The Extracellular Matrix

48

Robert K. Murray, MD, PhD, & Frederick W. Keeley, PhD

BIOMEDICAL IMPORTANCE

Most mammalian cells are located in tissues where they are surrounded by a complex **extracellular matrix (ECM)** often referred to as **"connective tissue."** The ECM contains three major classes of biomolecules: (1) the **structural proteins,** collagen, elastin, and fibrillin; (2) certain **specialized proteins** such as fibrillin, fibronectin, and laminin; and (3) **proteoglycans,** whose chemical natures are described below. The ECM has been found to be involved in many normal and pathologic processes—eg, it plays important roles in development, in inflammatory states, and in the spread of cancer cells. Involvement of certain components of the ECM has been documented in both rheumatoid arthritis and **osteoarthritis.** Several diseases (eg, osteogenesis imperfecta and a number of types of the Ehlers-Danlos syndrome) are due to genetic disturbances of the synthesis of collagen. Specific components of proteoglycans (the glycosaminoglycans; GAGs) are affected in the group of genetic disorders known as the mucopolysaccharidoses. Changes occur in the ECM during the aging process. This chapter describes the basic biochemistry of the three major classes of biomolecules found in the ECM and illustrates their biomedical significance. Major biochemical features of two specialized forms of ECM—bone and cartilage—and of a number of diseases involving them are also briefly considered.

COLLAGEN IS THE MOST ABUNDANT PROTEIN IN THE ANIMAL WORLD

Collagen, the major component of most connective tissues, constitutes approximately 25% of the protein of mammals. It provides an extracellular framework for all metazoan animals and exists in virtually every animal tissue. At least 19 distinct types of collagen made up of 30 distinct polypeptide chains (each encoded by a separate gene) have been identified in human tissues. Although several of these are present only in small proportions, they may play important roles in determining the physical properties of specific tissues. In addition, a number of proteins (eg, the C1q component of the complement system, pulmonary surfactant proteins SP-A and SP-D) that are not classified as collagens have collagen-like domains in their structures; these proteins are sometimes referred to as "noncollagen collagens."

Table 48–1 summarizes the types of collagens found in human tissues; the nomenclature used to designate types of collagen and their genes is described in the footnote.

The 19 types of collagen mentioned above can be subdivided into a number of classes based primarily on the structures they form (Table 48–2). In this chapter, we shall be primarily concerned with the fibril-forming collagens I and II, the major collagens of skin and bone and of cartilage, respectively. However, mention will be made of some of the other collagens.

COLLAGEN TYPE I IS COMPOSED OF A TRIPLE HELIX STRUCTURE & FORMS FIBRILS

All collagen types have a **triple helical structure.** In some collagens, the entire molecule is triple helical, whereas in others the triple helix may involve only a fraction of the structure. Mature collagen type I, containing approximately 1000 amino acids, belongs to the former type; in it, each polypeptide subunit or alpha chain is twisted into a left-handed helix of three residues per turn (Figure 48–1). Three of these alpha chains are then wound into a right-handed superhelix, forming a rod-like molecule 1.4 nm in diameter and about 300 nm long. A striking characteristic of collagen is the occurrence of **glycine** residues at every third position of the triple helical portion of the alpha chain. This is necessary because glycine is the only amino acid small enough to be accommodated in the limited space available down the central core of the triple helix. This repeating structure, represented as $(Gly\text{-}X\text{-}Y)_n$, is an absolute requirement for the formation of the triple helix. While X and Y can be any other amino acids, about 100 of the X positions are proline and about 100 of the Y positions are hydroxyproline. Proline and hydroxyproline confer rigidity on the collagen molecule. **Hydroxyproline** is formed by the posttranslational hydroxylation of peptide-bound proline residues catalyzed by the enzyme **prolyl hydroxylase,** whose cofactors are **ascorbic acid** (vitamin C) and α-ketoglutarate. Lysines

Table 48–1. Types of collagen and their genes.[1,2]

Type	Genes	Tissue
I	*COL1A1, COL1A2*	Most connective tissues, including bone
II	*COL2A1*	Cartilage, vitreous humor
III	*COL3A1*	Extensible connective tissues such as skin, lung, and the vascular system
IV	*COL4A1–COL4A6*	Basement membranes
V	*COL5A1–COL5A3*	Minor component in tissues containing collagen I
VI	*COL6A1–COL6A3*	Most connective tissues
VII	*COL7A1*	Anchoring fibrils
VIII	*COL8A1–COL8A2*	Endothelium, other tissues
IX	*COL9A1–COL9A3*	Tissues containing collagen II
X	*COL10A1*	Hypertrophic cartilage
XI	*COL11A1, COL11A2, COL2A1*	Tissues containing collagen II
XII	*COL12A1*	Tissues containing collagen I
XIII	*COL13A1*	Many tissues
XIV	*COL14A1*	Tissues containing collagen I
XV	*COL15A1*	Many tissues
XVI	*COL16A1*	Many tissues
XVII	*COL17A1*	Skin hemidesmosomes
XVIII	*COL18A1*	Many tissues (eg, liver, kidney)
XIX	*COL19A1*	Rhabdomyosarcoma cells

[1]Adapted slightly from Prockop DJ, Kivirrikko KI: Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem 1995;64:403.
[2]The types of collagen are designated by Roman numerals. Constituent procollagen chains, called proα chains, are numbered using Arabic numerals, followed by the collagen type in parentheses. For instance, type I procollagen is assembled from two proα1(I) and one proα2(I) chain. It is thus a heterotrimer, whereas type 2 procollagen is assembled from three proα1(II) chains and is thus a homotrimer. The collagen genes are named according to the collagen type, written in Arabic numerals for the gene symbol, followed by an A and the number of the proα chain that they encode. Thus, the COL1A1 and COL1A2 genes encode the α1 and α2 chains of type I collagen, respectively.

Table 48–2. Classification of collagens, based primarily on the structures that they form.[1]

Class	Type
Fibril-forming	I, II, III, V, and XI
Network-like	IV, VIII, X
FACITs[2]	IX, XII, XIV, XVI, XIX
Beaded filaments	VI
Anchoring fibrils	VII
Transmembrane domain	XIII, XVII
Others	XV, XVIII

[1]Based on Prockop DJ, Kivirrikko KI: Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem 1995;64:403.
[2]FACITs = fibril-associated collagens with interrupted triple helices.

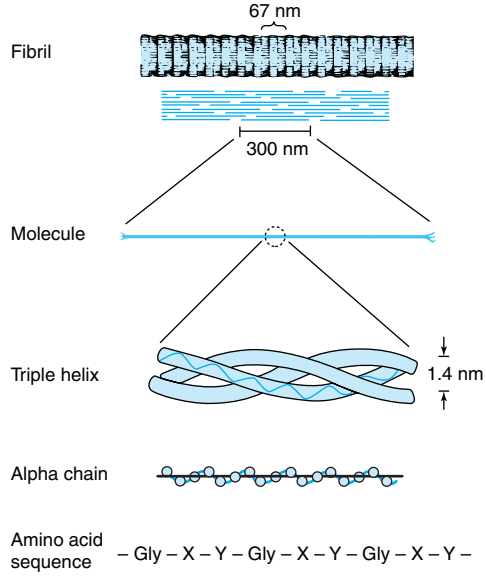

Figure 48–1. **Molecular features of collagen structure from primary sequence up to the fibril.** (Slightly modified and reproduced, with permission, from Eyre DR: Collagen: Molecular diversity in the body's protein scaffold. Science 1980;207:1315. Copyright © 1980 by the American Association for the Advancement of Science.)

in the Y position may also be posttranslationally modified to hydroxylysine through the action of **lysyl hydroxylase,** an enzyme with similar cofactors. Some of these hydroxylysines may be further modified by the addition of galactose or galactosyl-glucose through an **O-glycosidic linkage,** a glycosylation site that is unique to collagen.

Collagen types that form long rod-like fibers in tissues are assembled by lateral association of these triple helical units into a "quarter staggered" alignment such that each is displaced longitudinally from its neighbor by slightly less than one-quarter of its length (Figure 48–1, upper part). This arrangement is responsible for the banded appearance of these fibers in connective tissues. Collagen fibers are further stabilized by the formation of **covalent cross-links,** both within and between the triple helical units. These cross-links form through the action of **lysyl oxidase,** a copper-dependent enzyme that oxidatively deaminates the ε-amino groups of certain lysine and hydroxylysine residues, yielding reactive aldehydes. Such aldehydes can form aldol condensation products with other lysine- or hydroxylysine-derived aldehydes or form Schiff bases with the ε-amino groups of unoxidized lysines or hydroxylysines. These reactions, after further chemical rearrangements, result in the stable covalent cross-links that are important for the tensile strength of the fibers. Histidine may also be involved in certain cross-links.

Several collagen types do not form fibrils in tissues (Table 48–2). They are characterized by interruptions of the triple helix with stretches of protein lacking Gly-X-Y repeat sequences. These non-Gly-X-Y sequences result in areas of globular structure interspersed in the triple helical structure.

Type IV collagen, the best-characterized example of a collagen with discontinuous triple helices, is an important component of **basement membranes,** where it forms a mesh-like network.

Collagen Undergoes Extensive Posttranslational Modifications

Newly synthesized collagen undergoes extensive posttranslational modification before becoming part of a mature extracellular collagen fiber (Table 48–3). Like most secreted proteins, collagen is synthesized on ribosomes in a precursor form, **preprocollagen,** which contains a leader or signal sequence that directs the polypeptide chain into the lumen of the endoplasmic reticulum. As it enters the endoplasmic reticulum, this leader sequence is enzymatically removed. **Hydroxylation** of proline and lysine residues and **glycosylation** of hydroxylysines in the **procollagen** molecule also take place at this site. The procollagen molecule contains

Table 48–3. Order and location of processing of the fibrillar collagen precursor.

Intracellular
1. Cleavage of signal peptide
2. Hydroxylation of prolyl residues and some lysyl residues; glycosylation of some hydroxylysyl residues
3. Formation of intrachain and interchain S–S bonds in extension peptides
4. Formation of triple helix
Extracellular
1. Cleavage of amino and carboxyl terminal propeptides
2. Assembly of collagen fibers in quarter-staggered alignment
3. Oxidative deamination of ε-amino groups of lysyl and hydroxylysyl residues to aldehydes
4. Formation of intra- and interchain cross-links via Schiff bases and aldol condensation products

polypeptide extensions (extension peptides) of 20–35 kDa at both its amino and carboxyl terminal ends, neither of which is present in mature collagen. Both extension peptides contain cysteine residues. While the amino terminal propeptide forms only intrachain disulfide bonds, the carboxyl terminal propeptides form both intrachain and interchain disulfide bonds. Formation of these disulfide bonds assists in the registration of the three collagen molecules to form the triple helix, winding from the carboxyl terminal end. After formation of the triple helix, no further hydroxylation of proline or lysine or glycosylation of hydroxylysines can take place. **Self-assembly** is a cardinal principle in the biosynthesis of collagen.

