Collagens and proteoglycans of the corneal extracellular matrix

Y.M. Michelacci

Abstract

The cornea is a curved and transparent structure that provides the initial focusing of a light image into the eye. It consists of a central stroma that constitutes 90% of the corneal depth, covered anteriorly with epithelium and posteriorly with endothelium. Its transparency is the result of the regular spacing of collagen fibers with remarkably uniform diameter and interfibrillar space. Corneal collagen is composed of heterotypic fibrils consisting of type I and type V collagen molecules. The cornea also contains unusually high amounts of type VI collagen, which form microfibrillar structures, FACIT collagens (XII and XIV), and other nonfibrillar collagens (XIII and XVIII). FACIT collagens and other molecules, such as leucine-rich repeat proteoglycans, play important roles in modifying the structure and function of collagen fibrils. Proteoglycans are macromolecules composed of a protein core with covalently linked glycosaminoglycan side chains. Four leucine-rich repeat proteoglycans are present in the extracellular matrix of corneal stroma: decorin, lumican, mimecan and keratocan. The first is a dermatan sulfate proteoglycan, and the other three are keratan sulfate proteoglycans. Experimental evidence indicates that the keratan sulfate proteoglycans are involved in the regulation of collagen fibril diameter, and dermatan sulfate proteoglycan participates in the control of interfibrillar spacing and in the lamellar adhesion properties of corneal collagens. Heparan sulfate proteoglycans are minor components of the cornea, and are synthesized mainly by epithelial cells. The effect of injuries on proteoglycan synthesis is discussed.

Cornea

The cornea is a curved and transparent structure that, together with the sclera, composes the fibrous tunic of the eye. The fully developed cornea consists of a thick central portion known as the stroma, which constitutes about 90% of the corneal depth, covered anteriorly with stratified epithelium and posteriorly with endothelium. Both cell layers are separated from the stroma by basement membranes (Figure 1). The corneal curvature and transparency provide the initial focusing of light image into the eye. Corneal transparency is a unique physical property that results from the regular spacing of collagen fibers with remarkably uniform diameter and interfibrillar space. The specific fibrillar pattern appears in the cornea during embryonic development. Although many
relevant aspects of collagen biosynthesis have been recognized in recent years, we are still far from understanding the mechanisms underlying this uniformly oriented and perfectly calibrated synthesis and deposition of collagen fibers in the corneal matrix. Proteoglycans are thought to play an important role in this process. The present review deals with the most prominent macromolecules of the corneal extracellular matrix: collagen and proteoglycans.

**Collagen**

Collagen is the most abundant protein in vertebrates. It is a major component of the extracellular matrix and 26 different types of collagen have been described in vertebrates thus far (1,2). Table 1 summarizes the families and types of collagen. The fibril-forming collagens are the most abundant and include types I, II, III, V, VI and XI (3). Other collagens, known as FACIT collagens (fibril-associated collagens with interrupted triple helices), associate with the surface of collagen fibrils and modify their interactive properties (4,5). In addition, nonfibrillar collagens such as types XIII, XVII and XXV have been reported to be localized on the cell surface (6,7).

Despite the differences among collagens, all share a triple-helix structure composed of three polypeptides, known as α chains, consisting of Gly-X-Y repeats, where X is any amino acid, and Y is frequently proline or hydroxyproline. Each chain is a left-handed helix, and the three chains wind around each other in a right-handed superhelix.

**Collagen biosynthesis**

Collagen α chains are synthesized in the rough endoplasmic reticulum (RER) as long precursors named proα chains. For example, type I procollagen is a heterotrimer that contains two proα1(I) chains encoded by the COL1A1 gene, and one proα2(I) chain encoded by the COL1A2 gene (Figure 2). The preproα chains contain different structural domains: a) a signal sequence of 20 amino acids that is cleaved during RER transit, b) the N-terminal propeptide, c) a short non-triple-helix domain that contains the site of proteolytic cleavage of N-terminal peptides, d) the major triple-helix domain containing 1014 amino acid residues (each chain), characterized by the Gly-X-Y repeats, e) a 28-residue telopeptide, and f) the final 220 residues that form a globular structure containing disulfide bonds (C-terminal propeptide).

A large number of enzymes are involved in the post-translation modifications of the α chains (Figure 3). Some of these modifications, such as hydroxylation of proline to form 3- or 4-hydroxyproline, hydroxylation of lysine to hydroxylysine, and glycosylation of hydroxylysine, occur during the elongation of the peptide in the RER, and are
Figure 2. Genes that encode type I preprocollagen α chains (top) and domain structure of procollagen molecule (bottom). A, Signal sequence that is cleaved during rough endoplasmic reticulum transit; B, N-terminal propeptide; C, short non-triple-helix domain that contains the site of proteolytic cleavage of N-terminal peptides; D, major triple-helix domain, characterized by the Gly-X-Y repeats; E, telopeptide; F, N-terminal propeptides that contain disulfide bonds.

