Preface

Collagen type I is the most abundant protein in mammals. It has outstanding mechanical properties and is present in virtually every extracellular tissue with mechanical function. In tendons and ligaments, collagen transmits the force from muscles to bones and stores elastic energy. Smooth walking would not be possible without these properties. Collagen also represents most of the organic matrix of bones and tooth dentin and confers them their fracture resistance. It is a major constituent of skin and blood vessels and is even present in muscles, which could not function without a collagen-rich matrix around the contractile cells. A slightly different type of collagen (type II) is a critical component of a tissue as soft as articular cartilage. The function of collagen is not only mechanical. In the cornea of the eye, for example, the ordering of collagen fibrils confers transparency in addition to mechanical stability.

The versatility of collagen as a building material is mainly due to its complex hierarchical structure. Adaptation is possible at every level, leading to a great variety of properties, to serve a given function. The basic building block of collagen-rich tissues is the collagen fibril, a fiber with 50 to a few hundred nanometer thickness. These fibrils are assembled into composite materials with a variety of more complex structures, which may have anisotropic or nearly isotropic mechanical properties, depending on fiber arrangement. In bone and dentin, collagen is combined with mineral to yield very stiff tissues. In tendon or cornea, collagen is combined with other organic molecules, such as proteoglycans.

Given the size of the collagen fibrils, the important structures are often in the nanometer scale, and recent progress in characterization methods has revealed many details of these structures and of how they relate to the mechanical behavior. Moreover, a large number of molecules have recently been identified as members of the collagen family. This is described in the first part of this book (Chapters 2–4), which is devoted to the structure and biochemistry of collagens. The second part of this book (Chapters 5–8) discusses mechanical properties and the mechanisms at the origin of deformation, fatigue and fracture of collagen-based materials, both from an empirical and from a theoretical viewpoint. The major part of the book (Chapters 9–17) addresses a particular collagen-based tissue per chapter, including tendons and ligaments, artery walls, cornea, bone and dentin, among others. The last two chapters focus on the more special issues of genetic collagen diseases and

collagen-based tissue engineering. Chapter 1, finally, is meant as an introduction to the subject in general and to the other chapters in the book. It also gives elementary definitions of mechanical quantities needed throughout.

In this way, this text approaches collagen-based tissues from very different perspectives, highlighting structure and biochemistry, general principles of mechanical behavior, as well as structure, composition and mechanical behavior of a number of important tissues in our body. We hope that by this interdisciplinary approach, the book will be useful as an introduction and as a reference for advanced students, researchers and engineers in very diverse fields, such as materials science and engineering, collagen biochemistry and biophysics, as well as tissue engineering and regenerative medicine. This diversity is also reflected in the different backgrounds of the authors, who are all well-recognized specialists in the fields which they are covering. In addition to providing the best introduction and up-to-date reference, we really hope to transfer to the reader some of our excitement about the beauty of structure and mechanics of collagen-based tissues.

Chapter 2
Collagen Diversity, Synthesis and Assembly

D.J.S. Hulmes

Abstract The vertebrate collagen superfamily now includes over 50 collagens and collagen-like proteins. Here, their different structures are described, as well as their diverse forms of supramolecular assembly. Also presented here are the various steps in collagen biosynthesis, both intracellular and extracellular, and the functions of the collagen-specific post-translational modifications. Assembly of collagen fibrils, both in vitro and in vivo, is reviewed, including the mechanisms that control this process and the interactions involved. Finally, recent developments in the supramolecular assembly of collagen-like peptides are discussed.

2.1 Introduction

Collagens come in all shapes and sizes. The hallmark of a collagen is a molecule that is composed of three polypeptide chains, each of which contains one or more regions characterized by the repeating amino acid motif (Gly-X-Y), where X and Y can be any amino acid. This motif allows the chains to form a right-handed triple-helical structure (Fig. 2.1), with all glycine residues buried within the core of the protein, and residues X and Y exposed on the surface. Depending on the genetic type of collagen, this triple-helical motif can be a major or minor part of the molecule, other regions consisting of different non-collagenous domains. While the presence of a triple-helical motif is a necessary condition for being called a collagen, it is not a sufficient one. Some proteins contain these motifs but are not called collagens, mainly because they were named on the basis of their specific biological functions, rather than the structural role traditionally ascribed to collagens. But here also the distinction has become blurred, with new biological functions being discovered for both collagenous and non-collagenous domains. The question of deciding what is and what is not a collagen is best avoided; here I refer simply to the collagen superfamily, being all proteins containing a collagen-like triple-helical motif.

P. Fratzl (ed.), *Collagen: Structure and Mechanics*,

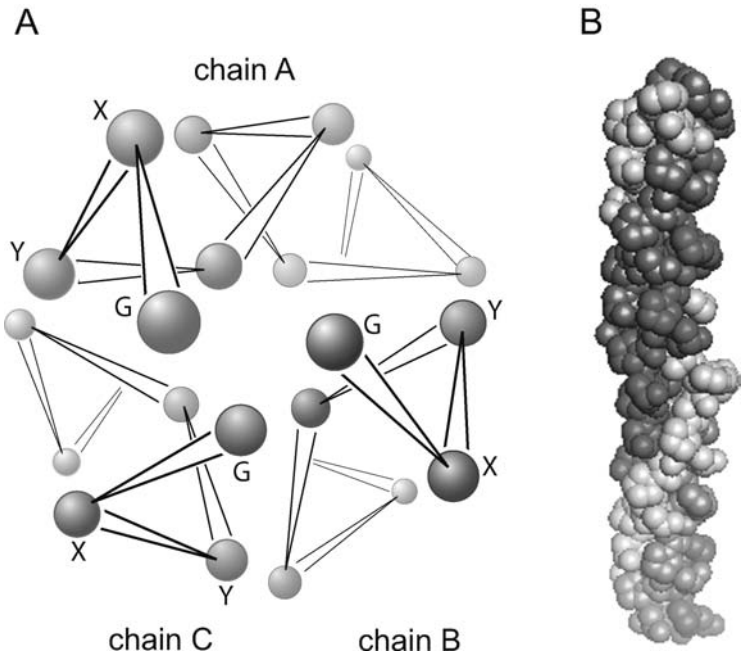

Fig. 2.1 Triple-helical structure of collagen. **A** Viewed along the molecular axis (α-carbons only), showing the paths of the individual polypeptide chains and the locations of residues in the Gly-X-Y triplets (G = Gly; from (Beck and Brodsky, 1998), with permission). **B** Viewed from the side (molecule tilted toward the reader at the top; space-filling representation), showing the right-handed helical twist

2.2 Fibrillar Collagens

In humans, there are currently 28 different proteins known as collagens, as well as about the same number of other members of the collagen superfamily (Kadler et al., 2007; Myllyharju and Kivirikko, 2004; Ricard-Blum et al., 2005). They can be grouped into a number of subfamilies (Table 2.1). From the biomechanical point of view, of most interest are the fibrillar collagens (Hulmes, 2002; Kadler, 1995). These proteins give rise to classical collagen fibrils, as seen by electron microscopy (Fig. 2.2), characterized by a repeating banding pattern with a so-called *D* periodicity of 64–67 nm, depending on the tissue. Within the fibril, collagen molecules of length 300 nm and width 1.5 nm are staggered with respect to their neighbors by multiples of *D*. The three-dimensional structure of collagen fibrils is discussed in greater detail in Chapter 3. The classical fibrillar collagens include: type I, the most widely occurring collagen found in skin, tendon, bone, cornea, lung and the vasculature; type II, which has a more specific tissue distribution being limited essentially to cartilage; type III, found in relatively elastic tissues such as embryonic skin, lung

Table 2.1 Collagens and collagen-like proteins in vertebrates

Sub-family	Members
Fibrillar collagens	Types I, II, III, V, XI, XXIV and XXVII
Fibril associated and related collagens	Types IX, XII, XIV, XVI, XIX, XX, XXI and XXII
Beaded filament forming collagen	Type VI
Basement membrane and associated collagens	Type IV, VII, XV and XVIII
Short chain collagens and related proteins	Types VIII and X; C1q; hibernation-related proteins HP-20, HP-25 and HP-27; emilins 1 and 2; adiponectin; CTRPs 1-7: inner ear (saccular) collagen
Transmembrane collagens and collagen-like proteins	Types XIII, XVII, XXIII and XXV/CLAC-P; ectodysplasins; macrophage scavenger receptors I-III; MARCO; SRCL; gliomedin; CL-P1
Collectins and ficolins	Mannan binding protein; surfactant proteins A and D; conglutinin; CL-43; CL-46; CL-L1; CL-P1; L-, M- and H-ficolins
Other collagens and collagen-like proteins	Emu1; collagen XXVI/Emu2; collagen XXVIII; acetylcholinesterase tail subunit

and blood vessels; type V, found as a quantitatively minor collagen in association with collagen I, with particularly high amounts in cornea; and type XI, a quantitatively minor component of cartilage in association with collagen II.

As in all collagens, each fibrillar collagen molecule consists of three polypeptide chains, called α chains. Molecules can be homotrimeric, consisting of three identical α chains, as in collagens II and III, or heterotypic, consisting of up to three genetically distinct α chains. Individual α chains are identified by the following nomenclature: α*n*(N), where N is the Roman numeral indicating collagen type and *n* is the number of the α chain. Thus, the chain composition of collagen II is $[\alpha1(II)]_3$, while that of collagen I, a heterotrimer with two identical α1 chains and a third distinct α2 chain, is $[\alpha1(I)]_2\alpha2(I)$. While collagen I has only two different α chains, collagen V has three, leading to different chain stoichiometries such as $[\alpha1(V)]_2\alpha2(V)$ and α1(V)α2(V)α3(V). Collagen XI has three different α chains, where the α3 chain is essentially the same as α1(II). In addition, there are several variants of the α1(XI) and α2(XI) chains due to alternative splicing, corresponding to the variable region (Fig. 2.2), of the corresponding mRNAs. To even further increase diversity of these quantitatively minor fibrillar collagens, hybrid forms of collagen V and XI molecules have also been reported.

The common characteristic of the classical fibrillar collagens is a long central triple-helical region in each α chain, consisting of a continuous $(Gly\text{-}X\text{-}Y)_n$ repeat, where *n* is 337–343 (depending on collagen type). In the case of collagens I, II and III, this region is flanked by short non-helical regions called telopeptides, typically

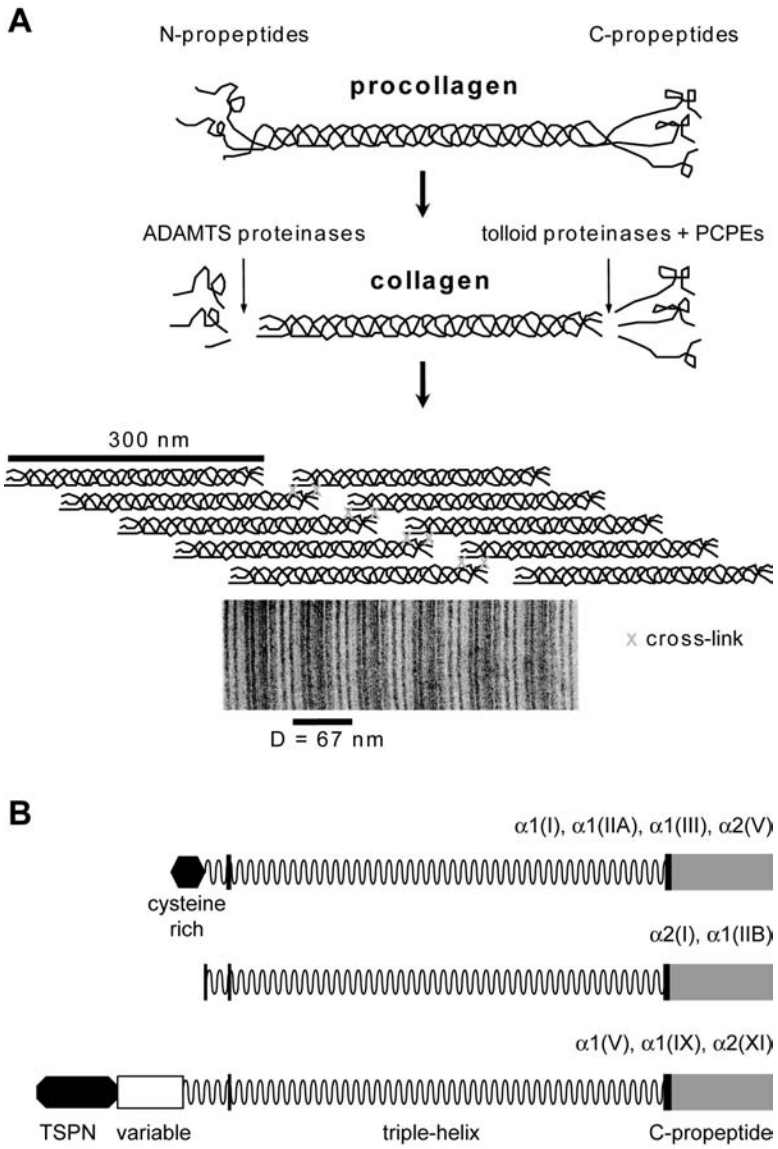

Fig. 2.2 Fibrillar procollagens and fibril assembly. **A** N- and C-terminal processing leads to spontaneous assembly of collagen fibrils, which are subsequently stabilized by the formation of covalent cross-links. **B** Domain structures of fibrillar procollagen chains

about 20 residues in length, at both N- and C-termini. All fibrillar collagens are synthesized in the form of soluble precursor molecules, called procollagens, with large N- and C-terminal propeptide domains. The C-propeptides are removed during the later stages of biosynthesis (Section 2.4.4), usually by specific metalloproteinases, leaving the short C-telopeptides. The extent of N-terminal processing depends on

collagen type. While for procollagens I, II and III the N-propeptides are completely removed (albeit relatively slowly in the case of procollagen III) leaving short N-telopeptides, N-terminal processing of procollagens V and XI leaves large N-terminal extensions, which help modulate fibril formation (Section 2.5.2).