Following **secretion** from the cell by way of the Golgi apparatus, extracellular enzymes called **procollagen aminoproteinase** and **procollagen carboxyproteinase** remove the extension peptides at the amino and carboxyl terminal ends, respectively. Cleavage of these propeptides may occur within crypts or folds in the cell membrane. Once the propeptides are removed, the triple helical collagen molecules, containing approximately 1000 amino acids per chain, spontaneously assemble into collagen fibers. These are further stabilized by the formation of inter- and intrachain cross-links through the action of lysyl oxidase, as described previously.

The same cells that secrete collagen also secrete **fibronectin,** a large glycoprotein present on cell surfaces, in the extracellular matrix, and in blood (see below). Fibronectin binds to aggregating precollagen fibers and alters the kinetics of fiber formation in the pericellular matrix. Associated with fibronectin and procollagen in

this matrix are the **proteoglycans** heparan sulfate and chondroitin sulfate (see below). In fact, type IX collagen, a minor collagen type from cartilage, contains attached proteoglycan chains. Such interactions may serve to regulate the formation of collagen fibers and to determine their orientation in tissues.

Once formed, collagen is relatively metabolically stable. However, its breakdown is increased during starvation and various inflammatory states. Excessive production of collagen occurs in a number of conditions, eg, hepatic cirrhosis.

A Number of Genetic Diseases Result From Abnormalities in the Synthesis of Collagen

About 30 genes encode collagen, and its pathway of biosynthesis is complex, involving at least eight enzyme-catalyzed posttranslational steps. Thus, it is not surprising that a number of diseases (Table 48–4) are due to **mutations in collagen genes** or in **genes encoding some of the enzymes** involved in these posttranslational modifications. The diseases affecting bone (eg, osteogenesis imperfecta) and cartilage (eg, the chondrodysplasias) will be discussed later in this chapter.

Ehlers-Danlos syndrome comprises a group of inherited disorders whose principal clinical features are hyperextensibility of the skin, abnormal tissue fragility, and increased joint mobility. The clinical picture is variable, reflecting underlying extensive genetic heterogeneity. At least 10 types have been recognized, most but not all of which reflect a variety of lesions in the synthesis of collagen. **Type IV** is the most serious because of its tendency for spontaneous rupture of arteries or the bowel, reflecting abnormalities in type III collagen. Patients with **type VI,** due to a deficiency of lysyl hydroxylase, exhibit marked joint hypermobility and a tendency to ocular rupture. A deficiency of procollagen N-proteinase, causing formation of abnormal thin, irregular collagen fibrils, results in **type VIIC,** manifested by marked joint hypermobility and soft skin.

Alport syndrome is the designation applied to a number of genetic disorders (both X-linked and autosomal) affecting the structure of type IV collagen fibers, the major collagen found in the basement membranes of the renal glomeruli (see discussion of laminin, below). Mutations in several genes encoding type IV collagen fibers have been demonstrated. The presenting sign is hematuria, and patients may eventually develop end-stage renal disease. Electron microscopy reveals characteristic abnormalities of the structure of the basement membrane and lamina densa.

In **epidermolysis bullosa,** the skin breaks and blisters as a result of minor trauma. The dystrophic form is due to mutations in *COL7A1,* affecting the structure of type VII collagen. This collagen forms delicate fibrils that anchor the basal lamina to collagen fibrils in the dermis. These anchoring fibrils have been shown to be markedly reduced in this form of the disease, probably resulting in the blistering. Epidermolysis bullosa simplex, another variant, is due to mutations in keratin 5 (Chapter 49).

Scurvy affects the structure of collagen. However, it is due to a deficiency of ascorbic acid (Chapter 45) and is not a genetic disease. Its major signs are bleeding

Table 48–4. Diseases caused by mutations in collagen genes or by deficiencies in the activities of posttranslational enzymes involved in the biosynthesis of collagen.[1]

Gene or Enzyme	Disease[2]
COL1A1, COL1A2	Osteogenesis imperfecta, type 1[3] (MIM 1566200) Osteoporosis[4] (MIM 166710) Ehlers-Danlos syndrome type VII autosomal dominant (130060)
COL2A1	Severe chondrodysplasias Osteoarthritis[4] (MIM 120140)
COL3A1	Ehlers-Danlos syndrome type IV (MIM 130050)
COL4A3–COL4A6	Alport syndrome (including both autosomal and X-linked forms) (MIM 104200)
COL7A1	Epidermolysis bullosa, dystrophic (MIM 131750)
COL10A1	Schmid metaphysial chondrodysplasia (MIM 156500)
Lysyl hydroxylase	Ehlers-Danlos syndrome type VI (MIM 225400)
Procollagen *N*-proteinase	Ehlers-Danlos syndrome type VII autosomal recessive (MIM 225410)
Lysyl hydroxylase	Menkes disease[5] (MIM 309400)

[1]Adapted from Prockop DJ, Kivirrikko KI: Collagens: molecular biology, diseases, and potentials for therapy. Annu Rev Biochem 1995;64:403.
[2]Genetic linkage to collagen genes has been shown for a few other conditions not listed here.
[3]At least four types of osteogenesis imperfecta are recognized; the great majority of mutations in all types are in the *COL1A1* and *COL1A2* genes.
[4]At present applies to only a relatively small number of such patients.
[5]Secondary to a deficiency of copper (Chapter 50).

gums, subcutaneous hemorrhages, and poor wound healing. These signs reflect impaired synthesis of collagen due to deficiencies of prolyl and lysyl hydroxylases, both of which require ascorbic acid as a cofactor.

ELASTIN CONFERS EXTENSIBILITY & RECOIL ON LUNG, BLOOD VESSELS, & LIGAMENTS

Elastin is a connective tissue protein that is responsible for properties of extensibility and elastic recoil in tissues. Although not as widespread as collagen, elastin is present in large amounts, particularly in tissues that require these physical properties, eg, lung, large arterial blood vessels, and some elastic ligaments. Smaller quantities of elastin are also found in skin, ear cartilage, and several other tissues. In contrast to collagen, there appears to be only one genetic type of elastin, although variants arise by alternative splicing (Chapter 37) of the hnRNA for elastin. Elastin is synthesized as a soluble monomer of 70 kDa called **tropoelastin.** Some of the prolines of tropoelastin are hydroxylated to **hydroxyproline** by prolyl hydroxylase, though hydroxylysine and glycosylated hydroxylysine are not present. Unlike collagen, tropoelastin is not synthesized in a pro- form with extension peptides. Furthermore, elastin does not contain repeat Gly-X-Y sequences, triple helical structure, or carbohydrate moieties.

After secretion from the cell, certain lysyl residues of tropoelastin are oxidatively deaminated to aldehydes by **lysyl oxidase,** the same enzyme involved in this process in collagen. However, the major cross-links formed in elastin are the **desmosines,** which result from the condensation of three of these lysine-derived aldehydes with an unmodified lysine to form a tetrafunctional cross-link unique to elastin. Once cross-linked in its mature, extracellular form, elastin is highly insoluble and extremely stable and has a very low turnover rate. Elastin exhibits a variety of random coil conformations that permit the protein to stretch and subsequently recoil during the performance of its physiologic functions.

Table 48–5 summarizes the main differences between collagen and elastin.

Deletions in the elastin gene (located at 7q11.23) have been found in approximately 90% of subjects with **Williams syndrome,** a developmental disorder affecting connective tissue and the central nervous system. The mutations, by affecting synthesis of elastin, probably play a causative role in the supravalvular aortic stenosis often found in this condition. A number of skin diseases (eg, scleroderma) are associated with accumulation of elastin. Fragmentation or, alternatively, a decrease of elastin is found in conditions such as pulmonary emphysema, cutis laxa, and aging of the skin.

Table 48–5. Major differences between collagen and elastin.

Collagen	Elastin
1. Many different genetic types	One genetic type
2. Triple helix	No triple helix; random coil conformations permitting stretching
3. $(Gly\text{-}X\text{-}Y)_n$ repeating structure	No $(Gly\text{-}X\text{-}Y)_n$ repeating structure
4. Presence of hydroxylysine	No hydroxylysine
5. Carbohydrate-containing	No carbohydrate
6. Intramolecular aldol cross-links	Intramolecular desmosine cross-links
7. Presence of extension peptides during biosynthesis	No extension peptides present during biosynthesis

MARFAN SYNDROME IS DUE TO MUTATIONS IN THE GENE FOR FIBRILLIN, A PROTEIN PRESENT IN MICROFIBRILS

Marfan syndrome is a relatively prevalent inherited disease affecting connective tissue; it is inherited as an autosomal dominant trait. It affects the eyes (eg, causing dislocation of the lens, known as ectopia lentis), the skeletal system (most patients are tall and exhibit long digits [arachnodactyly] and hyperextensibility of the joints), and the cardiovascular system (eg, causing weakness of the aortic media, leading to dilation of the ascending aorta). Abraham Lincoln may have had this condition. Most cases are caused by mutations in the gene (on chromosome 15) for fibrillin; missense mutations have been detected in several patients with Marfan syndrome.

Fibrillin is a large glycoprotein (about 350 kDa) that is a structural component of microfibrils, 10- to 12-nm fibers found in many tissues. Fibrillin is secreted (subsequent to a proteolytic cleavage) into the extracellular matrix by fibroblasts and becomes incorporated into the insoluble microfibrils, which appear to provide a scaffold for deposition of elastin. Of special relevance to Marfan syndrome, fibrillin is found in the zonular fibers of the lens, in the periosteum, and associated with elastin fibers in the aorta (and elsewhere); these locations respectively explain the ectopia lentis, arachnodactyly, and cardiovascular problems found in the syndrome. Other proteins (eg, emelin and two microfibril-associated proteins) are also present in these microfibrils, and it appears likely that abnormalities of them may cause other connective tissue disorders. An-

other gene for fibrillin exists on chromosome 5; mutations in this gene are linked to causation of congenital contractural arachnodactyly but not to Marfan syndrome. The probable sequence of events leading to Marfan syndrome is summarized in Figure 48–2.