Figure 3. Biosynthesis of fibrillar collagens. Modifications of the procollagen peptide in the rough endoplasmic reticulum (RER) include hydroxylation, glycosylation, and disulfide bond formation. Interchain disulfide bonds between the C-terminal peptides align the three chains and initiate formation of the triple helix. The process propagates towards the N-terminus of the molecule end. N-linked oligosaccharides are transferred to the propeptides in the RER and processed in the Golgi apparatus. Upon secretion, propeptides are cleaved, allowing lateral alignment and crosslinking. See text for details.
followed by formation of the triple helix. Prolyl 4-hydroxylase, located either in the lumen or on the inner surface of the RER membrane, is a tetramer containing two α subunits (64 kDa) and two β subunits (60 kDa) which are products of different genes (8). The β subunits contain the C-terminal amino acid sequence Lys-Asp-Glu-Leu (KDEL) that confers the ability to reside permanently in the RER. Prolyl 3-hydroxylase has a molecular mass of 160 kDa, and details about the structure of this enzyme have been recently reported (9). Lysyl hydroxylase is a homodimer, the monomer of which has a molecular mass of 85 kDa (10, see Ref. 11 for a review on these enzymes).

The extent of post-translation modification of prolyl and lysyl residues in the diverse collagen types differs substantially. For example, the relative extent of lysyl hydroxylation in type IV collagen is very high (more than 90% of the residues at position Y are hydroxylated), while in types I and III only 15% are hydroxylated. The three collagen hydroxylases require Fe2+, 2-oxoglutarate, molecular oxygen (O2) and ascorbate as co-factors.

The hydroxylation of proline at the 4 position is essential to provide thermal stability to the triple helix. Hydroxylation of lysyl residues provides a substrate for both glycosylation and formation of stable cross-links, important to the tissue tensile strength (12).

The glycosylation of hydroxylysyl residues in collagen requires two enzymes: hydroxylysyl galactosyltransferase and galactosylhydroxylysyl glucosyltransferase (13). The former adds a galactose residue to the 5-OH of hydroxylysine, and the latter transfers glucose to the galactose residue. The distribution of monosaccharides and disaccharides is influenced by the collagen type and is carried out only on non-triple-helix substrates.

Furthermore, asparagine-linked oligosaccharides may be present in the propeptide extensions (14). These oligosaccharides are synthesized on a dolichol lipid intermediate in the RER membrane and transferred intact to the proα chain (15). There is initial cleavage of the oligosaccharide while the chain is still in the RER, and further processing in the Golgi apparatus (16).

After synthesis is completed, the globular domains of the proα chains fold and are stabilized by disulfide bonds. The triple-helix formation begins at the C-terminal end of the molecule and is propagated towards the N-terminal end. The propagation of the triple-helix structure requires all peptide bonds involving the prolyl residues to be in the trans form, a configuration possibly produced by the enzyme peptidyl prolyl cis-trans isomerase (17).

Procollagen molecules are translocated to the Golgi apparatus and then packaged in secretory vesicles that fuse with the cell membrane and release their contents into the extracellular environment. In the Golgi apparatus, N-linked oligosaccharides are trimmed and synthesis of a high-mannose structure occurs. In addition, phosphorylation of certain serine residues and sulfation of some tyrosine residues occur in some collagens.

Once outside the cell, procollagens undergo proteolytic conversion to collagen, form fibrils, interact with other proteins, are stabilized by intermolecular crosslinks, and are eventually degraded. Propeptides are removed by two enzymes: procollagen N-proteinase and procollagen C-proteinase (18). Once cleaved, collagen molecules rapidly aggregate into ordered structures. Fibril formation is a nonenzymatic process and the nature of the interactions that govern fibril diameter is not completely understood. Interactions with other types of collagen or with proteoglycans are thought to control the rate of fibrillogenesis and the fibril diameter (19).

Collagen molecules in fibrillar array become substrate for the enzyme lysyl oxidase
that oxidatively deaminates certain lysyl and hydroxylysyl residues (20), forming reactive aldehydes that condense with lysyl or hydroxylysyl residues in adjacent molecules to form divalent crosslinks (Figure 3). Lysyl oxidase is a monomeric enzyme of 32 kDa (21) and requires pyridoxyl and copper as co-factors.

**Corneal collagens**

Table 2 summarizes the different collagen types described in vertebrate corneas, as well as the relative amounts of the most abundant ones. Normal human corneal stroma is rich in type I collagen, but also contains relatively large amounts of type V (22) and type VI collagens (23). Type III is present at low proportions but increases during wound healing, inflammation and several pathological conditions.