Little is known about the recently discovered fibrillar-like collagens XXIV and XXVII, though these appear to be associated with types I and II containing tissues, respectively. They differ from the classical fibrillar collagens in that the (Gly-X-Y) region is relatively short (329 triplets), and this region is interrupted by one (collagen XXIV) or two (collagen XVII) short imperfections in the (Gly-X-Y) repeat. Collagen XXVII has recently been shown to form non-striated filamentous structures (Plumb et al., 2007).

2.3 Non-fibrillar Collagens

While fibrillar collagens are the main focus of this chapter, it is of interest to compare these with the diverse forms of assembly of the non-fibrillar collagens (Ricard-Blum et al., 2000).

2.3.1 Basement Membrane and Associated Collagens

The non-fibrillar collagen, about which most is known, is type IV or basement membrane collagen (Hudson et al., 2003). Basement membranes (or basal laminae) are specialized structures found at tissue boundaries, underlying epithelial, endothelial, fat, muscle and nerve cells. Seen by electron microscopy, they are relatively thin sheets (typically 40–50 nm) composed of collagen IV, laminins, heparin sulphate proteoglycans and nidogens. Collagen IV molecules are longer than the fibrillar collagens, and they contain several discontinuities in the (Gly-X-Y) repeat. Also, unlike fibrillar collagens, there is no precursor form of collagen IV and the N- and C-terminal extensions, called respectively the 7S and NC1 domains, are intimately involved in supramolecular assembly. Collagen IV molecules associate via tetramerization of 7S domains, dimerization of NC1 domains, triple-helical interactions and interactions with other basement membrane components to form an open meshwork structure (Fig. 2.3). This structure is stabilized by disulphide bonding in the 7S region, as well as by lysine-derived cross-links initiated by lysyl oxidase (Section 2.4.5; also Chapter 4). There are six different collagen IV α chains, resulting in three types of heterotrimer: $[\alpha1(IV)]_2\alpha2(IV)$, $\alpha3(IV)\alpha4(IV)\alpha5(IV)$ and $[\alpha5(IV)]_2\alpha6(IV)$.

Collagen VII has the longest triple-helical region amongst vertebrate collagens, about 420 nm in length, including interruptions, and is flanked by non-collagenous NC1 and NC2 domains at the N- and C-termini, respectively. This collagen is found underlying the basement membrane at the dermal–epidermal junction, in the form of anchoring filaments (Fig. 2.3). Mutations in collagen VII are the cause of

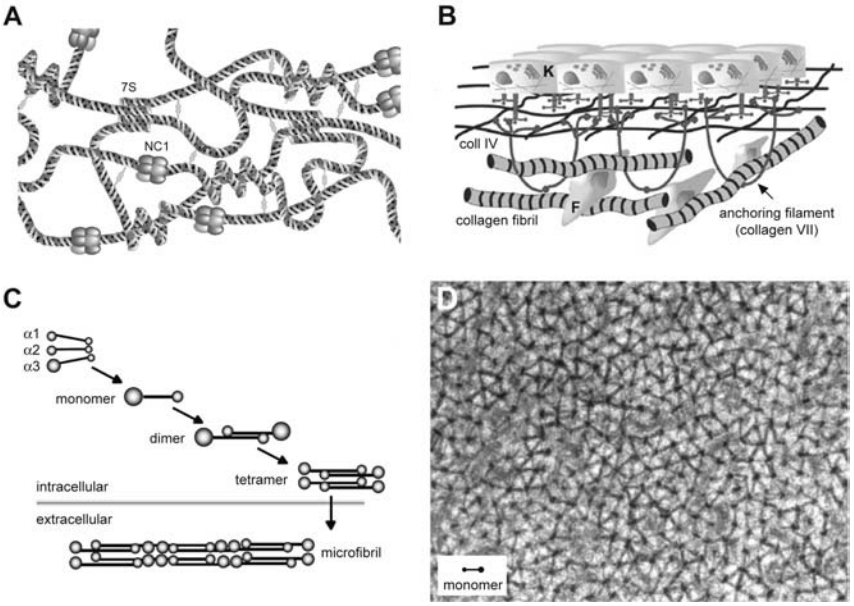

Fig. 2.3 Assembly of non-fibrillar collagens. **A** Basement membrane collagen IV (from (Hudson et al., 2003), with permission). **B** Collagen VII anchoring filaments connecting the epidermal basement membrane to collagen fibrils in the dermis (from (Brittingham et al., 2006), with permission). **C** Intra- and extracellular steps in the assembly of collagen VI microfibrils (from (Baldock et al., 2003), with permission). **D** Hexagonal network formed by collagen VIII in Descemet's membrane of the cornea (courtesy of R. Bruns)

severe blistering disorders called epidermolysis bullosa. Within anchoring filaments, about two molecules in length, bundles of collagen VII molecules aligned in register are arranged tail to tail with a short C-terminal overlap. Antiparallel assembly is triggered by proteolytic processing of the NC2 domain, while the large NC1 domain remains intact and interacts with collagen IV and laminin V (Brittingham et al., 2006).

Collagens XV and XVIII consist of several collagenous domains and together are referred to as multiplexins (for *multiple* triple *he*lices with *in*terruptions). Both are associated with basement membranes, and both carry covalently linked glycosaminoglycan chains, and are therefore also proteoglycans. In addition, both include a C-terminal fragment, endostatin, that is proteolytically cleaved from the molecule and has anti-angiogenic properties (Ricard-Blum et al., 2005).

2.3.2 Collagen VI

Collagen VI is a relatively ubiquitous collagen with important roles in maintaining tissue integrity (Baldock et al., 2003; Knupp and Squire, 2005). The molecules contain relatively short triple-helical regions, about one third the length of fibrillar

collagens, with large N- and C-terminal regions made up mostly of von Willebrand factor A domains. Each molecule is a heterotrimer α1(VI)α2(VI)α3(VI). In the case of the α3(VI) chain, the triple-helical region makes up less than 10% of the entire amino acid sequence. As with several other collagens (e.g., types II, IX, XI, XII and XIII), further diversity in the α chains is introduced at the transcriptional level through the formation of alternatively spliced variants. Collagen VI molecules assemble into the so-called beaded filaments, with a periodicity of 110 nm (Fig. 2.3). Unlike fibrillar collagens, there is no enzymatic processing of the N- and C-terminal non-collagenous regions. In addition, unlike all other collagens, supramolecular assembly begins inside the cell with the formation of antiparallel dimers and tetramers. Tetramers stabilized by disulphide cross-linking then associate in the extracellular matrix to form beaded filaments or microfibrils.

2.3.3 Collagens VIII and X

Collagens VIII and X, the so-called short-chain collagens, have specific tissue locations, underlying endothelial cells and in the hypertrophic zone of cartilage during endochondral ossification, respectively. The triple-helical region, which contains many interruptions, is about half the length of a fibrillar collagen. While collagen X is an α1(X) homotrimer, collagen VII, with both α1(VIII) and α2(VIII) chains, appears to occur as both homotrimers and heterotrimers. Molecules have both N- and C-terminal non-collagenous regions, the latter corresponding to the widely occurring C1q domain (Ghai et al., 2007). There is no proteolytic processing of these regions, rather they help in the formation of hexagonal supramolecular networks in, for example, Descemet's membrane of the cornea (collagen VIII; (Stephan et al., 2004); Fig. 2.3) and calcifying cartilage (collagen X; (Kwan et al., 1991)).

2.3.4 FACITs

The collagens known as FACITs (Fibril Associated Collagens with Interrupted Triple helices) have become a relatively large group (currently eight different types; see Ricard-Blum et al., 2005). The founder member, collagen IX, is an important component of cartilage collagen fibrils, along with collagens II and XI. Unlike the latter, however, collagen IX does not self-assemble into fibrils. The collagen IX molecule is composed of three relatively short collagenous regions (with interruptions) and four non-collagenous regions. The molecule is an α1(IX)α2(IX)α3(IX) heterotrimer, and the α2(IX) chain sometimes carries a chondroitin/dermatan sulphate glycosaminoglycan chain, thereby making collagen IX a part-time proteoglycan. Collagen IX molecules coat the surface of collagen II/XI cartilage fibrils, with the large N-terminal NC4 domain of the α1(IX) chain being available for interactions with other extracellular matrix components. The presence of the NC4 domain in the α1(IX) chain is itself subject to regulation

due to the use of alternative transcriptional start sites. Collagen IX is covalently cross-linked to both collagens II and XI in the cartilage fibril, through lysine-derived cross-links; recent studies have shown how flexibility within the non-collagenous domains allows these cross-links to form (Eyre et al., 2004). Strongest homology within the FACITs is in the C-terminal collagenous (COL1) and non-collagenous (NC1) domains, and all have a thrombospondin N-terminal-like domain (also found in the fibrillar collagens V and XI) just before the first collagenous domain (starting from the N-terminus). Collagen IX can be considered as the founder member of a subgroup of FACITs with three or more collagenous domains, which also includes collagens XVI, XIX and XXII (Ricard-Blum et al., 2005). Unlike collagen IX, however, collagens XVI, XIX and XXII do not associate specifically with collagen fibrils and have distinct tissue distributions. Collagen XXII, for example, is specifically localized to tissue junctions (Koch et al., 2004).

Collagen XII has also been shown to be associated with the collagen fibril surface, and has been localized to dense connective tissues rich in collagen I, such as tendons, ligaments, skin and cornea. The molecule is an $[\alpha 1(XII)]_3$ homotrimer, each chain consisting of two collagenous domains (making up only about 10% of the amino acid sequence) and three non-collagenous domains. The non-terminal non-collagenous domain (NC3) is particularly long and includes multiple fibronectin type III and von Willebrand factor A domains. Two alternative spliced forms of collagen XII are known, differing in the length of the NC3 domain, the longest form being a proteoglycan-like collagen IX. Collagen XII expression is upregulated by mechanical stress (Chiquet et al., 2003). Collagen XII can be considered as the founder member of the second subgroup of FACITs, characterized by the presence of just two collagenous domains. Collagens XIV and XX are closely related, while collagen XXI is by far the smallest member of this subgroup.

2.3.5 Other Collagens and Collagen-Like Proteins

The transmembrane collagens and collagen-like proteins have become one of the largest groups and have diverse functions in cell adhesion and signaling (Franzke et al., 2005). Lastly, the collagen-like proteins of the immune system (C1q, collectins and ficolins) have collagen-like stalks and globular heads, the former leading to self-association into bundles containing up to six molecules (Holmskov et al., 2003).

2.4 Collagen Biosynthesis

Collagen biosynthesis has been studied in greatest detail with regard to the fibrillar collagens. The process is complex, involving numerous intracellular and extracellular steps, all of which contribute to the structure and biomechanical properties of the final fibrils. The first event following synthesis of procollagen α chains on the ribosome is their import into the rough endoplasmic reticulum. There they undergo a series of post-translational modifications resulting in the assembly of procollagen

molecules (Fig. 2.4) (Myllyharju, 2005). These steps include modification of proline residues to hydroxyprolines, modification of lysines to hydroxylysines, N- and O-linked glycosylation, trimerization, disulphide bonding, prolyl *cis–trans* isomerization and folding of the triple helix. Molecules then transit the Golgi network where they are packaged into secretory vesicles prior to export into the extracellular matrix. Procollagen processing occurs during or shortly after secretion followed by assembly of fibrils. Finally, fibrils are stabilized by the formation of covalent cross-links.