FIBRONECTIN IS AN IMPORTANT GLYCOPROTEIN INVOLVED IN CELL ADHESION & MIGRATION

Fibronectin is a major glycoprotein of the extracellular matrix, also found in a soluble form in plasma. It consists of two identical subunits, each of about 230 kDa, joined by two disulfide bridges near their carboxyl terminals. The gene encoding fibronectin is very large, containing some 50 exons; the RNA produced by its transcription is subject to considerable alternative splicing, and as many as 20 different mRNAs have been detected in various tissues. Fibronectin contains three types of repeating motifs (I, II, and III), which are organized into functional **domains** (at least seven); functions of these domains include binding **heparin** (see below) and fibrin, collagen, DNA, and cell surfaces (Figure 48–3). The amino acid sequence of the fibronectin receptor of fibroblasts has been derived, and the protein is a member of the transmembrane integrin class of proteins (Chapter 51). The integrins are heterodimers, containing various types of α and β polypeptide chains. Fibronectin contains an Arg-Gly-Asp (RGD) sequence that binds to the receptor. The RGD sequence is shared by a number of other proteins present in the ECM that bind to integrins present in cell surfaces. Synthetic peptides containing the RGD sequence inhibit the binding of fibronectin to cell surfaces. Figure 48–4 illustrates the interaction of collagen, fibronectin, and laminin, all major proteins of the ECM, with a typical cell (eg, fibroblast) present in the matrix.

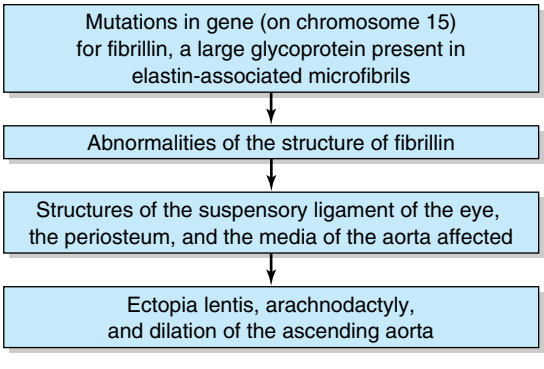

Figure 48–2. Probable sequence of events in the causation of the major signs exhibited by patients with Marfan syndrome (MIM 154700).

The fibronectin receptor interacts indirectly with **actin** microfilaments (Chapter 49) present in the cytosol (Figure 48–5). A number of proteins, collectively known as **attachment proteins,** are involved; these include talin, vinculin, an actin-filament capping protein, and α-actinin. Talin interacts with the receptor and vinculin, whereas the latter two interact with actin. The interaction of fibronectin with its receptor provides one route whereby the exterior of the cell can communicate with the interior and thus affect cell behavior. Via the interaction with its cell receptor, fibronectin plays an important role in the adhesion of cells to the ECM. It is also involved in cell migration by providing a binding site for cells and thus helping them to steer their way through the ECM. The amount of fibronectin around many transformed cells is sharply reduced, partly explaining their faulty interaction with the ECM.

LAMININ IS A MAJOR PROTEIN COMPONENT OF RENAL GLOMERULAR & OTHER BASAL LAMINAS

Basal laminas are specialized areas of the ECM that surround epithelial and some other cells (eg, muscle cells); here we discuss only the laminas found in the **renal glomerulus.** In that structure, the basal lamina is contributed by two separate sheets of cells (one endothelial and one epithelial), each disposed on opposite sides of the lamina; these three layers make up the **glomerular membrane.** The primary components of the basal lamina are three proteins—laminin, entactin, and type IV collagen—and the GAG **heparin** or **heparan sulfate.** These components are synthesized by the underlying cells.

Laminin (about 850 kDa, 70 nm long) consists of three distinct elongated polypeptide chains (A, B_1, and B_2) linked together to form an elongated cruciform shape. It has binding sites for type IV collagen, heparin, and integrins on cell surfaces. The collagen interacts with laminin (rather than directly with the cell surface), which in turn interacts with integrins or other laminin receptor proteins, thus anchoring the lamina to the cells. **Entactin,** also known as "nidogen," is a glycoprotein containing an RGD sequence; it binds to laminin and is a major cell attachment factor. The relatively thick basal lamina of the renal glomerulus has an important role in **glomerular filtration,** regulating the passage of large molecules (most plasma proteins) across the glomerulus into the renal tubule. The glomerular membrane allows small molecules, such as **inulin** (5.2 kDa), to pass through as easily as water. On the other hand, only a small amount of the protein **albumin** (69

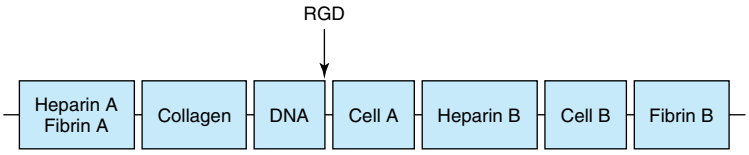

Figure 48–3. Schematic representation of fibronectin. Seven functional domains of fibronectin are represented; two different types of domain for heparin, cell-binding, and fibrin are shown. The domains are composed of various combinations of three structural motifs (I, II, and III), not depicted in the figure. Also not shown is the fact that fibronectin is a dimer joined by disulfide bridges near the carboxyl terminals of the monomers. The approximate location of the RGD sequence of fibronectin, which interacts with a variety of fibronectin integrin receptors on cell surfaces, is indicated by the arrow. (Redrawn after Yamada KM: Adhesive recognition sequences. J Biol Chem 1991;266:12809.)

kDa), the major plasma protein, passes through the normal glomerulus. This is explained by two sets of facts: (1) The pores in the glomerular membrane are large enough to allow molecules up to about 8 nm to pass through. (2) Albumin is smaller than this pore size, but it is prevented from passing through easily by the negative charges of heparan sulfate and of certain sialic acid-containing glycoproteins present in the lamina. These negative charges repel albumin and most plasma proteins, which are negatively charged at the pH of blood. The normal structure of the glomerulus may be severely damaged in certain types of **glomerulonephritis** (eg, caused by antibodies directed against various components of the glomerular membrane). This alters the pores and the amounts and dispositions of the negatively charged macromolecules referred to above, and relatively massive amounts of albumin (and of certain

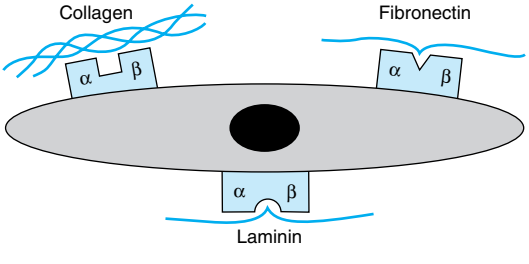

Figure 48–4. Schematic representation of a cell interacting through various integrin receptors with collagen, fibronectin, and laminin present in the ECM. (Specific subunits are not indicated.) (Redrawn after Yamada KM: Adhesive recognition sequences. J Biol Chem 1991;266:12809.)

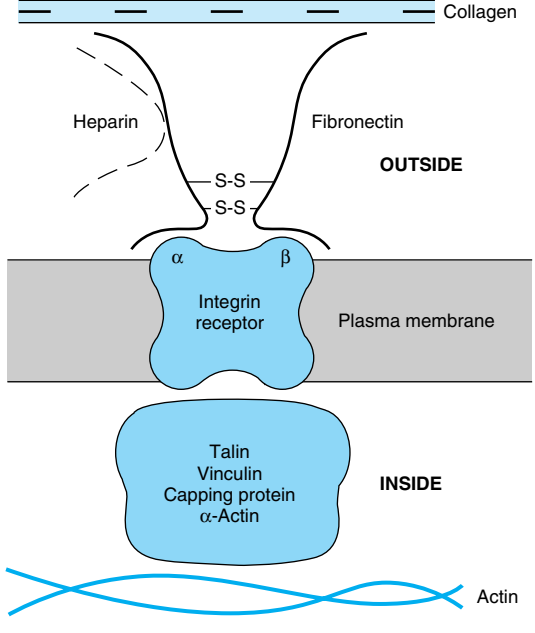

Figure 48–5. Schematic representation of fibronectin interacting with an integrin fibronectin receptor situated in the exterior of the plasma membrane of a cell of the ECM and of various attachment proteins interacting indirectly or directly with an actin microfilament in the cytosol. For simplicity, the attachment proteins are represented as a complex.

other plasma proteins) can pass through into the urine, resulting in severe **albuminuria.**

PROTEOGLYCANS & GLYCOSAMINOGLYCANS

The Glycosaminoglycans Found in Proteoglycans Are Built Up of Repeating Disaccharides

Proteoglycans are proteins that contain covalently linked glycosaminoglycans. A number of them have been characterized and given names such as syndecan, betaglycan, serglycin, perlecan, aggrecan, versican, decorin, biglycan, and fibromodulin. They vary in tissue distribution, nature of the core protein, attached glycosaminoglycans, and function. The proteins bound covalently to glycosaminoglycans are called **"core proteins";** they have proved difficult to isolate and characterize, but the use of recombinant DNA technology is beginning to yield important information about their structures. The amount of carbohydrate in a proteoglycan is usually much greater than is found in a glycoprotein and may comprise up to 95% of its weight. Figures 48–6 and 48–7 show the general structure of one particular proteoglycan, **aggrecan,** the major type found in cartilage. It is very large (about 2×10^3 kDa), with its overall structure resembling that of a bottle brush. It contains a long strand of hyaluronic acid (one type of GAG) to which link proteins are attached noncovalently. In turn, these latter interact noncovalently with core protein molecules from which chains of other GAGs (keratan sulfate and chondroitin sulfate in this case) project. More details on this macromolecule are given when cartilage is discussed below.

There are at least seven **glycosaminoglycans (GAGs):** hyaluronic acid, chondroitin sulfate, keratan sulfates I and II, heparin, heparan sulfate, and dermatan sulfate. A GAG is an unbranched polysaccharide made up of repeating disaccharides, one component of which is always an amino sugar (hence the name GAG), either D-glucosamine or D-galactosamine. The other component of the repeating disaccharide (except in the case of keratan sulfate) is a uronic acid, either L-glucuronic acid (GlcUA) or its 5′-epimer, L-iduronic acid (IdUA). With the exception of hyaluronic acid, all the GAGs contain sulfate groups, either as O-esters or as N-sulfate (in heparin and heparan sulfate). Hyaluronic acid affords another exception because there is no clear evidence that it is attached covalently to protein, as the definition of a proteoglycan given above specifies. Both GAGs and proteoglycans have proved difficult to work with, partly because of their complexity. However, they are major components of the ground substance; they have a number of important biologic roles; and they are involved in a number of disease processes—so that interest in them is increasing rapidly.

