chains are of variable length of up to 100 nm, and the core protein to which they are attached has a length of 30–50 nm. Original data for (**a**) and (**c**): Engel et al. (1981), (**b**): Spring et al. (1989), (**d**): Koch et al. (1992), (**e**): Morgelin et al. (1992), and (**f**): Paulsson et al. (1987)

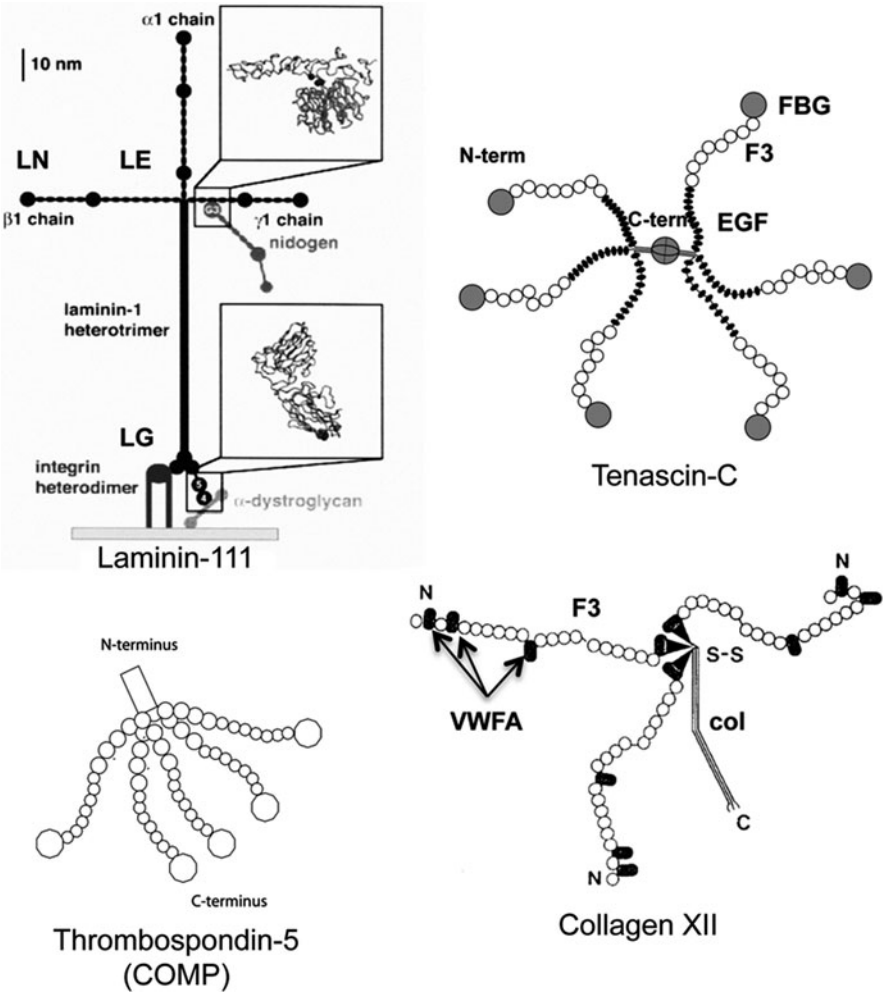

Fig. 1.5 Domain organization of multimeric ECM proteins laminin-111, tenascin-C, thrombospondin-5 (COMP), and collagen XII. Laminin-111 is a heterotrimer of an α1-, a β1-, and a γ1-chain, which are connected by a 77 nm long coiled-coil domain. At the C-terminus of the α-chain, 5 LG-domains are located, which are binding to integrin α6β1 and α-dystroglycan. The crystal structure of two LG domains was solved and is indicated in a *box*. The short arms are composed of many EGF-like (LE) domains and are terminated by LN domains involved in assembly. Two EGF-domains of the γ1-chain constitute the binding site for nidogen/entactin, which forms a very tight complex with laminin-111. The complex of the laminin-binding domain of nidogen with three EGF-domains was solved at atomic resolution and is shown in a box. Figure is taken from Sasaki et al. (2004).

Tenascin-C is a homo-hexamer whose 75 nm long subunits (for the avian 190 kDa splice variant) start with a FBG domain at the N-terminus (*gray circles*), followed by 8 F3 modules (*open circles*), and 11 EGF domains (*black ovals*). Three arms are joined by a short C-terminal three-stranded coiled-coil domain that is connected face to face to an identical trimeric assembly, giving rise to a six-armed ("hexabrachion") molecule with bilateral symmetry (Chiquet-Ehrismann et al. 1988; Spring et al. 1989).

in the extracellular space. All these events may lead to an activation of latent domains, large changes of function, and a degradation of ECM proteins. Proteolytic modifications are mediated by several hundred matrix proteases of the MMP, ADAM, ADAMTS, and other families (Tang and Hong 1999). Important for functional modifications and also for solubilization and structural stabilization are the N-glycosylation of asparagines and the O-glycosylation of serines and threonines. A special case is the attachment of long glycosaminoglycan (GAG) chains to the core protein of some ECM proteins. These proteins then exist as proteoglycans. Frequently two forms, one with and one without GAG chains are found. Furthermore, GAG chains are of variable length depending on age and tissue. All glycosylation occurs in the Golgi (exception: hyaluronan synthesis at the cell surface), and a large repertoire of enzymes is required.

It is not possible to deal with the large field of protein modifications in this chapter. However, it should be stressed that the building blocks of the ECM are dynamically remodeled by modifications and that large temporal and spatial variations exist. This leads to dramatic changes of functions and to a large increase of complexity. Examples are the degradation of collagens during pregnancy, the remodeling of bone, and the activation of thrombospondins. A static view on the matrix and its functions is therefore dangerous and functional prediction cannot be made without a detailed knowledge of the modification state.

1.2.7 Calcium Binding and Mineralization

Calcium and other divalent cations are essential for the regulation of many events inside cells. Here their concentration is in the nanomolar range and highly variable. Calcium-binding regulatory proteins such as calmodulin are changing their conformational state and function in response to the change in calcium concentration. In contrast, the calcium concentration is high in the ECM and rather uniform

Fig. 1.5 (continued) Thrombospondin-5, also called cartilage oligomeric matrix protein (COMP), has 5 about 35 nm long arms that are connected by a five-stranded coiled-coil domain (*box*, see Fig. 1.2). Each arm consists of 4 EGF domains (*medium size circles*) and 7 TSP3 domains (*small circles*) and is terminated by a TSPC domain (*large circle*). Crystallography showed that the TSPC forms a joint structure with the TSP3 domains (Kvansakul et al. 2004).
Collagen XII is composed of three identical subunits with very large noncollagenous "arms" about 70 nm in length, which are joined by a 50 nm long C-terminal collagen (col) triple helix with an interruption (visible as kink by EM; cf. Fig. 1.4). The N-terminal noncollagenous parts of the subunits consist of many F3 modules (*open circles*), interspersed with VWFA domains (*black ovals*), and a thrombospondin N-terminal like domain (TSPN; *black triangles*). The VWFA modules are recognizable in EM micrographs as globules protruding from the tandem F3 domains (cf. Fig. 1.4d). The scheme depicts the large variant of collagen XII (320 kDa per subunit), which is also a proteoglycan since it carries chondroitin sulfate chains at the N-terminus. Alternative splicing gives rise to a small variant (220 kDa) that lacks half of the noncollagenous domain and chondroitin sulfate (Koch et al. 1992; Koch et al. 1995)