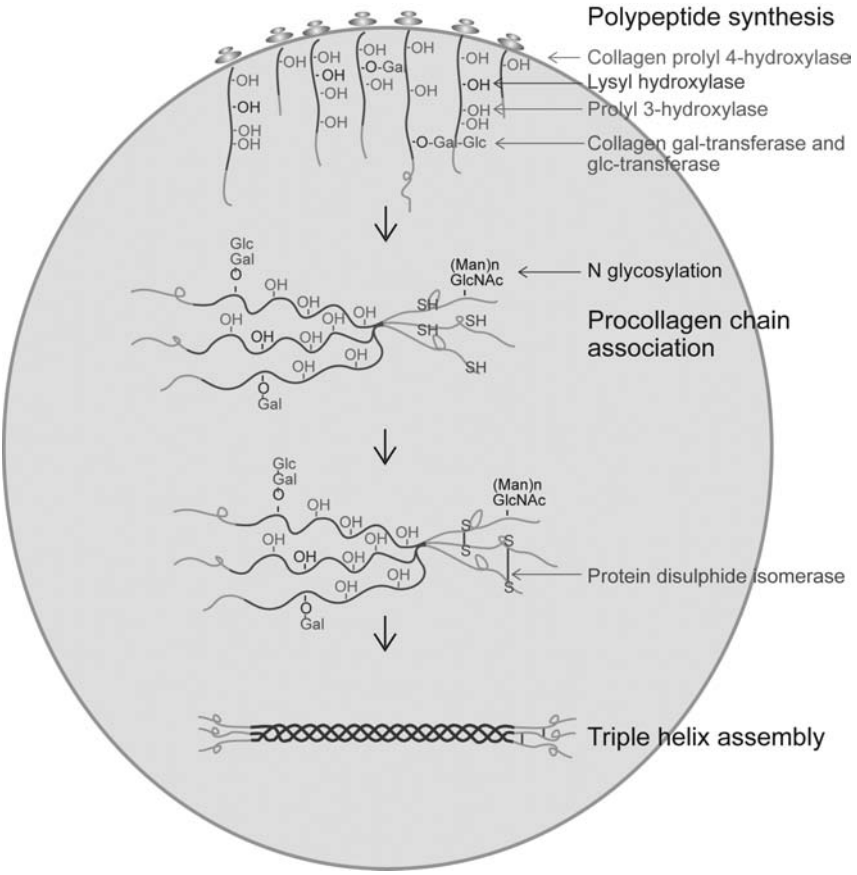

Fig. 2.4 Post-translational modifications and assembly of the procollagen molecule in the endoplasmic reticulum (from (Myllyharju, 2005), with permission)

2.4.1 Post-translational Modifications of Polypeptide Chains

It has long been known that stability of the collagen triple helix is related to the total content of the amino acids proline and hydroxyproline, which together make up about 20% of the total amino acids in human fibrillar collagens. This is

a result of the cyclic nature of the side chains that restricts flexibility about the peptide bond. Prolyl hydroxylation provides further stabilization, as shown by the approximately 30°C drop in melting temperature (see below) when this process is prevented. Hydroxylation occurs mainly on prolines in the Y position of the Gly-X-Y triplet, through the action of prolyl 4-hydroxylases (P4H), and also to a much lesser extent on prolines in the X-position, through the action of prolyl 3-hydroxylases (P3H) (Fig. 2.5). The enzymes responsible are 2-oxoglutarate and non-heme-Fe(II)-dependent dioxygenases, which require Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate for activity (Myllyharju, 2005). In humans, three forms of P4H (Myllyharju, 2005) and three forms of P3H (Morello et al., 2006) have been described. P4H is a tetramer consisting of two α subunits and two β subunits. The β subunit is common to all forms of the enzyme and also exists as an isolated protein with both enzyme and chaperone activity, protein disulphide isomerase (PDI). Different forms of P4H therefore differ in the α subunit, which shows tissue specificity, the type II enzyme being the major form in chondrocytes, osteoblasts and endothelial cells. While prolyl hydroxylation in the Y position is the norm, hydroxylation of prolines in the X-position is much less common, there being only one such residue per α chain in collagens I and II, in contrast to collagen IV which has several. Nevertheless, this modification performs a vital function, as recently shown by loss of prolyl 3-hydroxylase activity which leads to a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta (Cabral et al., 2007; Morello et al., 2006). It should be noted that for steric reasons, prolyl hydroxylation occurs only prior to assembly of the triple-helical procollagen molecule.

The mechanism of triple-helix stabilization by prolyl hydroxylation has generated much controversy but a general consensus is emerging that this is due to two effects: (i) increased hydrogen bonding through hydration networks and (ii) the electron-withdrawing (inductive) effect of the hydroxyl group that stabilizes the exo pucker of the (4R) stereoisomer in the Y position (Brodsky and Persikov, 2005). Denaturation or melting of the triple helix is a highly cooperative process, as measured by circular dichroism for example, which leads to a sharp transition with a characteristic melting temperature that is usually just a few degrees above body temperature. This is thought to be an evolutionary adaptation that permits the microunfolding that prevents misalignment and is necessary for interactions with other molecules, including degradative enzymes. It is known that the apparent melting temperature depends on the rate of heating. Indeed, recent observations have shown that collagen denaturation is an extremely slow process, and that the equilibrium melting temperature is actually below body temperature (Leikina et al., 2002). This means that collagen molecules only appear to be stable at body temperature, due to kinetic effects, and further stabilization comes from interactions with chaperones and supramolecular assembly (Section 2.4.2).

Lysyl hydroxylation is also an important post-translation modification in collagens. Hydroxylysine plays important roles in collagen cross-linking (see Chapter 4) and is a substrate for O-linked glycosylation (see below), thereby affecting fibril formation and other protein–protein interactions. Three forms of lysyl hydroxylase are known, LH1, LH2 and LH3, which like the prolyl hydroxylases reside

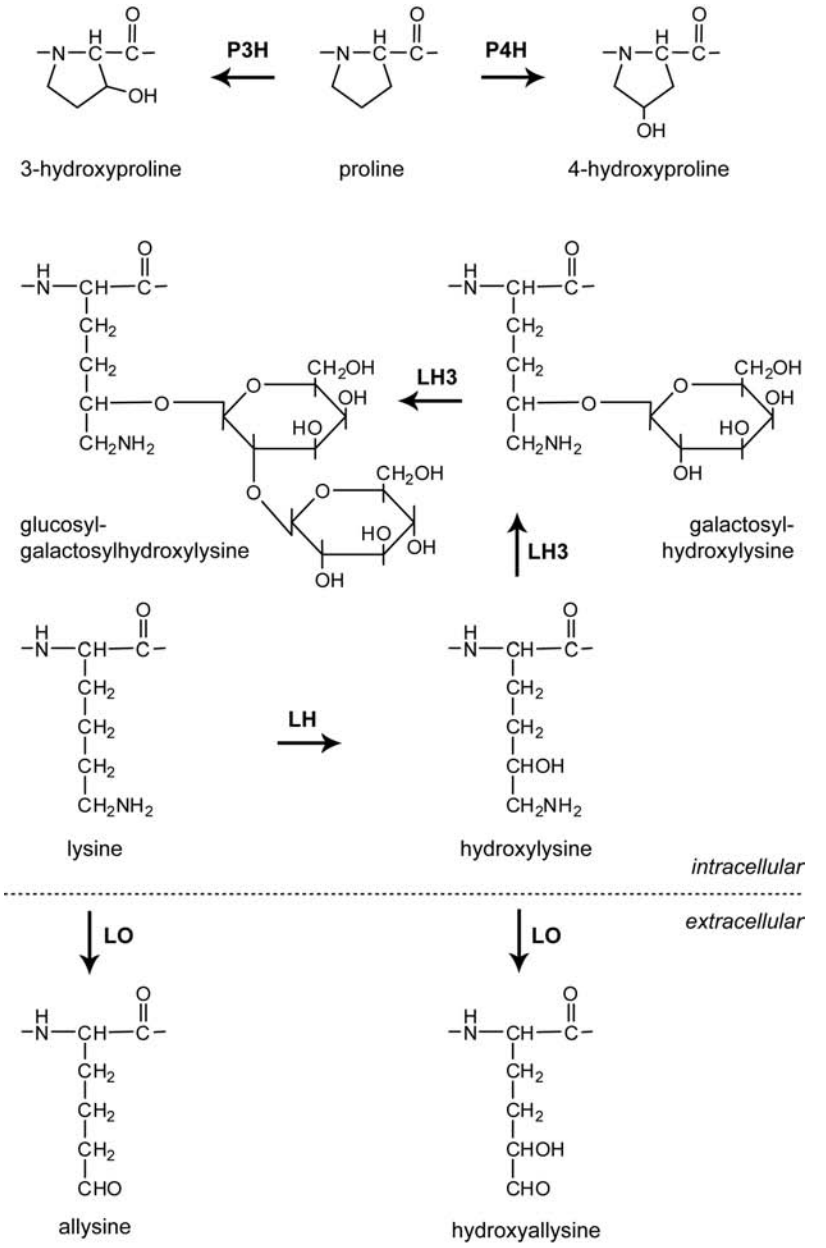

Fig. 2.5 Collagen-specific post-translational modifications of proline and lysine residues, and the enzymes involved. P3H = prolyl 3-hydroxylases; P4H = prolyl 4-hydroxylases; LH = lysyl hydroxylases; LO = lysyl oxidases

in the endoplasmic reticulum. Also like prolyl hydroxylases, these enzymes are 2-oxoglutarate dioxygenases requiring Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate for activity (Myllyla et al., 2007). The extent of lysyl hydroxylation varies between tissues, being relatively high, for example, in collagens IV and VI an also in embryonic tissues. Only lysines in the Y position of the Gly-X-Y triplet can be hydroxylated. LH2 has been shown to specifically hydroxylate lysine residues in the telopeptide regions. Telopeptide lysines and hydroxylysines are substrates for the cross-link initiating enzymes lysyl oxidases (Section 2.4.5), and the nature of the subsequent cross-linking is determined by the state of hydroxylation of the telopeptide lysines (see Chapter 4). This was recently demonstrated by a deficiency in LH2 which leads to Bruck syndrome, a form of osteogenesis imperfecta with joint contractures (van der Slot et al., 2003). Dysfunctional LH1, on the other hand, leads to Ehlers syndrome type VI, characterized by neonatal kyphoscoliosis, generalized joint laxity, skin fragility and severe muscle hypotonia at birth (Yeowell and Walker, 2000).

O-linked glycosylation of hydroxylysines is specific to collagenous proteins and involves the covalent attachment of galactose and then glucose from UDP carriers (Fig. 2.5). The enzyme(s) responsible was for a long time an enigma, but recent work has shown that the activity resides in the multifunctional enzyme LH3 which, in addition to being a lysyl hydroxylase, is also a galactosyltransferase (GT) and a glucosyltransferase (GGT). LH3 is therefore able to catalyze the three consecutive reactions required for the formation of glucosylgalactosyl-hydroxylysine. The molecule comprises a single polypeptide chain, with amino acids involved in the GT and GGT activities near the N-terminus and those involved in LH activity near the C-terminus. Identification of LH3 has allowed the functional significance of the LH and GGT activities to be investigated by genetic inactivation, either of the whole LH3 molecule or just the LH region (Ruotsalainen et al., 2006). The results showed that inactivation of LH had relatively little effect (probably due to partial compensation by other forms of LH), but inactivation of the entire molecule, including the GGT activity, was embryonic lethal, apparently due to disruption of basement membranes. Thus, O-linked glycosylation of collagens is essential for normal development.

2.4.2 Chain Association and Triple-Helix Formation

Following prolyl and lysyl hydroxylation and O-linked glycosylation in the rough endoplasmic reticulum, procollagen α chains then associate to form the procollagen molecule. In recent years, it has become clear that collagenous domains on their own have difficulty in assembling into trimers. Trimerization requires additional domains, adjacent to the $(Gly\text{-}X\text{-}Y)_n$ domains, to facilitate the assembly process. In the case of the fibrillar collagens, several lines of evidence point to a role for the procollagen C-propeptide domain in the initiation of trimerization. This involves two events: specific chain recognition to assure correct chain stoichiometry and formation of a stable nucleus to favor triple-helical folding. Control of chain stoichiometry is particularly important in cells that simultaneously produce several

collagen types, to result in the correct homotrimeric and heterotrimeric chain combinations. The amino acid sequences of the C-propeptide domains of the fibrillar procollagens are highly conserved, with the exception of a discontinuous sequence of 15 amino acids near the middle. It was shown some years ago (Lees et al., 1997), by swapping this region between different C-propeptide domains, that this sequence contains the information required for chain recognition. Regarding the formation of a stable nucleus, a further contribution was the recent identification of α-helical coiled-coil domains at the start of all known procollagen C-propeptide sequences (McAlinden et al., 2003). Alpha-helical coiled coils (not to be confused with collagen α chains!) are widely occurring oligomerization motifs (Parry, 2005), which result from the association of polypeptide chains containing the heptad repeating sequence $(abcdefg)_n$, where residues *a* and *d* are typically small hydrophobic amino acids and residues *e* and *g* are often charged. Formation of dimers, trimers or higher oligomers allows the α-helical chains to associate together in a rope-like structure (hence "coiled coils") thereby burying surface-exposed hydrophobic residues. In the procollagen molecule, four such heptad repeats are present at the start of the procollagen C-propeptides, separated from the collagen triple-helical region by the C-telopeptide. Similar juxtapositions of collagen triple-helical and coiled-coil domains are found in the collectins (Holmskov et al., 2003), where the coiled-coil domains have been shown to be essential for trimerization.