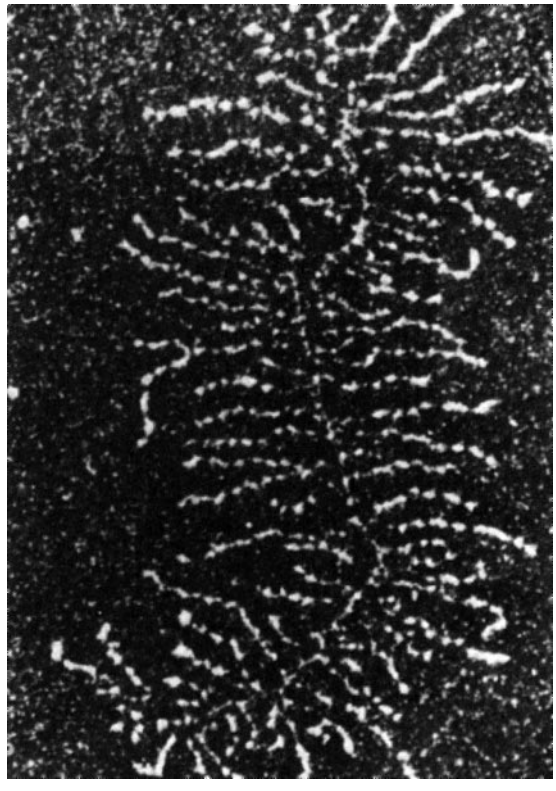

Figure 48–6. Dark field electron micrograph of a proteoglycan aggregate in which the proteoglycan subunits and filamentous backbone are particularly well extended. (Reproduced, with permission, from Rosenberg L, Hellman W, Kleinschmidt AK: Electron microscopic studies of proteoglycan aggregates from bovine articular cartilage. J Biol Chem 1975;250:1877.)

Biosynthesis of Glycosaminoglycans Involves Attachment to Core Proteins, Chain Elongation, & Chain Termination

A. Attachment to Core Proteins

The linkage between GAGs and their core proteins is generally one of three types.

1. An O-glycosidic bond between xylose (Xyl) and Ser, a bond that is unique to proteoglycans. This linkage is formed by transfer of a Xyl residue to Ser from UDP-xylose. Two residues of Gal are then added to the

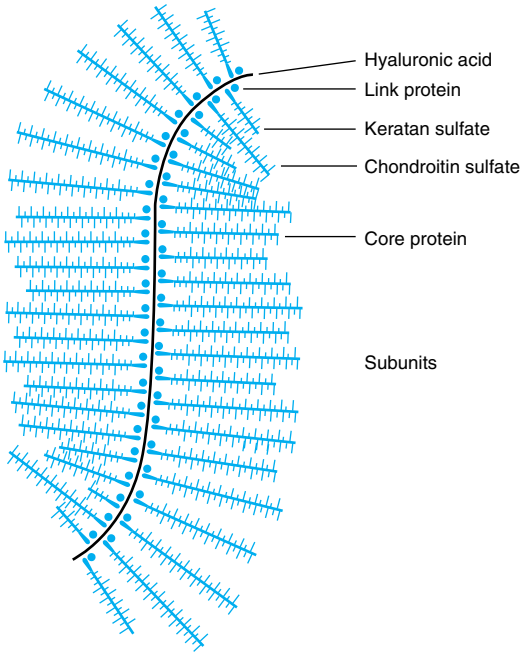

Figure 48–7. Schematic representation of the proteoglycan aggrecan. (Reproduced, with permission, from Lennarz WJ: *The Biochemistry of Glycoproteins and Proteoglycans.* Plenum Press, 1980.)

Xyl residue, forming a **link trisaccharide,** Gal-Gal-Xyl-Ser. Further chain growth of the GAG occurs on the terminal Gal.

2. An O-glycosidic bond forms between GalNAc (*N*-acetylgalactosamine) and Ser (Thr) (Figure 47–1[a]), present in keratan sulfate II. This bond is formed by donation to Ser (or Thr) of a GalNAc residue, employing UDP-GalNAc as its donor.

3. An N-glycosylamine bond between GlcNAc (*N*-acetylglucosamine) and the amide nitrogen of **Asn,** as found in N-linked glycoproteins (Figure 47–1[b]). Its synthesis is believed to involve dolichol-P-P-oligosaccharide.

The synthesis of the core proteins occurs in the **endoplasmic reticulum,** and formation of at least some of the above linkages also occurs there. Most of the later steps in the biosynthesis of GAG chains and their subsequent modifications occur in the **Golgi apparatus.**

B. Chain Elongation

Appropriate nucleotide sugars and highly specific Golgi-located glycosyltransferases are employed to synthesize the oligosaccharide chains of GAGs. The "one enzyme, one linkage" relationship appears to hold here, as in the case of certain types of linkages found in glycoproteins. The enzyme systems involved in chain elongation are capable of high-fidelity reproduction of complex GAGs.

C. Chain Termination

This appears to result from (1) sulfation, particularly at certain positions of the sugars, and (2) the progression of the growing GAG chain away from the membrane site where catalysis occurs.

D. Further Modifications

After formation of the GAG chain, numerous chemical modifications occur, such as the introduction of sulfate groups onto GalNAc and other moieties and the epimerization of GlcUA to IdUA residues. The enzymes catalyzing sulfation are designated **sulfotransferases** and use 3′-phosphoadenosine-5′-phosphosulfate (PAPS; active sulfate) as the sulfate donor. These Golgi-located enzymes are highly specific, and distinct enzymes catalyze sulfation at different positions (eg, carbons 2, 3, 4, and 6) on the acceptor sugars. An **epimerase** catalyzes conversions of glucuronyl to iduronyl residues.

The Various Glycosaminoglycans Exhibit Differences in Structure & Have Characteristic Distributions

The seven GAGs named above differ from each other in a number of the following properties: amino sugar composition, uronic acid composition, linkages between these components, chain length of the disaccharides, the presence or absence of sulfate groups and their positions of attachment to the constituent sugars, the nature of the core proteins to which they are attached, the nature of the linkage to core protein, their tissue and subcellular distribution, and their biologic functions.

The structures (Figure 48–8) and the distributions of each of the GAGs will now be briefly discussed. The major features of the seven GAGs are summarized in Table 48–6.

A. Hyaluronic Acid

Hyaluronic acid consists of an unbranched chain of repeating disaccharide units containing GlcUA and GlcNAc. Hyaluronic acid is present in bacteria and is widely distributed among various animals and tissues, including synovial fluid, the vitreous body of the eye, cartilage, and loose connective tissues.

B. Chondroitin Sulfates (Chondroitin 4-Sulfate & Chondroitin 6-Sulfate)

Proteoglycans linked to chondroitin sulfate by the Xyl-Ser O-glycosidic bond are prominent components of **cartilage** (see below). The repeating disaccharide is similar to that found in hyaluronic acid, containing

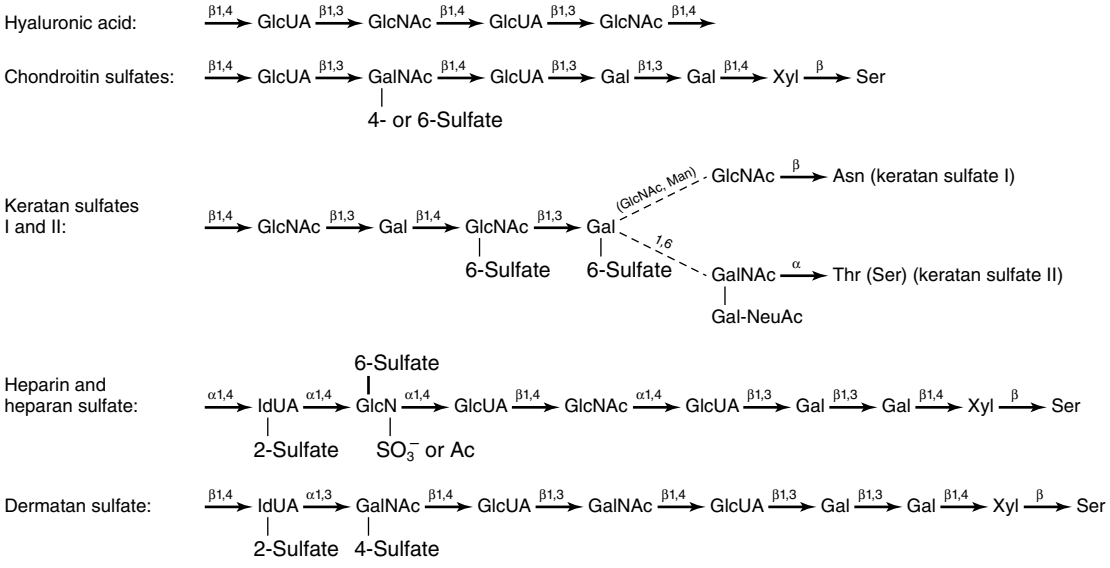

Figure 48–8. Summary of structures of glycosaminoglycans and their attachments to core proteins. (GlcUA, D-glucuronic acid; IdUA, L-iduronic acid; GlcN, D-glucosamine; GalN, D-galactosamine; Ac, acetyl; Gal, D-galactose; Xyl, D-xylose; Ser, L-serine; Thr, L-threonine; Asn, L-asparagine; Man, D-mannose; NeuAc, *N*-acetylneuraminic acid.) The summary structures are qualitative representations only and do not reflect, for example, the uronic acid composition of hybrid glycosaminoglycans such as heparin and dermatan sulfate, which contain both L-iduronic and D-glucuronic acid. Neither should it be assumed that the indicated substituents are always present, eg, whereas most iduronic acid residues in heparin carry a 2′-sulfate group, a much smaller proportion of these residues are sulfated in dermatan sulfate. The presence of link trisaccharides (Gal-Gal-Xyl) in the chondroitin sulfates, heparin, and heparan and dermatan sulfates is shown. (Slightly modified and reproduced, with permission, from Lennarz WJ: *The Biochemistry of Glycoproteins and Proteoglycans.* Plenum Press, 1980.)

Table 48–6. Major properties of the glycosaminoglycans.

GAG	Sugars	Sulfate[1]	Linkage of Protein	Location
HA	GlcNAc, GlcUA	Nil	No firm evidence	Synovial fluid, vitreous humor, loose connective tissue
CS	GalNAc, GlcUA	GalNAc	Xyl-Ser; associated with HA via link proteins	Cartilage, bone, cornea
KS I	GlcNAc, Gal	GlcNAc	GlcNAc-Asn	Cornea
KS II	GlcNAc, Gal	Same as KS I	GalNAc-Thr	Loose connective tissue
Heparin	GlcN, IdUA	GlcN GlcN IdUA	Ser	Mast cells
Heparan sulfate	GlcN, GlcUA	GlcN	Xyl-Ser	Skin fibroblasts, aortic wall
Dermatan sulfate	GalNAc, IdUA, (GlcUA)	GalNAc IdUa	Xyl-Ser	Wide distribution

[1]The sulfate is attached to various positions of the sugars indicated (see Figure 48–7).

GlcUA but with GalNAc replacing GlcNAc. The GalNAc is substituted with sulfate at either its 4′ or its 6′ position, with approximately one sulfate being present per disaccharide unit.

C. Keratan Sulfates I & II

As shown in Figure 48–8, the keratan sulfates consist of repeating Gal-GlcNAc disaccharide units containing sulfate attached to the 6′ position of GlcNAc or occasionally of Gal. Type I is abundant in **cornea,** and type II is found along with chondroitin sulfate attached to hyaluronic acid in **loose connective tissue.** Types I and II have different attachments to protein (Figure 48–8).

