Collagens, modifying enzymes and their mutations in humans, flies and worms

Johanna Myllyharju and Kari I. Kivirikko

Collagen Research Unit, Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology, University of Oulu, FIN-90014 Oulu, Finland

Collagens and proteins with collagen-like domains form large superfamilies in various species, and the numbers of known family members are increasing constantly. Vertebrates have at least 27 collagen types with 42 distinct polypeptide chains, >20 additional proteins with collagen-like domains and ~20 isoenzymes of various collagen-modifying enzymes. *Caenorhabditis elegans* has ~175 cuticle collagen polypeptides and two basement membrane collagens. *Drosophila melanogaster* has far fewer collagens than many other species but has ~20 polypeptides similar to the catalytic subunits of prolyl 4-hydroxylase, the key enzyme of collagen synthesis. More than 1300 mutations have so far been characterized in 23 of the 42 human collagen genes in various diseases, and many mouse models and *C. elegans* mutants are also available to analyse the collagen gene family and their modifying enzymes.

The collagens are a family of extracellular matrix proteins that play a dominant role in maintaining the structure of various tissues and also have many other important functions. For example, collagens are involved in cell adhesion, chemotaxis and migration, and the dynamic interplay between cells and collagens regulates tissue remodeling during growth, differentiation, morphogenesis and wound healing, and in many pathologic states.

All collagen molecules consist of three polypeptide chains, called α chains (Box 1), and contain at least one domain composed of repeating Gly-X-Y sequences in each of the constituent chains. In some collagens all three α chains are identical, whereas in others the molecules contain two or even three different α chains. The three α chains are each coiled into a left-handed helix and are then wound around a common axis to form a triple helix with a shallow right-handed superhelical pitch, so that the final structure is a rope-like rod. The presence of glycine, the smallest amino acid, in every third position is essential for the packing of this coiled-coil structure. The X and Y positions can have any amino acid other than glycine, but proline is often found in the X position and 4-hydroxyproline in the Y position. The 4-hydroxyprolines play a particularly important role because these residues are essential for the stability of the triple helix.

Collagens are the most abundant proteins in the human body, constituting ~30% of its protein mass. The important roles of these proteins have been clearly demonstrated by the wide spectrum of diseases caused by a large number of mutations found in collagen genes. This article will review the collagen superfamilies and their mutations in vertebrates, *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*. The genomes of these two model invertebrate species have been fully sequenced, and it is therefore possible to identify all of the collagen genes present in these species. Because of the extensive literature in these fields, this review will focus primarily on recent advances. More detailed accounts and more complete references can be found in previous reviews, for example Refs [1–6].

The collagen superfamily in vertebrates

Types of collagen

Vertebrates have at least 27 collagen types with 42 distinct α chains in total, and >20 additional proteins have collagen-like domains. All collagens also possess non-collagenous domains in addition to the actual collagen domains. Most collagens form supramolecular assemblies, such as fibrils and networks, and the superfamily can be

Box 1. Human collagen nomenclature

- **Collagen types I-XXVII**. Collagens are numbered with roman numerals in the order of their discovery.
- **Collagen polypeptide chains**. These are called α chains, each collagen molecule consisting of three of them. Depending on the collagen type, the three α chains can be either identical or the molecule can contain two or even three different α chains. The α chains of a specific collagen type are numbered with arabic numerals and the collagen type is given in parentheses. For example α1(I) and α2(I) are the α1 and α2 chains of type I collagen, and α1(II) is the α1 chain of type II collagen.
- **Procollagen**. The fibril-forming collagens (Figure 1a) are synthesized as procollagen molecules, which have propeptides at the N and C-terminal ends of their polypeptide chains, called proα chains.
- **Genes encoding the collagen chains**. These are named by the prefix COL followed by an arabic number for the collagen type, the letter 'A' stands for α chain and an arabic number for the chain. For example, the genes *COL1A1* and *COL1A2* encode the α1 and α2 chains of type I collagen, respectively, and the gene *COL2A1* encodes the α1 chain of type II collagen.

Corresponding author: Johanna Myllyharju (johanna.myllyharju@oulu.fi).

 0168-9525/$ - see front matter doi:10.1016/j.tig.2003.11.004

divided into several subfamilies on the basis of these assemblies or other features (Figure 1). Collagen types I–XIX have been discussed in many previous reviews [1–4], whereas collagen types XX–XXVII (Table 1) have been reported only during the past three years [7–14]. Some collagens have a restricted tissue distribution: for example, types II, IX and XI, which are found almost exclusively in cartilage; type X, found only in hypertrophic cartilage; the family of type IV collagens in basement membranes; type VII in the anchoring fibrils for basement membranes; and type XVII in skin hemidesmosomes. By contrast some collagen types are found in most extracellular matrices. The highly heterogeneous group of proteins that contain collagen domains but have not been defined as collagens (Figure 1,i) includes: the subcomponent C1q of complement, a C1q-like factor, adiponectin, at least eight collectins and three ficolins (humoral lectins of the innate immune defence system), the tail structure of acetylcholinesterase, three macrophage receptors, ectodysplasin, two EMILINS (elastic fibre-associated glycoproteins) and a src-homologous-and-collagen protein [3,4,6,15].