While triple-helix formation of fibrillar procollagens is initiated in the C-terminal region, with subsequent triple-helical folding in the C- to N-direction, this is not always the case in collagens. In transmembrane collagens and collagen-like proteins, for example, coiled-coils domains are found N-terminal to collagenous domains, where they help direct triple-helical folding in the N- to C-terminal direction (Snellman et al., 2007). Indeed, coiled-coil domains are almost ubiquitous in the collagen superfamily (McAlinden et al., 2003) and can occur C-terminal, N-terminal or between collagenous domains. Collagens without coiled-coil domains include collagen IV, C1q domain containing collagens and collagen-like proteins, and ficolins (which have C-terminal fibrinogen-like domains). Thus while coiled coils are likely to be the most widely used trimerization domains in collagens, alternative trimerization domains (collagen IV NC1, C1q, fibrinogen-like) may be used. One possible advantage of using trimerization domains to initiate triple helix assembly is to avoid misalignment of collagen chains, i.e., slippage of adjacent chains by multiples of the $(Gly\text{-}X\text{-}Y)_n$ repeat.

Following chain recognition and trimerization, a number of proteins come into play to guide the process of triple-helix formation during assembly of the procollagen molecule. These include prolyl 4-hydroxylase (P4H) which, in addition to its enzymatic function (Section 2.4.1), also acts as a chaperone protein by selectively binding to unfolded procollagen α chains, thereby preventing premature triple-helix formation (Walmsley et al., 1999). As previously mentioned, the β subunit of P4H is the same as another resident protein of the endoplasmic reticulum, protein disulphide isomerase (PDI), whose main function is to catalyze the formation and rearrangement of disulphide bonds (Ellgaard and Ruddock, 2005). In the case of procollagen assembly, PDI is involved in trimerization through the formation of

intra- and intermolecular disulphide bonds in the propeptide regions. PDI also binds selectively to unfolded chains thereby also performing a chaperone role (Bottomley et al., 2001). Additional chaperone proteins associated with quality control of intracellular collagen assembly are BiP/Grp78 and Grp94 (Koide and Nagata, 2005).

The peptide bond can exist in two conformations, *cis* or *trans*. For most amino acids, the *trans* conformation is strongly energetically favored. For prolyl- and hydroxyprolyl-containing peptide bonds, however, because of the cyclic nature of the side chains, the energy difference is smaller. Consequently, many of the proline and hydroxyprolines in newly synthesized procollagen α chains are in the *cis* conformation. Because only *trans* peptide bonds can be incorporated into the triple helix, *cis–trans* isomerization has been found to be a rate-limiting step in assembly of the procollagen molecule. This process can be accelerated by the enzymes known as prolyl *cis–trans* isomerases, or immunophilins, which includes the cyclophilins and the FK506 binding proteins (Barik, 2006). Like P4H and PDI, immunophilins also exhibit chaperone functions independent of their enzymatic activities.

In contrast to proteins that interact selectively with unfolded collagen chains, HSP47 is a collagen-specific chaperone that binds preferentially to folded triple helices (Koide and Nagata, 2005). HSP47 is a resident protein of the endoplasmic reticulum and is homologous to the serine protease inhibitors serpins. It is specifically expressed by cells that synthesize large amounts of collagen, where it performs a vital role as shown by the embryonic lethal phenotype of HSP47 knockout mice (Ishida et al., 2006). HSP47 binds to triple-helical collagen, preferentially to $(Gly\text{-}X\text{-}Y)_n$ repeats containing arginine at the Y position, and is thereby thought to stabilize the triple helix which is inherently unstable at physiological temperature (Section 2.4.1). By binding to the triple helix, it is also thought to prevent premature aggregation of procollagen molecules in the endoplasmic reticulum (Koide and Nagata, 2005). HSP47 dissociates from procollagen molecules as they transit the Golgi compartment (Section 2.4.3), probably as a result of a decrease in local pH.

2.4.3 Intracellular Transport and Secretion

From the endoplasmic reticulum, newly synthesized procollagen molecules proceed to the Golgi complex (or Golgi for short). Exit from the endoplasmic reticulum appears to occur via budding of carriers called ERGICs (Endoplasmic Reticulum to Golgi Intermediate Compartments) or VTCs (Vesicular Tubular Clusters) in a COPII-dependent process. COPII (coat protein complex II) is a multisubunit complex that consists of scaffold proteins, GTPases, GTPase activators and GDP/GTP exchange factors and has the ability to curve membranes, and thereby form vesicles (Fromme and Schekman, 2005). The transport of such large molecules as procollagen (300 nm in length) is difficult to reconcile with the conventional view of ER to Golgi transport involving 60–70 nm COPII-coated vesicles. Recent work (Mironov et al., 2003) has lead to the view that for such large cargos, while COPII is essential for budding, it is not itself incorporated into the carriers.

The state of aggregation of procollagen molecules during transport through the Golgi, the *trans*-Golgi network and subsequent secretion into the extracellular matrix has been a subject of investigation for over 30 years. Early studies (see Canty and Kadler, 2005) revealed bundles of procollagen molecules aligned in register (also known as SLS crystallites) in secretory vesicles, which were subsequently found to be secreted as bundles into the extracellular matrix (Hulmes et al., 1983). More recent studies have shown that procollagen molecules are found in this form throughout their transit through the Golgi, which occurs without leaving the lumen of the cisternae, thus supporting the "cisternal maturation" model (Bonfanti et al., 1998). It is not clear what makes procollagen molecules associate in this way within vacuoles, rather than in D-staggered array as in fibrils. With purified procollagen I, the presence of the propeptides impedes, but does not prevent, D-staggered assembly at the high concentrations likely to be found during intracellular transport (Mould et al., 1990). It is possible that interactions with other secreted components might favor non-staggered aggregation into bundles.

2.4.4 Procollagen Processing

Enzymatic removal of procollagen propeptides, or procollagen processing, is carried out by metalloproteinases belonging to the ADAMTS (Porter et al., 2005) and BMP1/Tolloid-like families (Hopkins et al., 2007), as well as the furin-like proprotein convertases (Seidah and Prat, 2005).

The ADAMTS (for "a disintegrin and a metalloproteinase with thrombospondin repeats") family has 20 members, of which ADAMTS-2, -3 and -14 are implicated in N-terminal processing of fibrillar procollagens (Porter et al., 2005). These enzymes are large multidomain proteins, including a prodomain (cleaved by furin), a reprolysin-like zinc metalloproteinase domain, disintegrin, cysteine-rich and spacer domains and multiple thrombospondin type-I domains. Defective procollagen N-proteinase activity is associated with a hereditary disorder originally found in cattle and sheep, called dermatosparaxis, and later found in humans, called Ehlers–Danlos syndrome type VIIC, characterized by extreme skin fragility and laxity (Section 2.5.2). All three enzymes cleave the amino-propeptides from procollagen I, while ADAMTS-2 cleaves procollagen III, and ADAMTS-3 cleaves procollagen II. They also have somewhat different tissue specificities, such as ADAMTS-2 in skin, lungs and aorta and ADAMTS-3 in cartilage and bone (Le Goff et al., 2006). It is of interest to note that ADAMTS-2 interacts with the FACIT collagen XIV, suggesting a possible role for fibril-associated collagens in procollagen processing (Colige et al., 1995).

BMP1/Tolloid-like proteinases (Hopkins et al., 2007) are largely responsible for the C-terminal processing of fibrillar procollagens. Originally identified as the products of genes involved in the induction of bone formation (BMP1, for bone morphogenetic protein-1) and in dorso-ventral patterning during embryonic development (Tolloid, originally identified in *Drosophila*), this family currently includes four members in mammals: BMP1, mTLD (mammalian Tolloid), mTLL1 (mTLD-like 1) and mTLL2 (mTLD-like 2). These are zinc-dependent metalloproteinases

with multiple domains, including a proregion (cleaved by furin), a catalytic astacin domain and a number of CUB (complement-uegf-BMP1) and EGF (epidermal growth factor)-like domains. While mTLD, mTLL1 and mTLL2 have five CUB domains and two EGF domains, BMP1 (a product of alternative splicing, encoded by the same gene as mTLD) lacks the two C-terminal CUB domains and the second EGF domain. All four enzymes can cleave the C-propeptides from the major fibrillar procollagens (types I–III), though BMP1 seems to be the most active in vitro. In the case of the minor fibrillar procollagens V and XI, C-terminal processing is complex, with furin-like cleavage appearing to play a major role in the proα1(V), proα1(XI) and proα2(XI) chains and BMP1/Tolloid-like proteinases cleaving the proα2(V) chain. Furthermore, unlike the major fibrillar procollagens, BMP1/Tolloid proteinases cleave within the large N-terminal regions of the proα1(V), proα1(XI) and proα2(XI) chains, between the TSPN and variable domains (Fig. 2.2).

In addition to the fibrillar collagens, BMP1/Tolloid-like proteinases cleave a large number of extracellular substrates (Hopkins et al., 2007), including the non-fibrillar collagen VII, the transmembrane collagen gliomedin, various proteoglycans, laminin 5 and prolysyl oxidases (Section 2.4.5). They are also involved in the activation of a large number of growth factors, including transforming growth factor-β, a key regulator of extracellular matrix deposition and turnover. The activity of BMP1/Tolloid-like proteinases is regulated by additional proteins of the extracellular matrix, procollagen C-proteinase enhancers (PCPE)-1 and -2 (Hopkins et al., 2007). Like BMP1/Tolloid-like proteinases, PCPEs also contain CUB domains, but are devoid of intrinsic proteolytic activity. Instead, they can enhance the activity of BMP1/Tolloid-like proteinases by up to 20-fold. It has recently been shown that PCPE-1 specifically enhances the action of BMP1 during C-terminal processing of the major fibrillar procollagens (Moali et al., 2005).

For a long time, it was thought that procollagen processing is an extracellular event. Recent studies, however, support the idea that processing and fibril assembly (Section 2.5.5) can begin in Golgi to plasma membrane carriers (Canty et al., 2004). Using pulse-chase and differential extraction procedures, these authors demonstrated the presence of procollagen processing intermediates, without N- or C-propeptides, in intracellular compartments, consistent with the observation that furin activation of proBMP1 and proADAMTS2 can also occur prior to secretion (Leighton and Kadler, 2003; Wang et al., 2003). Procollagen processing probably also occurs in close association with the plasma membrane. For example, targeted deletion of the extracellular matrix protein SPARC results in increased procollagen processing probably as a result of increased retention by cell-surface receptors (Rentz et al., 2007).

2.4.5 Covalent Cross-Linking

The final step in the biosynthesis of collagens is the introduction of covalent cross-links to stabilize the different forms of supramolecular assembly. In the case of the fibrillar collagens, cross-linking is for the most part initiated by members of the lysyl oxidase family of copper-dependent amine oxidases (Lucero and Kagan, 2006; Molnar et al., 2003). This includes five members: lysyl oxidase

(LOX), lysyl oxidase-like (LOXL), LOXL2, LOXL3 and LOXL4. All five are highly homologous in their C-terminal region, which contains the catalytic domain and a cytokine receptor-like domain. In the case of proLOX, the catalytic region is preceded by a proregion that is cleaved by members of the BMP1/Tolloid-like family. Since cleavage of the proLOX propeptide activates LOX, and cleavage of the procollagen C-propeptides trigger fibril formation (Section 2.5.2), BMP1/Tolloid-like proteinases regulate both collagen cross-linking and assembly. The precursor form of LOXL contains a unique proline-rich region and is also processed by BMP1 at different sites, leading to multiple forms of the mature enzyme (Borel et al., 2001). The remaining lysyl oxidase-like enzymes, LOXL2, LOXL3 and LOXL4, form a distinct subgroup that is characterized by the presence of four scavenger receptor domains, N-terminal to the catalytic domain. It is not known whether these enzymes are proteolytically processed.

The effect of lysyl oxidase activity on fibrillar collagens is to convert lysine or hydroxylysine residues in the N- and C-terminal telopeptide regions to corresponding peptidyl aldehydes (Fig. 2.5). Once formed, aldehydes spontaneously condense with other aldehydes or unreacted lysines and hydroxylysines to form a variety of intra- and intermolecular covalent cross-links (see Chapter 4). LOX activity is strongly increased by assembly of collagen into fibrils, presumably as a result of LOX binding to highly conserved sequences that become juxtaposed to the telopeptide regions when adjacent molecules are mutually staggered by 4×67 nm ($4D$) (Kuhn, 1987). LOXL is also active on a fibrillar collagen substrate (Borel et al., 2001). Non-fibrillar collagens taking part in lysyl oxidase initiated cross-linking include types IV, X and IX (Ricard-Blum et al., 2000). Little is known about the substrate specificity of the other lysyl oxidases. There is evidence for an association of LOXL2 with basement membrane collagen IV, and indeed LOXL2 also seems to act intracellularly through interactions with the transcription factor Snail, a regulator of the epithelial–mesenchymal transition (Peinado et al., 2005). LOX also seems to have an intracellular role, as a tumor suppressor, which appears to be mediated by its propeptide region (see Lucero and Kagan, 2006). Lysyl oxidase activity can be inhibited by lathyritic agents, the mostly widely used being β-aminopropionitrile (BAPN).

Another enzyme involved in collagen cross-linking in vivo is tissue transglutaminase (TG2), a multifunctional enzyme with both cross-linking (extracellular) and GTPase (intracellular) activities (Verderio et al., 2005). Cross-linking involves the formation of an isopeptide bond between the γ-carboxyamide group of specific peptidyl glutamine residues and ε-amino group of peptidyl lysines. Several collagens are cross-linked by TG2, including collagens II, III, V, VII and XI (Grenard et al., 2001).