Corneal collagen fibrils are composed of type I collagen molecules incorporated together with those of type V collagen into heterotypic fibrils (24). Type VI collagen forms microfibrillar structures by lateral aggregation, and types XII and XIV collagens are FACIT collagens (25). By analogy with type IX collagen in cartilage, the N-terminal domains of type XII and XIV collagens might be expected to protrude from the surface of corneal collagen fibrils where they would be able to interact with proteins in the corneal stroma.

Collagen XIII, a nonfibrillar collagen that is bound to the cell membrane through a single transmembrane domain, is widely expressed in the human eye and has been found in the posterior two thirds of the corneal stroma (7).

Collagen XVIII is a proteoglycan localized in the basement membrane of epithelia and vascular endothelium. Collagen XVIII is the only currently known collagen that carries heparan sulfate glycosaminoglycan side chains (26). Collagen XVIII gained public attention when it was discovered that part of its C-terminal non-triple-helix domain is endostatin, a potent inhibitor of angiogenesis and tumor growth released by cleavage of an Ala-His peptide bound (27). The processing of collagen XVIII to endostatin may represent a local control mechanism for the regulation of angiogenesis. In the cornea, collagen XVIII was immunolocalized to the human corneal epithelium, epithelial basement membrane, and Descemet membrane (28), and lack of collagen XVIII was shown to cause eye abnormalities (29).

**Proteoglycans**

In addition to FACIT collagens, other molecules have important roles in modifying the structure and function of collagen fibrils. These include some leucine-rich repeat proteoglycans (for a review, see Ref. 30).

Proteoglycans are macromolecules composed of a protein core with covalently linked glycosaminoglycan side chains (31). Studies of the glycosaminoglycans from bovine (32), rabbit (33), chicken (34), monkey (35) and human (36) corneas have shown that dermatan sulfate and keratan sulfate are the predominant components, with smaller amounts of heparan sulfate. The keratan sulfate chain

<table>
<thead>
<tr>
<th>Table 2. Corneal collagens.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td>V</td>
</tr>
<tr>
<td>VI</td>
</tr>
<tr>
<td>VII</td>
</tr>
<tr>
<td>VIII</td>
</tr>
<tr>
<td>IX</td>
</tr>
<tr>
<td>XII</td>
</tr>
<tr>
<td>XIII</td>
</tr>
<tr>
<td>XIV</td>
</tr>
<tr>
<td>XVII</td>
</tr>
<tr>
<td>XVIII</td>
</tr>
</tbody>
</table>
is attached to the proteoglycan core protein by an N-acetylglucosamine of a mannose-containing linkage oligosaccharide, which is N-linked to an asparagine residue of the protein (37). The linkage oligosaccharide is similar to biantennary complex-type N-linked oligosaccharides found in glycoproteins, with the keratan sulfate extending one branch and sialic acid terminating the second branch (38) (Figure 4).

In 1992, Blochberger et al. (39) isolated a cDNA clone encoding the core protein of a chick corneal keratan sulfate proteoglycan, which they named lumican. Three keratan sulfate proteoglycan core proteins with different primary structures, named 37A, 37B and 25, were isolated from bovine cornea. These proteins were also cloned, and 37B was identified as the bovine lumican (40), with 68% identity to chicken lumican. The keratan sulfate proteoglycans containing the 37A and the 25 cores were named keratocan and mimecan (or osteoglycin), respectively (41,42) (Figure 5). It seems that although expressed in other tissues, only in the cornea are lumican, keratocan, and mimecan glycosylated with sulfated keratan sulfate chains.

The corneal proteoglycans are thought to play a role in collagen fibrillogenesis and matrix assembly. Morphological studies have shown both dermatan sulfate and keratan sulfate proteoglycans to be associated with specific bands of corneal collagen fibrils (43,44), suggesting that proteoglycan-collagen interactions may play a role in corneal fibril assembly, matrix organization and ultimately corneal transparency.

The crucial role of lumican in the regulation of collagen assembly into fibrils has been established by studies on mice homozygous for a null mutation in lumican. These animals presented skin laxity and fragility and bilateral corneal opacity (45). In addition, it was shown that people who have mutations in KERA, the gene encoding keratocan, have cornea plana (46), in which the forward convex curvature is flattened leading to a decrease in refraction. Furthermore, it was recently shown that mimecan-deficient mice have thicker collagen fibrils in
Corneal collagens and proteoglycans

The fourth leucine-rich repeat proteoglycan of the cornea is a dermatan sulfate proteoglycan (Figure 4). The corneal dermatan sulfate chains are linked to a 40-kDa protein, forming a proteoglycan of 100-150 kDa containing one or two glycosaminoglycan side chains (52). The core protein was originally characterized by its antigenic properties as decorin (Figure 5), a proteoglycan present in most connective tissues and structurally related to other small interstitial proteoglycans. Later, Li et al. (53) confirmed this identity by cloning and sequencing a cDNA encoding the core protein of chick corneal dermatan sulfate proteoglycan.