D. Heparin

The repeating disaccharide contains **glucosamine** (GlcN) and either of the two uronic acids (Figure 48–9). Most of the amino groups of the GlcN residues are **N-sulfated,** but a few are acetylated. The GlcN also carries a C_6 sulfate ester.

Approximately 90% of the uronic acid residues are IdUA. Initially, all of the uronic acids are GlcUA, but a 5′-epimerase converts approximately 90% of the GlcUA residues to IdUA after the polysaccharide chain is formed. The protein molecule of the heparin proteoglycan is unique, consisting exclusively of serine and glycine residues. Approximately two-thirds of the serine residues contain GAG chains, usually of 5–15 kDa but occasionally much larger. Heparin is found in the granules of **mast cells** and also in liver, lung, and skin.

E. Heparan Sulfate

This molecule is present on many **cell surfaces** as a proteoglycan and is extracellular. It contains GlcN with fewer N-sulfates than heparin, and, unlike heparin, its predominant uronic acid is GlcUA.

F. Dermatan Sulfate

This substance is widely distributed in animal tissues. Its structure is similar to that of chondroitin sulfate, except that in place of a GlcUA in β-1,3 linkage to GalNAc it contains an IdUA in an α-1,3 linkage to GalNAc. Formation of the IdUA occurs, as in heparin and heparan sulfate, by 5′-epimerization of GlcUA. Because this is regulated by the degree of sulfation and because sulfation is incomplete, dermatan sulfate contains both IdUA-GalNAc and GlcUA-GalNAc disaccharides.

Deficiencies of Enzymes That Degrade Glycosaminoglycans Result in Mucopolysaccharidoses

Both exo- and endoglycosidases degrade GAGs. Like most other biomolecules, GAGs are subject to turnover, being both synthesized and degraded. In adult tissues, GAGs generally exhibit relatively slow turnover, their half-lives being days to weeks.

Understanding of the degradative pathways for GAGs, as in the case of glycoproteins (Chapter 47) and glycosphingolipids (Chapter 24), has been greatly aided by elucidation of the specific enzyme deficiencies that occur in certain **inborn errors of metabolism.** When GAGs are involved, these inborn errors are called **mucopolysaccharidoses** (Table 48–7).

Degradation of GAGs is carried out by a battery of lysosomal hydrolases. These include certain endoglycosidases, various exoglycosidases, and sulfatases, generally acting in sequence to degrade the various GAGs. A number of them are indicated in Table 48–7.

The **mucopolysaccharidoses** share a common mechanism of causation, as illustrated in Figure 48–10. They are inherited in an autosomal recessive manner, with Hurler and Hunter syndromes being perhaps the most widely studied. None are common. In some cases, a family history of a mucopolysaccharidosis is obtained. Specific laboratory investigations of help in their diagnosis are urine testing for the presence of increased

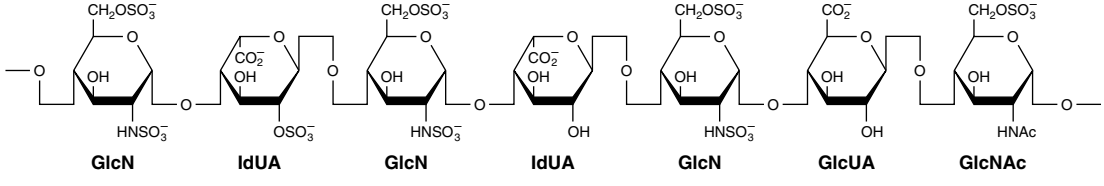

Figure 48–9. Structure of heparin. The polymer section illustrates structural features typical of heparin; however, the sequence of variously substituted repeating disaccharide units has been arbitrarily selected. In addition, non-O-sulfated or 3-O-sulfated glucosamine residues may also occur. (Modified, redrawn, and reproduced, with permission, from Lindahl U et al: Structure and biosynthesis of heparin-like polysaccharides. Fed Proc 1977;36:19.)

Table 48–7. Biochemical defects and diagnostic tests in mucopolysaccharidoses (MPS) and mucolipidoses (ML).[1]

Name	Alternative Designation[2,3]	Enzymatic Defect	Urinary Metabolites
Mucopolysaccharidoses			
Hurler, Scheie, Hurler-Scheie (MIM 252800)	MPS I	α-L-Iduronidase	Dermatan sulfate, heparan sulfate
Hunter (MIM 309900)	MPS II	Iduronate sulfatase	Dermatan sulfate, heparan sulfate
Sanfilippo A (MIM 252900)	MPS IIIA	Heparan sulfate *N*-sulfatase (sulfamidase)	Heparan sulfate
Sanfilippo B (MIM 252920)	MPS IIIB	α-*N*-Acetylglucosaminidase	Heparan sulfate
Sanfilippo C (MIM 252930)	MPS IIIC	Acetyltransferase	Heparan sulfate
Sanfilippo D (MIM 252940)	MPS IIID	*N*-Acetylglucosamine 6-sulfatase	Heparan sulfate
Morquio A (MIM 253000)	MPS IVA	Galactosamine 6-sulfatase	Keratan sulfate, chondroitin 6-sulfate
Morquio B (MIM 253010)	MPS IVB	β-Galactosidase	Keratan sulfate
Maroteaux-Lamy (MIM 253200)	MPS VI	*N*-Acetylgalactosamine 4-sulfatase (arylsulfatase B)	Dermatan sulfate
Sly (MIM 253220)	MPS VII	β-Glucuronidase	Dermatan sulfate, heparan sulfate, chondroitin 4-sulfate, chondroitin 6-sulfate
Mucolipidoses			
Sialidosis (MIM 256550)	M LI	Sialidase (neuraminidase)	Glycoprotein fragments
I-cell disease (MIM 252500)	ML II	UDP-*N*-acetylglucosamine: glycoprotein *N*-acetylglucosamininylphosphotransferase. (Acid hydrolases thus lack phosphomannosyl residues.)	Glycoprotein fragments
Pseudo-Hurler polydystrophy (MIM 252600)	ML III	As for ML II but deficiency is incomplete	Glycoprotein fragments

[1]Modified and reproduced, with permission, from DiNatale P, Neufeld EF: The biochemical diagnosis of mucopolysaccharidoses, mucolipidoses and related disorders. In: *Perspectives in Inherited Metabolic Diseases,* vol 2. Barr B et al (editors). Editiones Ermes (Milan), 1979.

[2]Fibroblasts, leukocytes, tissues, amniotic fluid cells, or serum can be used for the assay of many of the above enzymes. Patients with these disorders exhibit a variety of clinical findings that may include cloudy corneas, mental retardation, stiff joints, cardiac abnormalities, hepatosplenomegaly, and short stature, depending on the specific disease and its severity.

[3]The term MPS V is no longer used. The existence of MPS VIII (suspected glucosamine 6-sulfatase deficiency: MIM 253230) has not been confirmed. At least one case of hyaluronidase deficiency (MPS IX; MIM 601492) has been reported.

amounts of GAGs and assays of suspected enzymes in white cells, fibroblasts, or sometimes in serum. In certain cases, a tissue biopsy is performed and the GAG that has accumulated can be determined by electrophoresis. DNA tests are increasingly available. Prenatal diagnosis can be made using amniotic cells or chorionic villus biopsy.

The term **"mucolipidosis"** was introduced to denote diseases that combined features common to both mucopolysaccharidoses and sphingolipidoses (Chapter 24). Three mucolipidoses are listed in Table 48–7. In **sialidosis** (mucolipidosis I, ML-I), various oligosaccharides derived from glycoproteins and certain gangliosides can accumulate in tissues. **I-cell disease** (ML-II)

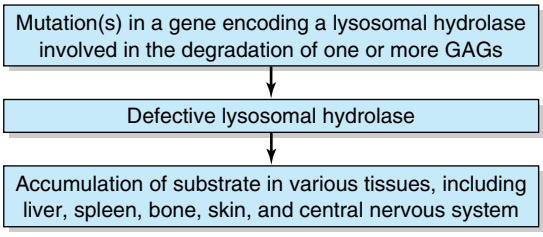

Figure 48–10. Simplified scheme of causation of a mucopolysaccharidosis, such as Hurler syndrome (MIM 252800), in which the affected enzyme is α-L-iduronidase. Marked accumulation of the GAGs in the tissues mentioned in the figure could cause hepatomegaly, splenomegaly, disturbances of growth, coarse facies, and mental retardation, respectively.

and **pseudo-Hurler polydystrophy** (ML-III) are described in Chapter 47. The term "mucolipidosis" is retained because it is still in relatively widespread clinical usage, but it is not appropriate for these two latter diseases since the mechanism of their causation involves mislocation of certain lysosomal enzymes. Genetic defects of the catabolism of the oligosaccharide chains of glycoproteins (eg, mannosidosis, fucosidosis) are also described in Chapter 47. Most of these defects are characterized by increased excretion of various fragments of glycoproteins in the urine, which accumulate because of the metabolic block, as in the case of the mucolipidoses.

Hyaluronidase is one important enzyme involved in the catabolism of both hyaluronic acid and chondroitin sulfate. It is a widely distributed endoglycosidase that cleaves hexosaminidic linkages. From hyaluronic acid, the enzyme will generate a tetrasaccharide with the structure $(GlcUA\text{-}\beta\text{-}1,3\text{-}GlcNAc\text{-}\beta\text{-}1,4)_2$, which can be degraded further by a β-glucuronidase and β-*N*-acetylhexosaminidase. Surprisingly, only one case of an apparent genetic deficiency of this enzyme appears to have been reported.

Proteoglycans Have Numerous Functions

As indicated above, proteoglycans are remarkably complex molecules and are found in every tissue of the body, mainly in the ECM or "ground substance." There they are associated with each other and also with the other major structural components of the matrix, collagen and elastin, in quite specific manners. Some proteoglycans bind to collagen and others to elastin. These interactions are important in determining the structural organization of the matrix. Some proteoglycans (eg, decorin) can also bind growth factors such as TGF-β, modulating their effects on cells. In addition, some of them interact with certain adhesive proteins such as fibronectin and laminin (see above), also found in the matrix. The GAGs present in the proteoglycans are polyanions and hence bind polycations and cations such as Na^+ and K^+. This latter ability attracts water by osmotic pressure into the extracellular matrix and contributes to its turgor. GAGs also gel at relatively low concentrations. Because of the long extended nature of the polysaccharide chains of GAGs and their ability to gel, the proteoglycans can act as sieves, restricting the passage of large macromolecules into the ECM but allowing relatively free diffusion of small molecules. Again, because of their extended structures and the huge macromolecular aggregates they often form, they occupy a large volume of the matrix relative to proteins.