2.5 Assembly of Fibrillar Collagens

2.5.1 Reconstitution of Fibrils In Vitro

The formation of collagen fibrils in vitro has been a subject of extensive research for over 50 years (see Kadler et al., 1996). Initial observations were on the reconstitution

of fibrils from cold solutions of collagen in dilute acid. The major fibrillar collagens are soluble at low pH, for example in dilute acetic acid. When the pH is adjusted to around neutral, and the temperature raised to around physiological, fibril formation occurs spontaneously resulting in banded fibrils. The structure of the resulting fibrils is influenced by several parameters, including buffer composition, intactness of the N- and C-telopeptides, the presence of other types of collagen, the presence of macromolecules other than collagens and the order of the initiating procedure.

The influence of buffer composition on fibril formation has been studied by several authors, notably by Williams et al. (1978) who recommended diluting cold solutions of collagen I at concentrations up to 1 mg/ml in 5 mM acetic acid with an equal volume of optimum buffer consisting of 60 mM TES (*N*-[tris (hydroxymethyl)methyl-2-amino]ethanesulphonic acid), 60 mM sodium phosphate, 270 mM NaCl, pH 7.3, and then raising the temperature to up to 30°C. In these conditions, fibril morphology is optimized, as judged by the presence of compact fibrils with a clear *D*-periodic banding pattern. The presence of phosphate is important for producing well-banded fibrils. Another important parameter is the temperature at which fibril formation is carried out, fibrils formed at lower temperatures (20°C) having larger diameters than those formed at higher temperatures (34°C). Finally, the initiation procedure also has an effect, the so-called "warm start" procedure (warming first and then buffer neutralization) giving better formed fibrils (i.e., less non-banded filaments) than with the "cold start" procedure (neutralization and then warming) (Holmes et al., 1986).

Intactness of the N- and C-telopeptides has a marked effect on fibril formation. Commercially available fibrillar collagens are of two types, either acid soluble or pepsin soluble. In acid-soluble collagen, the telopeptides are usually intact, and these help both to initiate fibril formation and to produce long cylindrical fibrils. Pepsin-soluble collagen is produced by preparing tissue extracts with pepsin, which digests most protein structures except for the collagen triple helix. As a consequence, the non-helical telopeptides are mostly removed by this procedure, making self-assembly more difficult. Selective removal of the N-telopeptides results in so-called *D*-periodic symmetric (DPS) fibrils, in which molecules assemble in an antiparallel manner throughout the fibril length, while partial loss of the C-telopeptides results in relatively short cigar-shaped *D*-periodic cigar-shaped tactoids (Kadler et al., 1996). The importance of telopeptide interactions has been further demonstrated by Prockop and Fertala (1998) who showed that exogenous peptides corresponding to N- or C-telopeptides could inhibit fibril formation.

Fibril formation is an entropy-driven process, in which self-assembly results in burying surface-exposed hydrophobic residues within the fibril, thereby increasing entropy in the solvent. Assembly proceeds by a nucleation and growth mechanism, as monitored typically by turbidimetry at a wavelength of around 300 nm. Light scattering, or turbidity, is proportional to both the concentration of collagen and the mass of the assembled structures. Turbidity curves classically show three phases: an initial lag phase with no change in turbidity, a rapid growth phase and finally a plateau region. During the lag phase, small numbers of collagen molecules associate

to form metastable nuclei, upon which further molecules accrete during the growth phase. The initial interaction seems to be a pair of 4*D* staggered molecules with a short N- and C-terminal overlap, mediated by the N- and C-telopeptides (Ward et al., 1986). Growth in fibril length and width then occurs via longitudinal and lateral interactions.

While fibrils formed in vitro appear at first glance to be approximately cylindrical, detailed observation of unstained fibrils by scanning transmission electron microscopy (STEM), which can measure the mass per unit length of individual fibrils, shows that they have pointed tips (see Kadler et al., 1996). When reconstituted from acid-soluble collagen, fibrils have relatively sharp N-terminal tips and blunt C-terminal tips.

2.5.2 Fibril Formation De Novo from Procollagen

As a more physiological alternative to the reconstitution of collagen fibrils from acid solutions, fibrils can also be formed in vitro from procollagen, in physiological buffers, in the presence of procollagen N-proteinase (ADAMTS-2) and procollagen C-proteinase (BMP-1). The propeptides impede assembly thereby increasing solubility, largely due to the presence of the C-propeptide region (see below). Because of this, it is convenient to begin with a partially processed form of procollagen with the N-propeptides removed, called pC-collagen, and trigger fibril formation by the addition of BMP-1. When this is done, the initially formed (unipolar) fibrils also have pointed tips, with a sharp N-terminus and a blunt C-terminus. Unlike fibrils reconstituted from acid-soluble collagen however, these de novo fibrils then go on to acquire pointed tips at both ends. This is due to the formation of bipolar fibrils in which molecules have their N-termini pointing toward one of the tips, with an abrupt change in molecular orientation somewhere in the middle. It was subsequently shown that such bipolar fibrils are relatively common in vivo, indicating that de novo fibril formation in vitro can reproduce at least some of the features of the in vivo system (Kadler et al., 1996). Also of note is that the shape of the tips is determined by the rate of C-terminal procollagen processing, fast processing generating blunter tips than slow processing (Holmes et al., 1996).

STEM analysis of de novo fibrils formed in vitro (Holmes et al., 1992) shows that the axial mass distributions for both sharp and blunt tips are linear, i.e., the number of molecules added per *D* period is constant (being greater for blunt tips). This is an important observation, which is at variance with simple conical shapes for the tips which would be expected to give quadratic axial mass distributions. The linear increase shows that the tips are paraboloidal in shape. Theoretical simulations have shown that this is consistent with fibril growth being determined by kinetic mechanisms, such as diffusion-limited aggregation (Parkinson et al., 1994).

Procollagen processing has a dramatic influence on fibril formation. This has been shown by studying procollagen I assembly in vitro, with or without the N- or C-propeptides. When the C-propeptides are removed from pC-collagen, fully

processed collagen molecules readily assemble into fibrils at concentrations greater than a so-called critical concentration of approximately 1 μg/ml. When only the C-propeptide is removed from procollagen, the processing intermediate pN-collagen begins to assemble at concentrations approximately 100-fold greater, but unlike fully processed collagen, pure pN-collagen forms *D*-periodic sheet-like structures. Such sheets can be several microns wide, but their thickness is limited to 8 nm (Hulmes et al., 1989). It seems that persistence of the N-propeptide region prevents lateral aggregation in 1*D*, probably by steric hindrance as a result of the N-propeptides folding back against the main triple-helical region. When both the N- and C-propeptides are intact, procollagen molecules have also been found to assemble into similar sheet-like structures, but only at concentrations greater than 1 mg/ml (Mould et al., 1990). Thus, the presence of the C-propeptides increases solubility by approximately 1000-fold.

As mentioned earlier, defective N-proteinase activity is associated with a hereditary disorder originally found in cattle and sheep (dermatosparaxis) and later found in humans (Ehlers–Danlos VIIc) characterized by extreme skin fragility. When examined by electron microscopy, collagen fibrils in the skin are no longer approximately cylindrical as in normal tissue, but instead are thin and branched. By treating mixtures of procollagen and pC-collagen with C-proteinase, resulting in mixtures of pN-collagen and collagen, respectively, it was possible to reproduce exactly the shape of the dermatosporactic fibrils in vitro, with a pN-collagen:collagen ratio of about 2:1 (Hulmes et al., 1989). In contrast, when the pN-collagen:collagen ratio was less than 1:5, fibrils were approximately circular in cross-section. These observations are consistent with the idea that persistence of the N-propeptide distorts fibril shape by increasing the surface area to volume ratio, in order for the N-propeptides to be surface located and not buried within the fibril. Another way of increasing the surface area to volume ratio is by making fibrils with small diameters. Thus, as suggested by Fleischmajer et al. (1990), it is likely that limited persistence of the N-propeptide has a role to play in fibril diameter regulation.

2.5.3 Heterotypic Fibril Assembly

Another important factor that influences fibril assembly is the formation of copolymers consisting of different collagen types. Such heterotypic fibrils are the norm in most connective tissues, for example, collagens I, III and V in skin, collagens I and V in cornea and collagens II and XI in cartilage. N-terminal processing of procollagen III is relatively slow, thus heterotypic collagen I/III fibrils in skin contain significant amounts of pN-collagen III, which have been shown to limit fibril diameter in vitro (Romanic et al., 1991), again probably due to steric effects. Similar reasoning applies to collagen I/V fibrils in cornea and collagen II/XI fibrils in cartilage. N-terminal processing of collagens V and XI is at best only partial, leaving large surface located N-terminal extensions which impede fibril growth, with the rest of the molecules being buried within the fibril interior (Fig. 2.6). The effects of

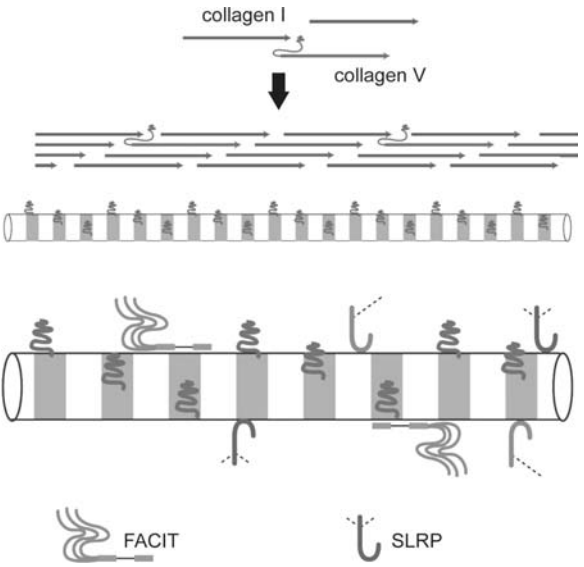

Fig. 2.6 The complex landscape of the corneal collagen fibril surface, showing N-terminal extensions of collagen V, surface bound FACITs and small leucine-rich proteoglycans (from (Birk and Bruckner, 2005), with permission)

collagens V and XI have also been demonstrated in vitro (Birk et al., 1990; Blaschke et al., 2000). In vivo, targeted down-regulation of the expression of collagens III (Liu et al., 1997) or V (Wenstrup et al., 2006), or a natural mutation that prevents collagen XI expression (Li et al., 1995), also results in the formation of large diameter fibrils, consistent with the steric blocking mechanism of fibril growth.

Steric blocking by N-terminal extensions is not the only mechanism by which the minor fibrillar collagens might limit heterotypic fibril diameter. Another mechanism is through altering the rate of fibril nucleation. This has been shown most dramatically in the case of collagen V, where complete deficiency of the proα1(V) chain (which results in death at embryonic day 10) leads to the absence of procollagen V molecules and an almost total lack of collagen fibrils in the dermis, despite their being normal levels of collagen I (Wenstrup et al., 2004). The minor fibrillar collagens therefore seem to act as initiators of fibril formation, during the early nucleation stage. Large amounts of collagens V and XI would therefore lead to greater numbers of heterotypic fibrils (with collagens I and II, respectively). For a given amount of collagen, the presence of a greater number of fibrils would result in the average fibril diameter being smaller.

2.5.4 Interactions with Proteoglycans and Other Components of the Extracellular Matrix

A large body of evidence shows that interactions with the so-called small leucine-rich proteoglycans (SLRPs) have marked effects on collagen assembly (Iozzo, 1999;

McEwan et al., 2006). At present, 15 different types of SLRPs are known, each one consisting of a single polypeptide chain (or protein core) with an N-terminal cysteine-rich cap, followed by several leucine-rich repeat motifs and in most cases a C-terminal disulphide bonded cap. In proteoglycan form, a small number of glycosaminoglycan (GAG) chains, either chondroitin/dermatan sulphate or keratan sulphate, are covalently attached to the protein core. The leucine-rich repeats (each one, 21–27 amino acid residues in length) give SLRPs on overall banana shape, as exemplified by the crystal structure of decorin (Scott et al., 2004). SLRPs known to affect collagen assembly include decorin, fibromodulin, lumican, biglycan, keratocan and osteoglycin/mimecan. Early in vitro studies showed that decorin and fibromodulin can markedly interfere with collagen fibril formation, resulting in delayed assembly and usually reduced fibril diameters. Removal of the GAG chains has little effect, showing that this is largely a property of the protein core. Lumican has similar effects in vitro, while biglycan appears to bind to collagen without affecting fibril assembly. In the case of osteoglycin/mimecan, it has recently been shown that processing by BMP1 potentiates the inhibiting effect on fibril assembly (Ge et al., 2004).