Hahn and Birk (54) studied the effects of β-D-xyloside on the fibril organization of avian cornea. This compound interferes with xylose-mediated O-linked proteoglycan synthesis and thus disturbs dermatan sulfate proteoglycan synthesis, but does not modify the keratan sulfate proteoglycan formation. The collagen fibril diameters were unaltered, but a disruption in lamellar organization occurred, suggesting that dermatan sulfate proteoglycans are not involved in the regulation

<table>
<thead>
<tr>
<th>Proteoglycan</th>
<th>Core protein (kDa)</th>
<th>Glycosaminoglycan (number of chains)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decorin</td>
<td>40</td>
<td>Dermatan/chondroitin sulfate (1)</td>
</tr>
<tr>
<td>Biglycan</td>
<td>40</td>
<td>Dermatan/chondroitin sulfate (2)</td>
</tr>
<tr>
<td>Fibromodulin</td>
<td>42</td>
<td>Keratan sulfate (2-3)</td>
</tr>
<tr>
<td>Lumican</td>
<td>38</td>
<td>Keratan sulfate (3-4)</td>
</tr>
<tr>
<td>Keratocan</td>
<td>38</td>
<td>Keratan sulfate (4-5)</td>
</tr>
<tr>
<td>Osteoadherin</td>
<td>42</td>
<td>Keratan sulfate (2-3)</td>
</tr>
<tr>
<td>Class II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epiphycan (PG-Lb)</td>
<td>35</td>
<td>Dermatan/chondroitin sulfate (2-3)</td>
</tr>
<tr>
<td>Mimecan (osteoglycin)</td>
<td>35</td>
<td>Keratan sulfate (2-3)</td>
</tr>
</tbody>
</table>

Figure 5. Domain structure of decorin, a prototype member of the leucine-rich repeat proteoglycan family, and classification of the leucine-rich repeat proteoglycans based on amino acid sequence homology. Roman numerals indicate decorin domains. In addition to the glycosaminoglycan chain (dermatan sulfate or chondroitin sulfate), N-linked oligosaccharides are also present.
of collagen fibril diameter, but are important to fibril-fibril spacing and lamellar cohesiveness. In corneal explants from embryonic chicken, an increased synthesis of keratan sulfate proteoglycan and a decreased synthesis of dermatan sulfate proteoglycan coincide with the onset of tissue transparency, again suggesting a correlation between proteoglycan composition and corneal transparency (55).

Although several groups have studied the glycosaminoglycan composition of bovine, rabbit, chicken, and monkey cornea, as well as the biosynthesis of proteoglycans by rabbit and embryonic chicken corneal explants (56), only a few reports concerning human corneal glycosaminoglycans have appeared (57). The glycosaminoglycan (36) and proteoglycan (58) composition of human cornea, as well as the glycosaminoglycans synthesized by human cornea explants under tissue culture conditions have been recently described.

Corneal explants submitted to tissue culture conditions for three days presented the same glycosaminoglycan composition as the fresh cornea (dermatan sulfate and keratan sulfate, about 50% each, 1.5 mg/g tissue wet weight). The glycosaminoglycan synthesis was maintained in vitro for weeks under these tissue culture conditions, as indicated by the metabolic incorporation of $[^{35}S]$-sulfate. Nevertheless, during a 24-h incubation the corneal explants synthesized glycosaminoglycans which varied in relative proportions from those present in intact corneas. The labeling of dermatan sulfate was higher than that of keratan sulfate. Furthermore, $[^{35}S]$-heparan sulfate also appeared, suggesting a higher synthesis rate for dermatan sulfate and heparan sulfate.

The main glycosaminoglycan synthesized by epithelial cells isolated from human cornea is heparan sulfate. Conversely, stromal cells synthesize mainly dermatan sulfate and keratan sulfate when both epithelial and endothelial cell layers are removed (36).

The quantitative and qualitative patterns of proteoglycan synthesis, and possibly of collagen fibril deposition, change when the cornea is submitted to injuries, such as epithelial debridement (36,58) and laser refractive surgery (59), suggesting that the keratocyte program for synthesis and deposition of extracellular matrix, which was followed during normal development leading to corneal transparency, is hardly recovered after injury.

Thus, studies to determine the structure of corneal fibrils and to identify in three dimensions the binding of FACIT collagens, leucine-rich repeat proteoglycans, and other macromolecules, are directly relevant to understanding the assembly of the corneal stroma in health and disease.

References

isomerase are products of the same gene. EMBO Journal, 6: 643-649.