A. Some Functions of Specific GAGs & Proteoglycans

Hyaluronic acid is especially high in concentration in embryonic tissues and is thought to play an important role in permitting cell migration during morphogenesis and wound repair. Its ability to attract water into the extracellular matrix and thereby "loosen it up" may be important in this regard. The high concentrations of hyaluronic acid and chondroitin sulfates present in cartilage contribute to its compressibility (see below).

Chondroitin sulfates are located at sites of calcification in endochondral bone and are also found in cartilage. They are also located inside certain neurons and may provide an endoskeletal structure, helping to maintain their shape.

Both **keratan sulfate I** and **dermatan sulfate** are present in the cornea. They lie between collagen fibrils and play a critical role in corneal transparency. Changes in proteoglycan composition found in corneal scars disappear when the cornea heals. The presence of dermatan sulfate in the sclera may also play a role in maintaining the overall shape of the eye. Keratan sulfate I is also present in cartilage.

Heparin is an important anticoagulant. It binds with factors IX and XI, but its most important interaction is with **plasma antithrombin III** (discussed in Chapter 51). Heparin can also bind specifically to lipoprotein lipase present in capillary walls, causing a release of this enzyme into the circulation.

Certain proteoglycans (eg, **heparan sulfate**) are associated with the plasma membrane of cells, with their core proteins actually spanning that membrane. In it they may act as receptors and may also participate in the mediation of cell growth and cell-cell communication. The attachment of cells to their substratum in culture is mediated at least in part by heparan sulfate. This proteoglycan is also found in the basement membrane of the kidney along with type IV collagen and laminin

(see above), where it plays a major role in determining the charge selectiveness of glomerular filtration.

Proteoglycans are also found in intracellular locations such as the nucleus; their function in this organelle has not been elucidated. They are present in some storage or secretory granules, such as the chromaffin granules of the adrenal medulla. It has been postulated that they play a role in release of the contents of such granules. The various functions of GAGs are summarized in Table 48–8.

B. Associations With Major Diseases & With Aging

Hyaluronic acid may be important in permitting **tumor cells** to migrate through the ECM. Tumor cells can induce fibroblasts to synthesize greatly increased amounts of this GAG, thereby perhaps facilitating their own spread. Some tumor cells have less heparan sulfate at their surfaces, and this may play a role in the lack of adhesiveness that these cells display.

The intima of the **arterial wall** contains hyaluronic acid and chondroitin sulfate, dermatan sulfate, and heparan sulfate proteoglycans. Of these proteoglycans, dermatan sulfate binds plasma low-density lipoproteins. In addition, dermatan sulfate appears to be the major GAG synthesized by arterial smooth muscle cells. Because it is these cells that proliferate in atherosclerotic lesions in arteries, dermatan sulfate may play an important role in development of the atherosclerotic plaque.

Table 48–8. Some functions of glycosaminoglycans and proteoglycans.[1]

- Act as structural components of the ECM
- Have specific interactions with collagen, elastin, fibronectin, laminin, and other proteins such as growth factors
- As polyanions, bind polycations and cations
- Contribute to the characteristic turgor of various tissues
- Act as sieves in the ECM
- Facilitate cell migration (HA)
- Have role in compressibility of cartilage in weight-bearing (HA, CS)
- Play role in corneal transparency (KS I and DS)
- Have structural role in sclera (DS)
- Act as anticoagulant (heparin)
- Are components of plasma membranes, where they may act as receptors and participate in cell adhesion and cell-cell interactions (eg, HS)
- Determine charge-selectiveness of renal glomerulus (HS)
- Are components of synaptic and other vesicles (eg, HS)

[1]ECM, extracellular matrix; HA, hyaluronic acid; CS, chondroitin sulfate; KS I, keratan sulfate I; DS, dermatan sulfate; HS, heparan sulfate.

In various types of **arthritis,** proteoglycans may act as autoantigens, thus contributing to the pathologic features of these conditions. The amount of chondroitin sulfate in cartilage diminishes with age, whereas the amounts of keratan sulfate and hyaluronic acid increase. These changes may contribute to the development of **osteoarthritis.** Changes in the amounts of cer-

Table 48–9. The principal proteins found in bone.[1]

Proteins	Comments
Collagens	
Collagen type I	Approximately 90% of total bone protein. Composed of two α1(I) and one α2(I) chains.
Collagen type V	Minor component.
Noncollagen proteins	
Plasma proteins	Mixture of various plasma proteins.
Proteoglycans[2]	
CS-PG I (biglycan)	Contains two GAG chains; found in other tissues.
CS-PG II (decorin)	Contains one GAG chain; found in other tissues.
CS-PG III	Bone-specific.
Bone SPARC[3] protein (osteonectin)	Not bone-specific.
Osteocalcin (bone Gla protein)	Contains γ-carboxyglutamate residues that bind to hydroxyapatite. Bone-specific.
Osteopontin	Not bone-specific. Glycosylated and phosphorylated.
Bone sialoprotein	Bone-specific. Heavily glycosylated, and sulfated on tyrosine.
Bone morphogenetic proteins (BMPs)	A family (eight or more) of secreted proteins with a variety of actions on bone; many induce ectopic bone growth.

[1]Various functions have been ascribed to the noncollagen proteins, including roles in mineralization; however, most of them are still speculative. It is considered unlikely that the noncollagen proteins that are not bone-specific play a key role in mineralization. A number of other proteins are also present in bone, including a tyrosine-rich acidic matrix protein (TRAMP), some growth factors (eg, TGFβ), and enzymes involved in collagen synthesis (eg, lysyl oxidase).
[2]CS-PG, chondroitin sulfate–proteoglycan; these are similar to the dermatan sulfate PGs (DS-PGs) of cartilage (Table 48–11).
[3]SPARC, secreted protein acidic and rich in cysteine.

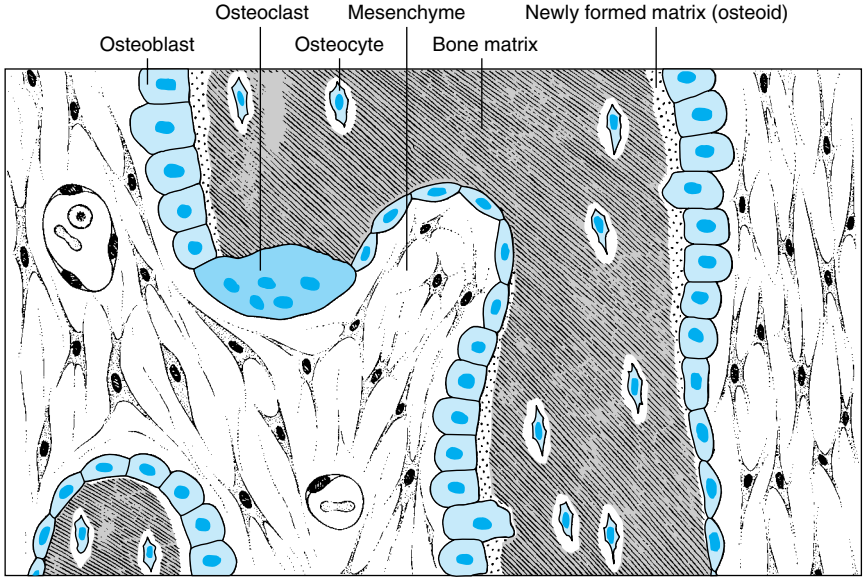

Figure 48–11. Schematic illustration of the major cells present in membranous bone. Osteoblasts (lighter color) are synthesizing type I collagen, which forms a matrix that traps cells. As this occurs, osteoblasts gradually differentiate to become osteocytes. (Reproduced, with permission, from Junqueira LC, Carneiro J: *Basic Histology*: *Text & Atlas,* 10th ed. McGraw-Hill, 2003.)

tain GAGs in the skin are also observed with **aging** and help to account for the characteristic changes noted in this organ in the elderly.

An exciting new phase in proteoglycan research is opening up with the findings that mutations that affect individual proteoglycans or the enzymes needed for their synthesis alter the regulation of specific signaling pathways in drosophila and *Caenorhabditis elegans,* thus affecting development; it already seems likely that similar effects exist in mice and humans.

BONE IS A MINERALIZED CONNECTIVE TISSUE

Bone contains both organic and inorganic material. The organic matter is mainly protein. The principal proteins of bone are listed in Table 48–9; **type I collagen** is the major protein, comprising 90–95% of the organic material. Type V collagen is also present in small amounts, as are a number of noncollagen proteins, some of which are relatively specific to bone. The inorganic or mineral component is mainly crystalline **hydroxyapatite**—$Ca_{10}(PO_4)_6(OH)_2$—along with sodium, magnesium, carbonate, and fluoride; approximately 99% of the body's calcium is contained in bone (Chapter 45). Hydroxyapatite confers on bone the strength and resilience required by its physiologic roles.

Bone is a dynamic structure that undergoes continuing cycles of remodeling, consisting of resorption followed by deposition of new bone tissue. This remodeling permits bone to adapt to both physical (eg, increases in weight-bearing) and hormonal signals.

The major cell types involved in bone resorption and deposition are **osteoclasts** and **osteoblasts** (Figure 48–11). The former are associated with resorption and the latter with deposition of bone. Osteocytes are descended from osteoblasts; they also appear to be involved in maintenance of bone matrix but will not be discussed further here.

Osteoclasts are multinucleated cells derived from pluripotent hematopoietic stem cells. Osteoclasts possess an apical membrane domain, exhibiting a ruffled border that plays a key role in bone resorption (Figure 48–12). A proton-translocating ATPase expels protons across the ruffled border into the resorption area, which is the microenvironment of low pH shown in the figure. This lowers the local pH to 4.0 or less, thus increasing the solubility of hydroxyapatite and allowing demineralization to occur. Lysosomal acid proteases are released that digest the now accessible matrix proteins.