In vivo, further insights have been gained by targeted disruption (knockout) of the corresponding genes. The decorin knockout is characterized by abnormal skin fragility associated with collagen fibrils having a much broader range of diameters than in normal skin, with non-uniform axial mass distributions and irregular contours resulting from fibril fusion (Danielson et al., 1997). This is consistent with a role for decorin molecules in coating the fibril surface thereby keeping them apart. In the case of the lumican knockout (Chakravarti et al., 1998), fibrils in cornea which are normally thin ($\sim$30 nm) and uniform in diameter show a similar loss of diameter control and organization resulting in opaque corneas (Fig. 2.7; see Chapter 13). The fibromodulin knockout results in tendon fibrils with reduced diameters, though interpretation of these data is complicated by a compensatory up-regulation of lumican expression (Svensson et al., 1999). Finally, the biglycan knockout leads to an osteoporosis-like defect in bone associated with large irregular fibrils (Xu et al., 1998). It should be noted that SLRPs have multiple biological effects independent of their effects on collagen assembly, so interpretation of the in vivo data is complex.

The interaction of decorin and collagen I has been studied in some detail (Graham et al., 2000). Unlike fibrils formed in vitro, fibrils isolated from tissues show diameter limitation for most of their length, as evidenced by a constant axial mass distribution. This is thought to be due, at least in part, to interactions with decorin that coats the fibrils preventing further accretion. Interestingly, the amount of decorin appears to be less at the fibril tips. This is to be expected if the stoichiometry of binding of decorin to collagen is constant, since the number of molecules exposed on the surface, as a proportion of the total number of molecules in a fibril cross-section, increases as the tip diameter becomes smaller. Thus, tips are coated with a relatively small amount of decorin that will therefore no longer prevent fibril growth, until a limiting diameter is reached. This provides a mechanism for fibril to grow in length

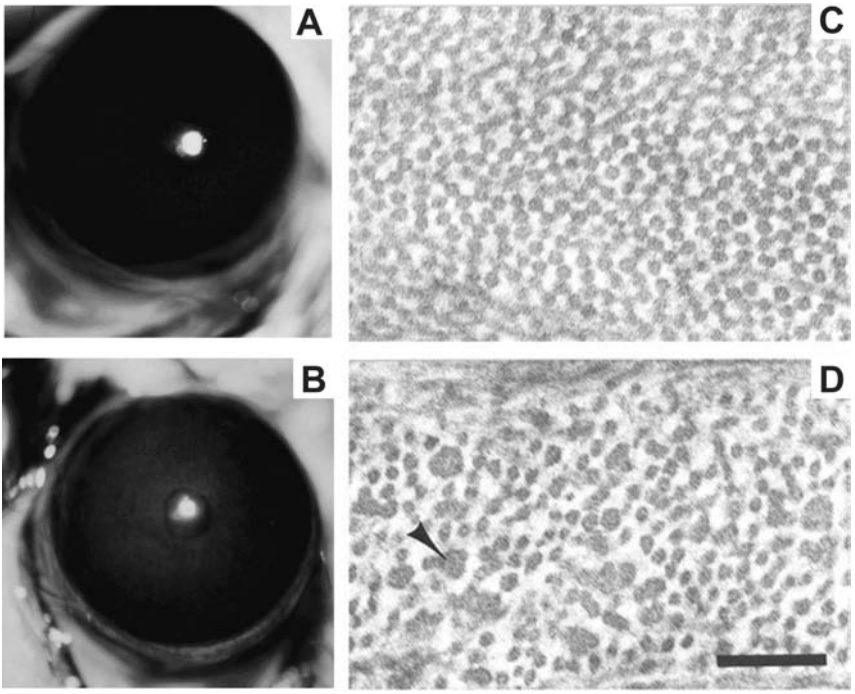

Fig. 2.7 Targeted disruption of lumican expression leads to corneal opacity as a result of loss of collagen fibril diameter control. **A** normal cornea and **C** corresponding electron micrograph of collagen fibrils (in cross-section) in the corneal stroma. **B** and **D** Lumican deficient cornea (from (Chakravarti et al., 1998), with permission)

whilst maintaining a constant diameter. It also permits early fibrils to fuse end to end, with the molecules pointing in the same direction, or in an antiparallel manner by C–C fusion, which creates bipolar fibrils with two N-terminal tips.

While SLRPs have usually been found to diminish the rate of fibril formation, this is not always the case. Recombinant lumican, for example, has been found to accelerate fibril formation in vitro (Neame et al., 2000), as have other extracellular matrix molecules, such as dermatopontin (MacBeath et al., 1993), tenascin-X (Minamitani et al., 2004), perlecan (Kvist et al., 2006), hevin (Sullivan et al., 2006) and cartilage oligomeric matrix protein (Halasz et al., 2007). It is likely that such interactions stabilize the initial nuclei that form during the lag phase prior to subsequent fibril growth. Fibronectin has also been found to accelerate collagen assembly in vitro (Speranza et al., 1987). Consistent with this observation, targeted disruption of fibronectin expression (Velling et al., 2002) or blocking fibronectin assembly with anti-$\alpha 5\beta 1$ integrin antibodies (Li et al., 2003) prevents collagen assembly in cell culture.

2.5.5 Cell Interactions and Long-Range Order

It is now well established that in vivo, procollagen processing, secretion and fibril formation are intimately associated with membrane-enclosed compartments. Prior to secretion (see Section 2.4.4), there is evidence that both procollagen processing and early fibril formation can occur in Golgi to plasma membrane carriers. Subsequently, these carriers appear to fuse with the plasma membrane forming cell-surface invaginations (Canty et al., 2004). Such structures were first observed by Birk, Trelstad and colleagues (Birk et al., 1989; Birk and Trelstad, 1986)(Fig. 2.8). Originally considered to be recesses within the plasma membrane, more recent studies using three-dimensional reconstruction techniques have shown early fibrils within membranous extrusions projecting from the cell surface, called fibropositors (Canty et al., 2004) (Fig. 2.8). It is thought that such structures deliver newly formed fibrils to the extracellular matrix, where they join existing fibril bundles within channels formed between neighboring cells. These structures are readily observed in embryonic tendon when ECM synthesis is high, though not in adult tissues. Tendon is a tissue where fibril alignment is particularly important in order to withstand tensile stress. Recent observations have shown that fibroblasts within chick embryo fibroblasts are connected by cadherin-mediated cell–cell junctions, which appear to define extracellular spaces that facilitate fibril alignment (Richardson et al., 2007).

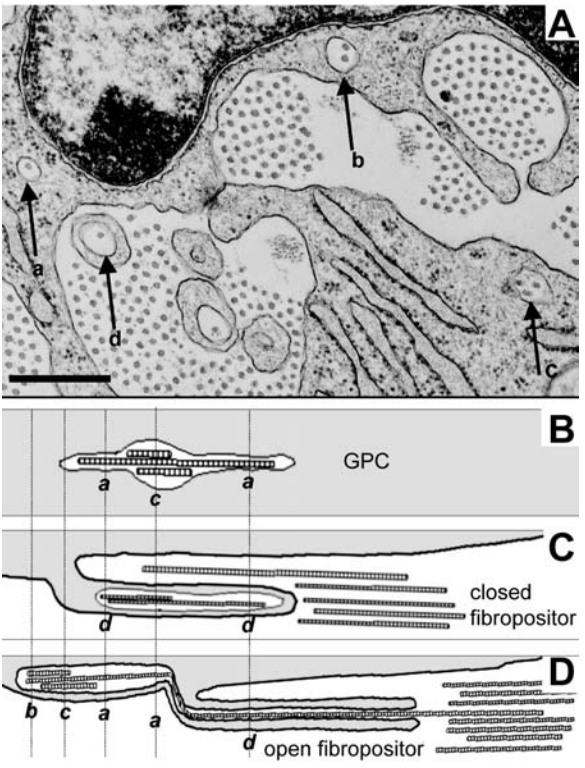

Fig. 2.8 Association of newly formed fibrils with cell membranes. **A** Electron micrograph of a transverse section through embryonic mouse tail showing bundles of extracellular collagen fibrils between adjacent cells, as well as membrane-bounded fibrils (or small groups of fibrils) within the cytoplasm (*a*, *b*, *c*) or in plasma membrane (PM) extensions called fibropositors (*d*). Scale bar 500 nm. **B**–**D** Schematic representations of collagen fibrils in a Golgi-to-PM transport compartment (**B**) , as well as in closed (**C**) and open (**D**) fibropositors, with individual fibrils or small groups of fibrils indicated as in (**A**). From Canty and Kadler (2005), with permission

Early fibrils in 12–16-day chick embryo tendons have diameters of about 30 nm and lengths in the range 20–30 μm (Birk et al., 1995). Thereafter, there is a sharp increase in length and an increase in fibril diameter. There is evidence that subsequent fibril growth occurs by a process of fibril fusion involving tapered fibril tips. As discussed previously (Section 2.5.2), fibrils in vivo occur in two forms, either N–C unipolar or N–N bipolar, with a change in orientation somewhere in the middle. Since N-to-N fusion is never observed, subsequent fusion of bipolar fibrils cannot occur (Kadler et al., 1996). This limits possible modes of fibril fusion to unipolar–unipolar or unipolar–bipolar, thereby providing a mechanism for controlling fusion within fibril bundles. Another factor controlling fibril fusion is the presence of small leucine-rich proteoglycans, notably decorin and lumican. For example, the rapid rise in fibril length seen in chick embryo tendon at 17 days is correlated with a sharp drop in decorin content (Birk et al., 1995).

While fibril diameters in tendon increase with age and can reach up to 500 nm in rat tail tendon, fibril diameters in corneal stroma remain at about 30 nm throughout life. This is essential for the maintenance of optical transparency (see Chapter 13). Furthermore, unlike the aligned fibril bundles seen in tendon, fibrils in cornea are arranged in lamellae, within each of which fibrils are aligned but between which there is an abrupt change in orientation. In chick corneal, the angle is almost 90° giving rise to a twist in orientation between pairs of lamella throughout the depth of the stroma. Such changes in orientation between lamellae are also encountered in other connective tissues, notably compact bone and intervertebral disc, and are reminiscent of cholesteric liquid crystals (Giraud-Guille, 1996). This similarity raises the question of whether liquid crystalline organization might have a role to play in connective tissue morphogenesis. Clearly, once formed, collagen fibrils are too big to reorient to any major degree in tissues. Collagen molecules, on the other hand, can show liquid crystalline organization in acetic acid solution, resulting in cholesteric phases (Giraud-Guille, 1992). More recently, soluble procollagen molecules have been found to form pre-cholesteric phases in a physiological buffer, thus raising the possibility that long-range order might arise prior to fibril formation (Martin et al., 2000). Whether such liquid crystalline ordering of procollagen occurs in vivo, or whether cells control long range order, is a matter of debate. The observation that orthogonal arrays of collagen fibrils form in the chick primary corneal stroma from procollagen molecules that must traverse a basement membrane after being secreted by an epithelial cell sheet (Linsenmayer et al., 1998), is an argument in favor of such a self-assembly process. Indeed, formation of fibropositors may be a consequence of collagen assembly, rather than vice versa (Birk and Bruckner, 2005).

2.6 Assembly of Collagen-Like Peptides

This chapter would not be complete without a discussion of a new and rapidly growing area of research, the use of self-assembling collagen-like peptides to make novel scaffolds for applications in tissue engineering and regenerative medicine. It

is now well established that the extracellular matrix provides signals that control many cell activities such as differentiation, adhesion, migration, proliferation and apoptosis. These signals are orchestrated by the three-dimensional organization of the matrix, which provides both specific interactions and also structural clues that are only beginning to be understood. The possibility of synthesizing collagen-like peptides with the ability to self-assemble into nanofibrous structures permits the construction of well-characterized scaffolds for the study of cell–matrix interactions and for applications in tissue repair (Koide, 2005).

Some of the pioneering works in this area was carried out by Fields and his colleagues, who developed "peptide-amphiphiles" consisting of peptides with a collagen-like sequence attached to a long-chain dialkyl ester lipid tail (Fields et al., 1998). The collagen sequence conforms to the host–guest peptide model (Brodsky and Persikov, 2005), with two $(Gly\text{-}Pro\text{-}Hyp)_4$ sequences (for triple helix stability) straddling a real collagen sequence (in this case, from the collagen $\alpha 1$(IV) chain). These peptide-amphiphiles form stable triple helices and assemble into monolayers, due to the presence of the lipid tail. A related approach was that of Kaplan and colleagues (Martin et al., 2003), who synthesized "collagen-triblock" peptides consisting of the sequence $(Gly\text{-}X\text{-}Hyp\text{-}Gly\text{-}Pro\text{-}Hyp)_6$, where X = Pro, Ala, Val or Ser, straddled by a pentaglutamate sequence at either end. All four peptides formed stable triple helices. Furthermore, at high concentrations ($> 20\,mg/ml$), all but the Ser-containing triple helices formed liquid crystals. The type of liquid crystal depended on the amino acid sequence: nematic when X = Val and cholesteric when X = Ala or Pro. In nematic phases, neighboring molecules are aligned in parallel, while in cholesteric phases, they are related by a twist. This shows that single amino acid substitutions can have important consequences for supramolecular assembly.