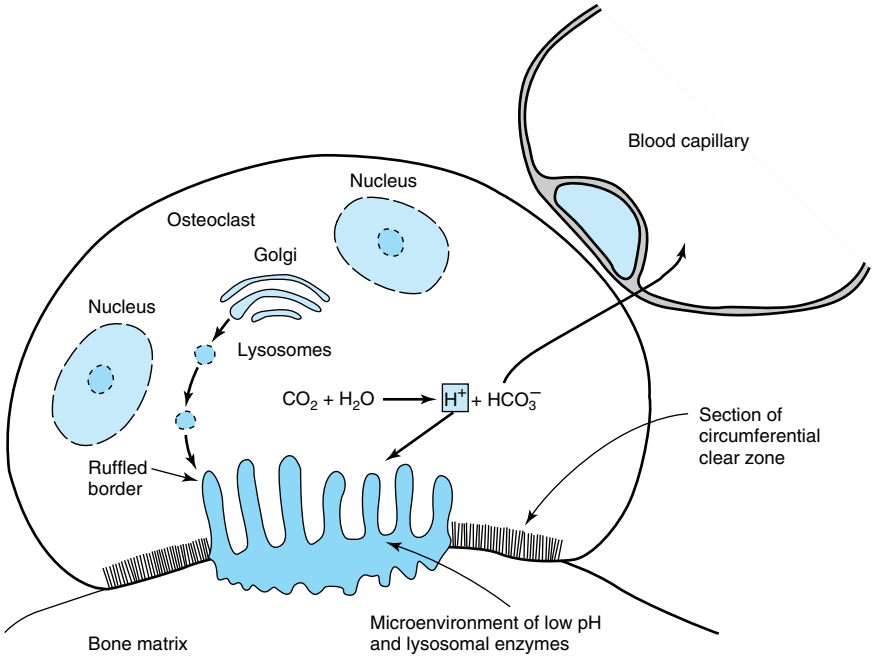

Figure 48–12. Schematic illustration of some aspects of the role of the osteoclast in bone resorption. Lysosomal enzymes and hydrogen ions are released into the confined microenvironment created by the attachment between bone matrix and the peripheral clear zone of the osteoclast. The acidification of this confined space facilitates the dissolution of calcium phosphate from bone and is the optimal pH for the activity of lysosomal hydrolases. Bone matrix is thus removed, and the products of bone resorption are taken up into the cytoplasm of the osteoclast, probably digested further, and transferred into capillaries. The chemical equation shown in the figure refers to the action of carbonic anhydrase II, described in the text. (Reproduced, with permission, from Junqueira LC, Carneiro J: *Basic Histology: Text & Atlas,* 10th ed. McGraw-Hill, 2003.)

Osteoblasts—mononuclear cells derived from pluripotent mesenchymal precursors—synthesize most of the proteins found in bone (Table 48–9) as well as various growth factors and cytokines. They are responsible for the deposition of new bone matrix (osteoid) and its subsequent mineralization. Osteoblasts control mineralization by regulating the passage of calcium and phosphate ions across their surface membranes. The latter contain alkaline phosphatase, which is used to generate phosphate ions from organic phosphates. The mechanisms involved in mineralization are not fully understood, but several factors have been implicated. Alkaline phosphatase contributes to mineralization but in itself is not sufficient. Small vesicles (matrix vesicles) containing calcium and phosphate have been described at sites of mineralization, but their role is not clear. Type I collagen appears to be necessary, with mineralization being first evident in the gaps between successive molecules. Recent interest has focused on acidic phosphoproteins, such as bone sialoprotein, acting as sites of nucleation. These proteins contain motifs (eg, poly-Asp and poly-Glu stretches) that bind calcium and may provide an initial scaffold for mineralization. Some macromolecules, such as certain proteoglycans and glycoproteins, can also act as inhibitors of nucleation.

It is estimated that approximately 4% of compact bone is renewed annually in the typical healthy adult, whereas approximately 20% of trabecular bone is replaced.

Many factors are involved in the regulation of bone metabolism, only a few of which will be mentioned here. Some stimulate osteoblasts (eg, parathyroid hormone and 1,25-dihydroxycholecalciferol) and others inhibit them (eg, corticosteroids). Parathyroid hormone and 1,25-dihydroxycholecalciferol also stimulate osteoclasts, whereas calcitonin and estrogens inhibit them.

Table 48–10. Some metabolic and genetic diseases affecting bone and cartilage.

Disease	Comments
Dwarfism	Often due to a deficiency of growth hormone, but has many other causes.
Rickets	Due to a deficiency of vitamin D during childhood.
Osteomalacia	Due to a deficiency of vitamin D during adulthood.
Hyperparathyroidism	Excess parathormone causes bone resorption.
Osteogenesis imperfecta (eg, MIM 166200)	Due to a variety of mutations in the *COL1A1* and *COL1A2* genes affecting the synthesis and structure of type I collagen.
Osteoporosis	Commonly postmenopausal or in other cases is more gradual and related to age; a small number of cases are due to mutations in the *COL1A1* and *COL1A2* genes and possibly in the vitamin D receptor gene (MIM 166710)
Osteoarthritis	A small number of cases are due to mutations in the *COL1A* genes.
Several chondrodysplasias	Due to mutations in *COL2A1* genes.
Pfeiffer syndrome[1] (MIM 100600)	Mutations in the gene encoding fibroblast growth receptor 1 (FGFR1).
Jackson-Weiss (MIM 123150) and Crouzon (MIM 123500) syndromes[1]	Mutations in the gene encoding FGFR2.
Achondroplasia (MIM 100800) and thanatophoric dysplasia[2] (MIM 187600)	Mutations in the gene encoding FGFR3.

[1]The Pfeiffer, Jackson-Weiss, and Crouzon syndromes are craniosynostosis syndromes; craniosynostosis is a term signifying premature fusion of sutures in the skull.
[2]Thanatophoric (Gk *thanatos* "death" + *phoros* "bearing") dysplasia is the most common neonatal lethal skeletal dysplasia, displaying features similar to those of homozygous achondroplasia.

MANY METABOLIC & GENETIC DISORDERS INVOLVE BONE

A number of the more important examples of metabolic and genetic disorders that affect bone are listed in Table 48–10.

Osteogenesis imperfecta (brittle bones) is characterized by abnormal fragility of bones. The scleras are often abnormally thin and translucent and may appear blue owing to a deficiency of connective tissue. Four types of this condition (mild, extensive, severe, and variable) have been recognized, of which the extensive type occurring in the newborn is the most ominous. Affected infants may be born with multiple fractures and not survive. Over 90% of patients with osteogenesis imperfecta have mutations in the *COL1A1* and *COL1A2* genes, encoding proα1(I) and proα2(I) chains, respectively. Over 100 mutations in these two genes have been documented and include partial gene deletions and duplications. Other mutations affect RNA splicing, and the most frequent type results in the replacement of glycine by another bulkier amino acid, affecting formation of the triple helix. In general, these mutations result in decreased expression of collagen or

Table 48–11. The principal proteins found in cartilage.

Proteins	Comments
Collagen proteins	
Collagen type II	90–98% of total articular cartilage collagen. Composed of three α1(II) chains.
Collagens V, VI, IX, X, XI	Type IX cross-links to type II collagen. Type XI may help control diameter of type II fibrils.
Noncollagen proteins	
Proteoglycans	
Aggrecan	The major proteoglycan of cartilage.
Large non-aggregating proteoglycan	Found in some types of cartilage.
DS-PG I (biglycan)[1]	Similar to CS-PG I of bone.
DS-PG II (decorin)	Similar to CS-PG II of bone.
Chondronectin	May play role in binding type II collagen to surface of cartilage.
Anchorin C II	May bind type II collagen to surface of chondrocyte.

[1]The core proteins of DS-PG I and DS-PG II are homologous to those of CS-PG I and CS-PG II found in bone (Table 48–9). A possible explanation is that osteoblasts lack the epimerase required to convert glucuronic acid to iduronic acid, the latter of which is found in dermatan sulfate.

in structurally abnormal proα chains that assemble into abnormal fibrils, weakening the overall structure of bone. When one abnormal chain is present, it may interact with two normal chains, but folding may be prevented, resulting in enzymatic degradation of all of the chains. This is called "procollagen suicide" and is an example of a dominant negative mutation, a result often seen when a protein consists of multiple different subunits.

Osteopetrosis (marble bone disease), characterized by increased bone density, is due to inability to resorb bone. One form occurs along with renal tubular acidosis and cerebral calcification. It is due to mutations in the gene (located on chromosome 8q22) encoding carbonic anhydrase II (CA II), one of four isozymes of carbonic anhydrase present in human tissues. The reaction catalyzed by carbonic anhydrase is shown below:

Reaction II is spontaneous. In osteoclasts involved in bone resorption, CA II apparently provides protons to neutralize the OH^- ions left inside the cell when H^+ ions are pumped across their ruffled borders (see above). Thus, if CA II is deficient in activity in osteoclasts, normal bone resorption does not occur, and osteopetrosis results. The mechanism of the cerebral calcification is not clear, whereas the renal tubular acidosis reflects deficient activity of CA II in the renal tubules.

Osteoporosis is a generalized progressive reduction in bone tissue mass per unit volume causing skeletal weakness. The ratio of mineral to organic elements is unchanged in the remaining normal bone. Fractures of various bones, such as the head of the femur, occur very easily and represent a huge burden to both the affected patients and to the health care budget of society. Among other factors, estrogens and interleukins-1 and -6 appear to be intimately involved in the causation of osteoporosis.

THE MAJOR COMPONENTS OF CARTILAGE ARE TYPE II COLLAGEN & CERTAIN PROTEOGLYCANS

The principal proteins of hyaline cartilage (the major type of cartilage) are listed in Table 48–11. Type II collagen is the principal protein (Figure 48–13), and a number of other minor types of collagen are also present. In

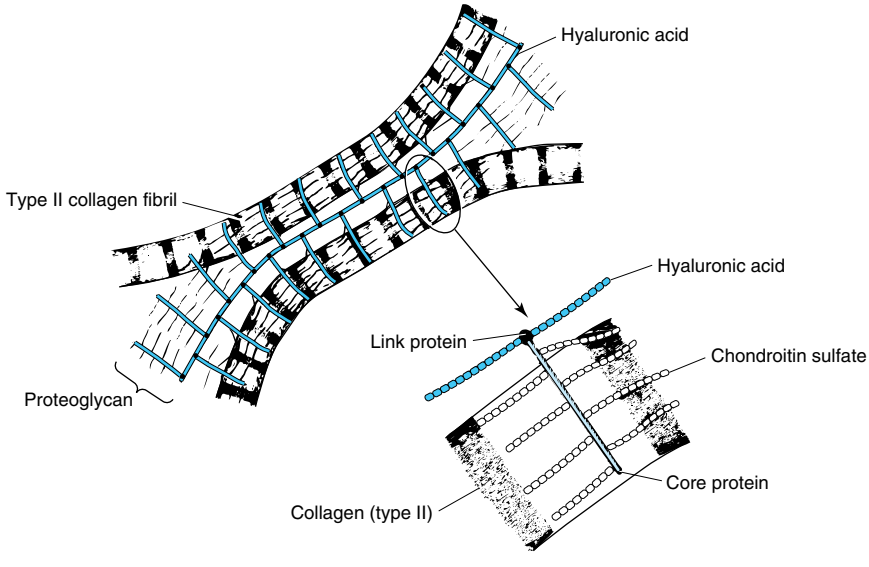

Figure 48–13. Schematic representation of the molecular organization in cartilage matrix. Link proteins noncovalently bind the core protein (lighter color) of proteoglycans to the linear hyaluronic acid molecules (darker color). The chondroitin sulfate side chains of the proteoglycan electrostatically bind to the collagen fibrils, forming a cross-linked matrix. The oval outlines the area enlarged in the lower part of the figure. (Reproduced, with permission, from Junqueira LC, Carneiro J: ***Basic Histology***: *Text & Atlas,* 10th ed. McGraw-Hill, 2003.)

addition to these components, elastic cartilage contains elastin and fibroelastic cartilage contains type I collagen. Cartilage contains a number of proteoglycans, which play an important role in its compressibility. **Aggrecan** (about 2×10^3 kDa) is the major proteoglycan. As shown in Figure 48–14, it has a very complex structure, containing several GAGs (hyaluronic acid, chondroitin sulfate, and keratan sulfate) and both link and core proteins. The core protein contains three domains: A, B, and C. The hyaluronic acid binds noncovalently to domain A of the core protein as well as to the link protein, which stabilizes the hyaluronate–core protein interactions. The keratan sulfate chains are located in domain B, whereas the chondroitin sulfate chains are located in domain C; both of these types of GAGs are bound covalently to the core protein. The core protein also contains both O- and N-linked oligosaccharide chains.