In order to form high molecular weight polymers of collagen-like peptides, two approaches have been used. In the first method, host–guest peptides containing an integrin-binding sequence were constructed with an N-terminal cysteine and a C-terminal thioester (Paramonov et al., 2005). When incubated in an aqueous solvent at neutral pH, these peptides undergo triple-helix formation and also spontaneous chemical ligation to form polymers 10–20 nm in diameter and several microns in length. The second approach is inspired by the notion of "sticky ends" in DNA ligation, and it involves the synthesis of a molecule with a triple-helical core, with two of the chains having N-terminal $(Pro\text{-}Y\text{-}Gly)_5$ extensions and the third a C-terminal $(Pro\text{-}Y\text{-}Gly)_5$ extension (Kotch and Raines, 2006) or a variant thereof (Koide et al., 2005) (Fig. 2.9). Molecules then associate by self-complementary interactions to form near-continuous triple helices that are hundreds of nanometer in length. Interestingly, polymers are longer when Y is Hyp rather than Pro, suggesting that hydroxyproline favors self-assembly. This has recently been verified using classical $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptides that assemble into branched filamentous triple-helical networks in a temperature and concentration-dependent manner, unlike $(Pro\text{-}Pro\text{-}Gly)_{10}$ triple helices that do not (Kar et al., 2006).

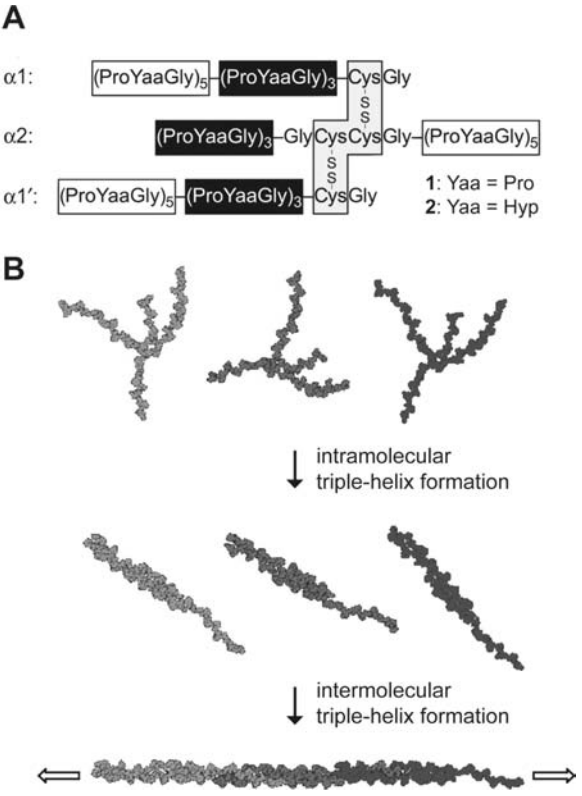

Fig. 2.9 Collagen-like peptides. **A** Amino acid sequences showing the central triple-helical zone and the overhanging "sticky ends". **B** Self-assembly of molecules shown in (**A**) each identified by a different color. From Kotch and Raines (2006), with permission

2.7 Conclusions

Less than 40 years ago, only one type of collagen was known. Since then, the number of different types has increased dramatically, in addition to the now large number of collagen-like proteins. The collagen triple-helical motif has turned out to remarkably be versatile in function, leading to the assembly of diverse supramolecular structures. The biosynthesis of collagens is a complex process, involving several post-translational modifications, the functions of which continue to be elucidated. While studies on collagen assembly have gradually switched from in vitro to in vivo, both approaches give complementary information on the molecular mechanisms involved. Armed with this vast amount of data, the time is ripe for developing new approaches to the use of collagens in tissue regeneration and repair.

References

Baldock C, Sherratt MJ, Shuttleworth CA, Kielty CM (2003) The supramolecular organization of collagen VI microfibrils. J Mol Biol 330: 297–307.

Barik S (2006) Immunophilins: for the love of proteins. Cell Mol Life Sci 63: 2889–2900.

Beck K, Brodsky B (1998) Supercoiled protein motifs: The collagen triple-helix and the alpha-helical coiled coil. J Struct Biol 122: 17–29.

Birk DE, Bruckner P (2005) Collagen superstuctures. Top Curr Chem 247: 185–205.

Birk DE, Trelstad RL (1986) Extracellular compartments in tendon morphogenesis: Collagen fibril, bundle, and macroaggregate formation. J Cell Biol 103: 231–240.

Birk DE, Fitch JM, Babiarz JP, Doane KJ, Linsenmayer TF (1990) Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. J Cell Sci 95 (Pt 4): 649–657.

Birk DE, Nurminskaya MV, Zycband EI (1995) Collagen fibrillogenesis in situ: Fibril segments undergo post-depositional modifications resulting in linear and lateral growth during matrix development. Dev Dyn 202: 229–243.

Birk DE, Zycband EI, Winkelmann DA, Trelstad RL (1989) Collagen fibrillogenesis in situ: fibril segments are intermediates in matrix assembly. Proc Natl Acad Sci U S A 86: 4549–4553.

Blaschke UK, Eikenberry EF, Hulmes DJS, Galla HJ, Bruckner P (2000) Collagen XI nucleates self-assembly and limits lateral growth of cartilage fibrils. J Biol Chem 275: 10370–10378.

Bonfanti L, Mironov AA, Jr., MartÆnez-Mençrguez JA, Martella O, Fusella A, Baldassarre M, Buccione R, Geuze HJ, Mironov AA, Luini A (1998) Procollagen traverses the Golgi stack without leaving the lumen of Cisternae: Evidence for cisternal maturation. Cell 95: 993–1003.

Borel A, Eichenberger D, Farjanel J, Kessler E, Gleyzal C, Hulmes DJS, Sommer P, Font B (2001) Lysyl oxidase-like protein from bovine aorta. Isolation and maturation to an active form by bone morphogenetic protein-1. J Biol Chem 276: 48944–48949.

Bottomley MJ, Batten MR, Lumb RA, Bulleid NJ (2001) Quality control in the endoplasmic reticulum. PDI mediates the ER retention of unassembled procollagen C-propeptides. Curr Biol 11: 1114–1118.

Brittingham R, Uitto J, Fertala A (2006) High-affinity binding of the NC1 domain of collagen VII to laminin 5 and collagen IV. Biochem Biophys Res Commun 343: 692–699.

Brodsky B, Persikov AV (2005) Molecular structure of the collagen triple helix. Adv Protein Chem 70: 301–339.

Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, Leikin S, Makareeva E, Kuznetsova NV, Rosenbaum KN, Tifft CJ, Bulas DI, Kozma C, Smith PA, Eyre DR, Marini JC (2007) Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet 39: 359–365.

Canty EG, Kadler KE (2005) Procollagen trafficking, processing and fibrillogenesis. J Cell Sci 118: 1341–1353.

Canty EG, Lu Y, Meadows RS, Shaw MK, Holmes DF, Kadler KE (2004) Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon. J Cell Biol 165: 553–563.

Chakravarti S, Magnuson T, Lass JH, Jepsen KJ, LaMantia C, Carroll H (1998) Lumican regulates collagen fibril assembly: Skin fragility and corneal opacity in the absence of lumican. J Cell Biol 141: 1277–1286.

Chiquet M, Renedo AS, Huber F, Fluck M (2003) How do fibroblasts translate mechanical signals into changes in extracellular matrix production? Matrix Biol 22: 73–80.

Colige A, Beschin A, Samyn B, Goebels Y, van Beeumen J, Nusgens BV, Lapiere CM (1995) Characterization and partial amino acid sequencing of a 107-kDa procollagen I N-proteinase purified by affinity chromatography on immobilized type XIV collagen. J Biol Chem 270: 16724–16730.

Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, Iozzo RV (1997) Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. J Cell Biol 136: 729–743.

Ellgaard L, Ruddock LW (2005) The human protein disulphide isomerase family: Substrate interactions and functional properties. EMBO Rep 6: 28–32.

Eyre DR, Pietka T, Weis MA, Wu JJ (2004) Covalent cross-linking of the NC1 domain of collagen type IX to collagen type II in cartilage. J Biol Chem 279: 2568–2574.

Fields GB, Lauer JL, Dori Y, Forns P, Yu YC, Tirrell M (1998) Protein-like molecular architecture: Biomaterial applications for inducing cellular receptor binding and signal transduction. Biopolymers 47: 143–151.

Fleischmajer R, MacDonald ED, Perlish JS, Burgeson RE, Fisher LW (1990) Dermal collagen fibrils are hybrids of type I and type III collagen molecules. J Struct Biol 105: 162–169.

Franzke CW, Bruckner P, Bruckner-Tuderman L (2005) Collagenous transmembrane proteins: recent insights into biology and pathology. J Biol Chem 280: 4005–4008.

Fromme JC, Schekman R (2005) COPII-coated vesicles: flexible enough for large cargo? Curr Opin Cell Biol 17: 345–352.

Ge G, Seo NS, Liang X, Hopkins DR, Hook M, Greenspan DS (2004) Bone morphogenetic protein-1/tolloid-related metalloproteinases process osteoglycin and enhance its ability to regulate collagen fibrillogenesis. J Biol Chem 279: 41626–41633.

Ghai R, Waters P, Roumenina LT, Gadjeva M, Kojouharova MS, Reid KB, Sim RB, Kishore U (2007) C1q and its growing family. Immunobiology 212: 253–266.

Giraud-Guille MM (1992) Liquid crystallinity in condensed type I collagen solutions. A clue to the packing of collagen in extracellular matrices. J Mol Biol 224: 861–873.

Giraud-Guille MM (1996) Twisted liquid crystalline supramolecular arrangements in morphogenesis. Int Rev Cytol 166: 59–101.

Graham HK, Holmes DF, Watson RB, Kadler KE (2000) Identification of collagen fibril fusion during vertebrate tendon morphogenesis. The process relies on unipolar fibrils and is regulated by collagen–proteoglycan interaction. J Mol Biol 295: 891–902.

Grenard P, Bresson-Hadni S, El Alaoui S, Chevallier M, Vuitton DA, Ricard-Blum S (2001) Transglutaminase-mediated cross-linking is involved in the stabilization of extracellular matrix in human liver fibrosis. J Hepatol 35: 367–375.

Halasz K, Kassner A, Morgelin M, Heinegard D (2007) COMP acts as a catalyst in collagen fibrillogenesis. J Biol Chem, 282: 31166–31173.

Holmes DF, Capaldi MJ, Chapman JA (1986) Reconstitution of collagen fibrils in vitro; the assembly process depends on the initiating procedure. Int J Biol Macromol 8: 161–166.

Holmes DF, Chapman JA, Prockop DJ, Kadler KE (1992) Growing tips of type-I collagen fibrils formed in vitro are near-paraboloidal in shape, implying a reciprocal relationship between accretion and diameter. Proc Nat Acad Sci U S A 89: 9855–9859.

Holmes DF, Watson RB, Chapman JA, Kadler KE (1996) Enzymic control of collagen fibril shape. J Mol Biol 261: 93–97.

Holmskov U, Thiel S, Jensenius JC (2003) Collections and ficolins: humoral lectins of the innate immune defense. Annu Rev Immunol 21: 547–578.

Hopkins DR, Keles S, Greenspan DS (2007) The bone morphogenetic protein 1/Tolloid-like metalloproteinases. Matrix Biol, 26: 508–523.

Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG (2003) Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med 348: 2543–2556.

Hulmes DJS (2002) Building collagen molecules, fibrils, and suprafibrillar structures. J Struct Biol 137: 2–10.

Hulmes DJS, Bruns RR, Gross J (1983) On the state of aggregation of newly secreted procollagen. Proc Natl Acad Sci U S A 80: 388–392.

Hulmes DJS, Kadler KE, Mould AP, Hojima Y, Holmes DF, Cummings C, Chapman JA, Prockop DJ (1989) Pleomorphism in type I collagen fibrils produced by persistence of the procollagen N-propeptide. J Mol Biol 210: 337–345.

Iozzo RV (1999) The biology of the small leucine-rich proteoglycans – Functional network of interactive proteins. J Biol Chem 274: 18843–18846.

Ishida Y, Kubota H, Yamamoto A, Kitamura A, Bachinger HP, Nagata K (2006) Type I collagen in Hsp47-null cells is aggregated in endoplasmic reticulum and deficient in N-propeptide processing and fibrillogenesis. Mol Biol Cell 17: 2346–2355.

Kadler KE (1995) Extracellular matrix 1: fibril-forming collagens. Protein Profile 2: 491–619.

Kadler KE, Holmes DF, Trotter JA, Chapman JA (1996) Collagen fibril formation. Biochem J 316: 1–11.

Kadler KE, Baldock C, Bella J, Boot-Handford RP (2007) Collagens at a glance. J Cell Sci 120: 1955–1958.

Kar K, Amin P, Bryan MA, Persikov AV, Mohs A, Wang YH, Brodsky B (2006) Self-association of collagen triple helix peptides into higher order structures. J Biol Chem 281: 33283–33290.

Knupp C, Squire JM (2005) Molecular packing in network-forming collagens. Adv Protein Chem 70: 375–403.

Koch M, Schulze J, Hansen U, Ashwodt T, Keene DR, Brunken WJ, Burgeson RE, Bruckner P, Bruckner-Tuderman L (2004) A novel marker of tissue junctions: collagen XXII. J Biol Chem 279: 22514–22521.

Koide T (2005) Triple helical collagen-like peptides: engineering and applications in matrix biology. Connect Tissue Res 46: 131–141.

Koide T, Nagata K (2005) Collagen biosynthesis. Top Curr Chem 247: 85–114.