The other proteoglycans found in cartilage have simpler structures than aggrecan.

Chondronectin is involved in the attachment of type II collagen to chondrocytes.

Cartilage is an avascular tissue and obtains most of its nutrients from synovial fluid. It exhibits slow but continuous turnover. Various **proteases** (eg, collagenases and stromalysin) synthesized by chondrocytes can degrade collagen and the other proteins found in cartilage. Interleukin-1 (IL-1) and tumor necrosis factor α (TNFα) appear to stimulate the production of such proteases, whereas transforming growth factor β (TGFβ) and insulin-like growth factor 1 (IGF-I) generally exert an anabolic influence on cartilage.

THE MOLECULAR BASES OF THE CHONDRODYSPLASIAS INCLUDE MUTATIONS IN GENES ENCODING TYPE II COLLAGEN & FIBROBLAST GROWTH FACTOR RECEPTORS

Chondrodysplasias are a mixed group of hereditary disorders affecting cartilage. They are manifested by short-limbed dwarfism and numerous skeletal deformities. A number of them are due to a variety of mutations in the *COL2A1* gene, leading to abnormal forms of type II collagen. One example is **Stickler syndrome,** manifested by degeneration of joint cartilage and of the vitreous body of the eye.

The best-known of the chondrodysplasias is **achondroplasia,** the commonest cause of short-limbed dwarfism. Affected individuals have short limbs, nor-

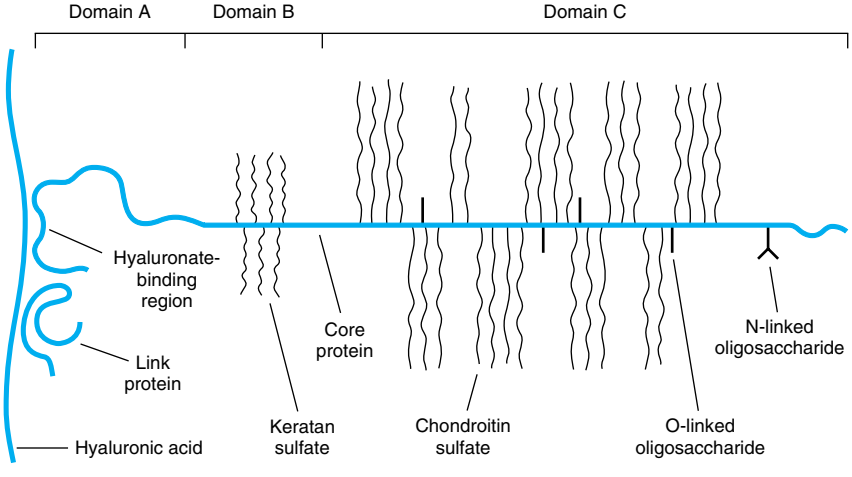

Figure 48–14. Schematic diagram of the aggrecan from bovine nasal cartilage. A strand of hyaluronic acid is shown on the left. The core protein (about 210 kDa) has three major domains. Domain A, at its amino terminal end, interacts with approximately five repeating disaccharides in hyaluronate. The link protein interacts with both hyaluronate and domain A, stabilizing their interactions. Approximately 30 keratan sulfate chains are attached, via GalNAc-Ser linkages, to domain B. Domain C contains about 100 chondroitin sulfate chains attached via Gal-Gal-Xyl-Ser linkages and about 40 O-linked oligosaccharide chains. One or more N-linked glycan chains are also found near the carboxyl terminal of the core protein. (Reproduced, with permission, from Moran LA et al: *Biochemistry,* 2nd ed. Neil Patterson Publishers, 1994.)

mal trunk size, macrocephaly, and a variety of other skeletal abnormalities. The condition is often inherited as an autosomal dominant trait, but many cases are due to new mutations. The molecular basis of achondroplasia is outlined in Figure 48–15. Achondroplasia is not a collagen disorder but is due to mutations in the gene encoding **fibroblast growth factor receptor 3 (FGFR3). Fibroblast growth factors** are a family of at least nine proteins that affect the growth and differentiation of cells of mesenchymal and neuroectodermal origin. Their receptors are transmembrane proteins and form a subgroup of the family of receptor tyrosine kinases. FGFR3 is one member of this subgroup and mediates the actions of FGF3 on cartilage. In almost all cases of achondroplasia that have been investigated, the mutations were found to involve nucleotide 1138 and resulted in substitution of arginine for glycine (residue number 380) in the transmembrane domain of the protein, rendering it inactive. No such mutation was found in unaffected individuals. As indicated in Table 48–10, other skeletal dysplasias (including certain craniosynostosis syndromes) are also due to mutations in genes encoding FGF receptors. Another type of skeletal dysplasia (diastrophic dysplasia) has been found to be due to mutation in a sulfate transporter. Thus, thanks to recombinant DNA technology, a new era in understanding of skeletal dysplasias has begun.

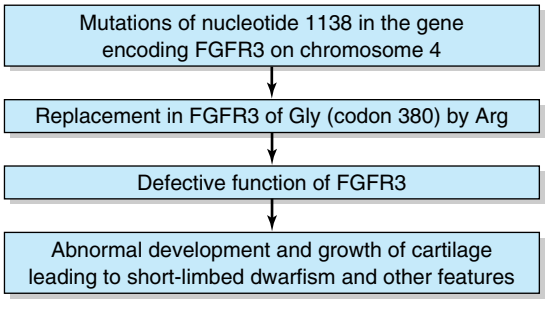

Figure 48–15. Simplified scheme of the causation of achondroplasia (MIM 100800). In most cases studied so far, the mutation has been a G to A transition at nucleotide 1138. In a few cases, the mutation was a G to C transversion at the same nucleotide. This particular nucleotide is a real "hot spot" for mutation. Both mutations result in replacement of a Gly residue by an Arg residue in the transmembrane segment of the receptor. A few cases involving replacement of Gly by Cys at codon 375 have also been reported.

SUMMARY

- The major components of the ECM are the structural proteins collagen, elastin, and fibrillin; a number of specialized proteins (eg, fibronectin and laminin); and various proteoglycans.
- Collagen is the most abundant protein in the animal kingdom; approximately 19 types have been isolated. All collagens contain greater or lesser stretches of triple helix and the repeating structure $(Gly\text{-}X\text{-}Y)_n$.
- The biosynthesis of collagen is complex, featuring many posttranslational events, including hydroxylation of proline and lysine.
- Diseases associated with impaired synthesis of collagen include scurvy, osteogenesis imperfecta, Ehlers-Danlos syndrome (many types), and Menkes disease.
- Elastin confers extensibility and elastic recoil on tissues. Elastin lacks hydroxylysine, Gly-X-Y sequences, triple helical structure, and sugars but contains desmosine and isodesmosine cross-links not found in collagen.
- Fibrillin is located in microfibrils. Mutations in the gene for fibrillin cause Marfan syndrome.
- The glycosaminoglycans (GAGs) are made up of repeating disaccharides containing a uronic acid (glucuronic or iduronic) or hexose (galactose) and a hexosamine (galactosamine or glucosamine). Sulfate is also frequently present.
- The major GAGs are hyaluronic acid, chondroitin 4- and 6-sulfates, keratan sulfates I and II, heparin, heparan sulfate, and dermatan sulfate.
- The GAGs are synthesized by the sequential actions of a battery of specific enzymes (glycosyltransferases, epimerases, sulfotransferases, etc) and are degraded by the sequential action of lysosomal hydrolases. Genetic deficiencies of the latter result in mucopolysaccharidoses (eg, Hurler syndrome).
- GAGs occur in tissues bound to various proteins (linker proteins and core proteins), constituting proteoglycans. These structures are often of very high molecular weight and serve many functions in tissues.
- Many components of the ECM bind to proteins of the cell surface named integrins; this constitutes one pathway by which the exteriors of cells can communicate with their interiors.
- Bone and cartilage are specialized forms of the ECM. Collagen I and hydroxyapatite are the major constituents of bone. Collagen II and certain proteoglycans are major constituents of cartilage.

- The molecular causes of a number of heritable diseases of bone (eg, osteogenesis imperfecta) and of cartilage (eg, the chondrodystrophies) are being revealed by the application of recombinant DNA technology.

REFERENCES

Bandtlow CE, Zimmermann DR: Proteoglycans in the developing brain: new conceptual insights for old proteins. Physiol Rev 2000;80:1267.

Bikle DD: Biochemical markers in the assessment of bone diseases. Am J Med 1997;103:427.

Burke D et al: Fibroblast growth factor receptors: lessons from the genes. Trends Biochem Sci 1998;23:59.

Compston JE: Sex steroids and bone. Physiol Rev 2001;81:419.

Fuller GM, Shields D: *Molecular Basis of Medical Cell Biology.* Appleton & Lange, 1998.

Herman T, Horvitz HR: Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. Proc Natl Acad Sci U S A 1999;96:974.

Prockop DJ, Kivirikko KI: Collagens: molecular biology, diseases, and potential therapy. Annu Rev Biochem 1995;64:403.

Pyeritz RE: Ehlers-Danlos syndrome. N Engl J Med 2000;342:730.

Sage E: Regulation of interactions between cells and extracellular matrix: a command performance on several stages. J Clin Invest 2001;107:781. (This article introduces a series of six articles on cell-matrix interaction. The topics covered are cell adhesion and de-adhesion, thrombospondins, syndecans, SPARC, osteopontin, and Ehlers-Danlos syndrome. All of the articles can be accessed at www.jci.org.)

Scriver CR et al (editors): *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed. McGraw-Hill, 2001 (This comprehensive four-volume text contains chapters on disorders of collagen biosynthesis and structure, Marfan syndrome, the mucopolysaccharidoses, achondroplasia, Alport syndrome, and craniosynostosis syndromes.)

Selleck SB: Genetic dissection of proteoglycan function in *Drosophila* and *C. elegans.* Semin Cell Dev Biol 2001;12:127.