Koide T, Homma DL, Asada S, Kitagawa K (2005) Self-complementary peptides for the formation of collagen-like triple helical supramolecules. Bioorg Med Chem Lett 15: 5230–5233.

Kotch FW, Raines RT (2006) Self-assembly of synthetic collagen triple helices. Proc Natl Acad Sci U S A 103: 3028–3033.

Kuhn K (1987) The classical collagens: types I, II and III. In *Structure and Function of Collagen Types*, Mayne R, Burgeson RE (eds) pp. 1–42. Academic Press: Orlando.

Kvist AJ, Johnson AE, Morgelin M, Gustafsson E, Bengtsson E, Lindblom K, Aszodi A, Fassler R, Sasaki T, Timpl R, Aspberg A (2006) Chondroitin sulfate perlecan enhances collagen fibril formation. Implications for perlecan chondrodysplasias. J Biol Chem 281: 33127–33139.

Kwan APL, Cummings CE, Chapman JA, Grant ME (1991) Macromolecular organization of chicken type X collagen in vitro. J Cell Biol 114: 597–604.

Lees JF, Tasab M, Bulleid NJ (1997) Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. EMBO J 16: 908–916.

Le Goff C, Somerville RP, Kesteloot F, Powell K, Birk DE, Colige AC, Apte SS (2006) Regulation of procollagen amino-propeptide processing during mouse embryogenesis by specialization of homologous ADAMTS proteases: insights on collagen biosynthesis and dermatosparaxis. Development 133: 1587–1596.

Leighton M, Kadler KE (2003) Paired basic/Furin-like proprotein convertase cleavage of Pro-BMP-1 in the trans-Golgi network. J Biol Chem 278: 18478–18484.

Leikina E, Mertts MV, Kuznetsova N, Leikin S (2002) Type I collagen is thermally unstable at body temperature. Proc Natl Acad Sci U S A 99: 1314–1318.

Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ninomiya Y, Oxford JT, Morris NP, Andrikopoulos K, Ramirez F, Wardell BB, Lifferth GD, Teuscher C, Woodward SR, Taylor BA, Seegmiller RE, Olsen BR (1995) A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. Cell 80: 423–430.

Li S, Van Den DC, D'Souza SJ, Chan BM, Pickering JG (2003) Vascular smooth muscle cells orchestrate the assembly of type I collagen via alpha2beta1 integrin, RhoA, and fibronectin polymerization. Am J Pathol 163: 1045–1056.

Linsenmayer TF, Fitch JM, Gordon MK, Cai CX, Igoe F, Marchant JK, Birk DE (1998) Development and roles of collagenous matrices in the embryonic avian cornea. Prog Retin Eye Res 17: 231–265.

Liu X, Wu H, Byrne M, Krane S, Jaenisch R (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. Proc Natl Acad Sci U S A 94: 1852–1856.

Lucero HA, Kagan HM (2006) Lysyl oxidase: an oxidative enzyme and effector of cell function. Cell Mol Life Sci 63: 2304–2316.

MacBeath JR, Shackleton DR, Hulmes DJS (1993) Tyrosine-rich acidic matrix protein (TRAMP) accelerates collagen fibril formation in vitro. J Biol Chem 268: 19826–19832.

Martin R, Farjanel J, Eichenberger D, Colige A, Kessler E, Hulmes DJS, Giraud-Guille MM (2000) Liquid crystalline ordering of procollagen as a determinant of three-dimensional extracellular matrix architecture. J Mol Biol 301: 11–17.

Martin R, Waldmann L, Kaplan DL (2003) Supramolecular assembly of collagen triblock peptides. Biopolymers 70: 435–444.

McAlinden A, Smith TA, Sandell LJ, Ficheux D, Parry DAD, Hulmes DJS (2003) α-helical coiled-coil oligomerization domains are almost ubiquitous in the collagen superfamily. J Biol Chem 278: 42200–42207.

McEwan PA, Scott PG, Bishop PN, Bella J (2006) Structural correlations in the family of small leucine-rich repeat proteins and proteoglycans. J Struct Biol 155: 294–305.

Minamitani T, Ikuta T, Saito Y, Takebe G, Sato M, Sawa H, Nishimura T, Nakamura F, Takahashi K, Ariga H, Matsumoto K (2004) Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen. Exp Cell Res 298: 305–315.

Mironov AA, Mironov AA, Jr., Beznoussenko GV, Trucco A, Lupetti P, Smith JD, Geerts WJ, Koster AJ, Burger KN, Martone ME, Deerinck TJ, Ellisman MH, Luini A (2003) ER-to-Golgi carriers arise through direct en bloc protrusion and multistage maturation of specialized ER exit domains. Dev Cell 5: 583–594.

Moali C, Font B, Ruggiero F, Eichenberger D, Rousselle P, Francois V, Oldberg A, Bruckner-Tuderman L, Hulmes DJS (2005) Substrate-specific modulation of a multisubstrate proteinase. C-terminal processing of fibrillar procollagens is the only BMP-1-dependent activity to be enhanced by PCPE-1. J Biol Chem 280: 24188–24194.

Molnar J, Fong KS, He QP, Hayashi K, Kim Y, Fong SF, Fogelgren B, Szauter KM, Mink M, Csiszar K (2003) Structural and functional diversity of lysyl oxidase and the LOX-like proteins. Biochim Biophys Acta 1647: 220–224.

Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M, Castagnola P, Rauch F, Glorieux FH, Vranka J, Bachinger HP, Pace JM, Schwarze U, Byers PH, Weis M, Fernandes RJ, Eyre DR, Yao Z, Boyce BF, Lee B (2006) CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. Cell 127: 291–304.

Mould AP, Hulmes DJS, Holmes DF, Cummings C, Sear CH, Chapman JA (1990) D-periodic assemblies of type I procollagen. J Mol Biol 211: 581–594.

Myllyharju J (2005) Intracellular post-translational modifications of collagens. Top Curr Chem 247: 115–247.

Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20: 33–43.

Myllyla R, Wang C, Heikkinen J, Juffer A, Lampela O, Risteli M, Ruotsalainen H, Salo A, Sipila L (2007) Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). J Cell Physiol 212: 323–329.

Neame PJ, Kay CJ, McQuillan DJ, Beales MP, Hassell JR (2000) Independent modulation of collagen fibrillogenesis by decorin and lumican. Cell Mol Life Sci 57: 859–863.

Paramonov SE, Gauba V, Hartgerink JD (2005) Synthesis of collagen-like peptide polymers by native chemical ligation. Macromolecules 38: 7555–7561.

Parkinson J, Kadler KE, Brass A (1994) Simple physical model of collagen fibrillogenesis based on diffusion limited aggregation. J Mol Biol 247: 823–831.

Parry DA (2005) Structural and functional implications of sequence repeats in fibrous proteins. Adv Protein Chem 70: 11–35.

Peinado H, Del Carmen Iglesias-de la Cruz M, Olmeda D, Csiszar K, Fong KS, Vega S, Nieto MA, Cano A, Portillo F (2005) A molecular role for lysyl oxidase-like 2 enzyme in snail regulation and tumor progression. EMBO J 24: 3446–3458.

Plumb DA, Dhir V, Mironov A, Ferrara L, Poulsom R, Kadler KE, Thornton DJ, Briggs MD, Boot-Handford RP (2007) Collagen XXVII is developmentally regulated and forms thin fibrillar structures distinct from those of classical vertebrate fibrillar collagens. J Biol Chem 282: 12791–12795.

Porter S, Clark IM, Kevorkian L, Edwards DR (2005) The ADAMTS metalloproteinases. Biochem J 386: 15–27.

Prockop DJ, Fertala A (1998) Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides – Demonstration that assembly is driven by specific binding sites on the monomers. J Biol Chem 273: 15598–15604.

Rentz TJ, Poobalarahi F, Bornstein P, Sage EH, Bradshaw AD (2007) SPARC regulates processing of procollagen I and collagen fibrillogenesis in dermal fibroblasts. J Biol Chem 282: 22062–22071.

Ricard-Blum S, Dublet B, van der Rest M (2000) *Unconventional Collagens*. Oxford University Press: Oxford

Ricard-Blum S, Ruggiero F, van der Rest M (2005) The collagen superfamily. Top Curr Chem 247: 35–84.

Richardson SH, Starborg T, Lu Y, Humphries SM, Meadows RS, Kadler KE (2007) Tendon development requires regulation of cell condensation and cell shape via cadherin-11-mediated cell–cell junctions. Mol Cell Biol 27: 6218–6228.

Romanic AM, Adachi E, Kadler KE, Hojima Y, Prockop DJ (1991) Copolymerization of pNcolagen III and collagen I. J Biol Chem 266: 12703–12709.

Ruotsalainen H, Sipila L, Vapola M, Sormunen R, Salo AM, Uitto L, Mercer DK, Robins SP, Risteli M, Aszodi A, Fassler R, Myllyla R (2006) Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. J Cell Sci 119: 625–635.

Scott PG, McEwan PA, Dodd CM, Bergmann EM, Bishop PN, Bella J (2004) Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. Proc Natl Acad Sci U S A 101: 15633–15638.

Seidah NG, Prat A (2005) Proprotein convertases in the secretory pathway, cytosol and extracellular milieu. Essays Biochem 38: 79–94.

Snellman A, Tuomisto A, Koski A, Latvanlehto A, Pihlajaniemi T (2007) The role of disulfide bonds and alpha-helical coiled-coils in the biosynthesis of type XIII collagen and other collagenous transmembrane proteins. J Biol Chem 282: 14898–14905.

Speranza ML, Valentini G, Calligaro A (1987) Influence of fibronectin on the fibrillogenesis of type I and type III collagen. Coll Relat Res 7: 115–123.

Stephan S, Sherratt MJ, Hodson N, Shuttleworth CA, Kielty CM (2004) Expression and supramolecular assembly of recombinant alpha1(viii) and alpha2(viii) collagen homotrimers. J Biol Chem 279: 21469–21477.

Sullivan MM, Barker TH, Funk SE, Karchin A, Seo NS, Hook M, Sanders J, Starcher B, Wight TN, Puolakkainen P, Sage EH (2006) Matricellular hevin regulates decorin production and collagen assembly. J Biol Chem 281: 27621–27632.

Svensson L, Aszùdi A, Reinholt FP, Fèssler R, Heinegard D, Oldberg A (1999) Fibromodulin-null mice have abnormal collagen fibrils, tissue organization, and altered lumican deposition in tendon. J Biol Chem 274: 9636–9647.

van der Slot AJ, Zuurmond AM, Bardoel AF, Wijmenga C, Pruijs HE, Sillence DO, Brinckmann J, Abraham DJ, Black CM, Verzijl N, DeGroot J, Hanemaaijer R, TeKoppele JM, Huizinga TW, Bank RA (2003) Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. J Biol Chem 278: 40967–40972.

Velling T, Risteli J, Wennerberg K, Mosher DF, Johansson S (2002) Polymerization of type I and III collagens is dependent on fibronectin and enhanced by integrins alpha 11beta 1 and alpha 2beta 1. J Biol Chem 277: 37377–37381.

Verderio EA, Johnson TS, Griffin M (2005) Transglutaminases in wound healing and inflammation. Prog Exp Tumor Res 38: 89–114.

Walmsley AR, Batten MR, Lad U, Bulleid NJ (1999) Intracellular retention of procollagen within the endoplasmic reticulum is mediated by prolyl 4-hydroxylase. J Biol Chem 274: 14884–14892.

Wang WM, Lee S, Steiglitz BM, Scott IC, Lebares CC, Allen ML, Brenner MC, Takahara K, Greenspan DS (2003) Transforming growth factor-beta induces secretion of activated ADAMTS-2. A procollagen III N-proteinase. J Biol Chem 278: 19549–19557.

Ward NP, Hulmes DJS, Chapman JA (1986) Collagen self-assembly in vitro: electron microscopy of initial aggregates formed during the lag phase. J Mol Biol 190: 107–112.

Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, Birk DE (2004) Type V collagen controls the initiation of collagen fibril assembly. J Biol Chem 279: 53331–53337.

Wenstrup RJ, Florer JB, Davidson JM, Phillips CL, Pfeiffer BJ, Menezes DW, Chervoneva I, Birk DE (2006) Murine model of the Ehlers–Danlos syndrome. col5a1 haploinsufficiency disrupts collagen fibril assembly at multiple stages. J Biol Chem 281: 12888–12895.

Williams BR, Gelman RA, Poppke DC, Piez KA (1978) Collagen fibril formation: optimal in vitro conditions and preliminary kinetic results. J Biol Chem 235: 6578–6585.

Xu T, Bianco P, Fisher LW, Longenecker G, Smith E, Goldstein S, Bonadio J, Boskey A, Heegaard AM, Sommer B, Satomura K, Dominguez P, Zhao C, Kulkarni AB, Robey PG, Young MF (1998) Targeted disruption of the biglycan gene leads to an osteoporosis-like phenotype in mice. Nat Genet 20: 78–82.

Yeowell HN, Walker LC (2000) Mutations in the lysyl hydroxylase 1 gene that result in enzyme deficiency and the clinical phenotype of Ehlers-Danlos syndrome type VI. Mol Genet Metab 71: 212–224.