Collagen fibrils often consist of more than one collagen type. For example, the type I collagen fibrils often contain small amounts of types III, V and XII, whereas the type II collagen fibrils of cartilage also contain types IX and XI. Collagen types V and XI can also form hybrid molecules [e.g. having an α1(XI) and an α2(V) chain in the same molecule]. The six α chains of type IV form at least three types of molecule $[\alpha1(IV)]_2\alpha2(IV)$, α3(IV)α4(IV)α5(IV) and $[\alpha6(IV)]_2\alpha5(IV)$ [16]. Further heterogeneity within the superfamily is caused by alternative splicing of the transcripts of many of the genes and the use of alternative promoters in some genes. The large number of structures present in members of the superfamily implies that they are involved in numerous different biological functions [1–4].

The non-collagenous domains of many collagens also have important functions. Major interest has been focused on endostatin, a proteolytically derived 20 kDa C-terminal fragment of collagen XVIII, and restin, a corresponding fragment of collagen XV, which inhibit endothelial cell migration and angiogenesis and reduce tumour growth in animal models [3,4,17]. The C-terminal non-collagenous domain of collagen IV also inhibits angiogenesis and tumour growth [3,17], whereas other functions have been described for the non-collagenous domains of certain other collagens [3].

Biosynthesis and modifying enzymes

Collagen synthesis involves many post-translational modifications that require three collagen hydroxylases [18,19], two collagen glycosyltransferases [1,3,4], two specific proteinases [20] to cleave the N and C propeptides from the procollagen molecules (family a in Figure 1) and one specific oxidase [21] to initiate crosslink formation (Figure 2). Other enzymes include peptidyl proline *cis-trans* isomerase and protein disulfide isomerase (PDI), which has at least three functions: (i) to catalyze the formation of intrachain and interchain disulfide bonds; (ii) to serve as the β subunit in collagen prolyl 4-hydroxylases; and (iii) to act as a chaperone that binds nascent collagen chains and prevents their aggregation [3,18,19]. Collagen synthesis also involves a specific chaperone, Hsp47. Homozygous knockout of this gene in mice is lethal at the embryonic stage, indicating that this protein is essential for normal development [22,23].

Collagen prolyl 4-hydroxylase, an $\alpha_2\beta_2$ tetramer located within the lumen of the endoplasmic reticulum, plays a central role in collagen synthesis because 4-hydroxyproline residues are essential for the formation of triple-helical molecules *in vivo*. Collagen prolyl 4-hydroxylase has at least three isoenzymes in humans, with distinct α subunits but all have PDI as their β subunit [18,19,24,25]. A novel family of three cytoplasmic prolyl 4-hydroxylases has recently been shown to play a key role in the regulation of the hypoxia-inducible transcription factor HIF [19,26–28]. These enzymes have no PDI subunit, have different requirements with respect to the sequence flanking the prolines that are hydroxylated and have markedly higher K_m values for oxygen than the collagen prolyl 4-hydroxylases [26–29]. Lysyl hydroxylase also has at least three isoenzymes [3,4,30,31], whereas lysyl oxidase has at least five [3,21,32,33]. The two proteinases that cleave the N and C propeptides from procollagen molecules in large transport vesicles close to the plasma membrane [34] and in the extracellular matrix each have at least three isoenzymes [3,4,20,35]. The C proteinases also process various other precursor proteins of the extracellular matrix and belong to the tolloid family. The main C proteinase isoenzyme is identical to a protein previously called bone morphogenic protein-1 (BMP-1) [3,4,20]. The two specific collagen glycosyltransferases have not been cloned but one of the lysyl hydroxylase isoenzymes has also been shown to possess small amounts of these enzyme activities [36,37]. Work is

Table 1. Recently identified collagen types XX–XXVII

Type	Chain	Residues	Location	Group[a]	Ref.
XX	α1(XX)	1473	Corneal epithelium, skin, cartilage and tendon	b	[7]
XXI	α1(XXI)	957	Many tissues	b	[8]
XXII[b]	α1(XXII)	1616	Tissue junctions[c]	b	
XXIII	α1(XXIII)	540	Metastatic tumour cells	g	[9]
XXIV	α1(XXIV)	1714	Developing bone and cornea	a	[10]
XXV	α1(XXV)	666	Neurons	g	[11]
XXVI	α1(XXVI)	438	Testis, ovary	b	[12]
XXVII	α1(XXVII)	1860	Cartilage, eye, ear and lung	a	[13,14]

[a]Group in Figure 1.
[b]GenBank accession number AF406780.
[c]M. Koch and L. Bruckner-Tuderman, unpublished.

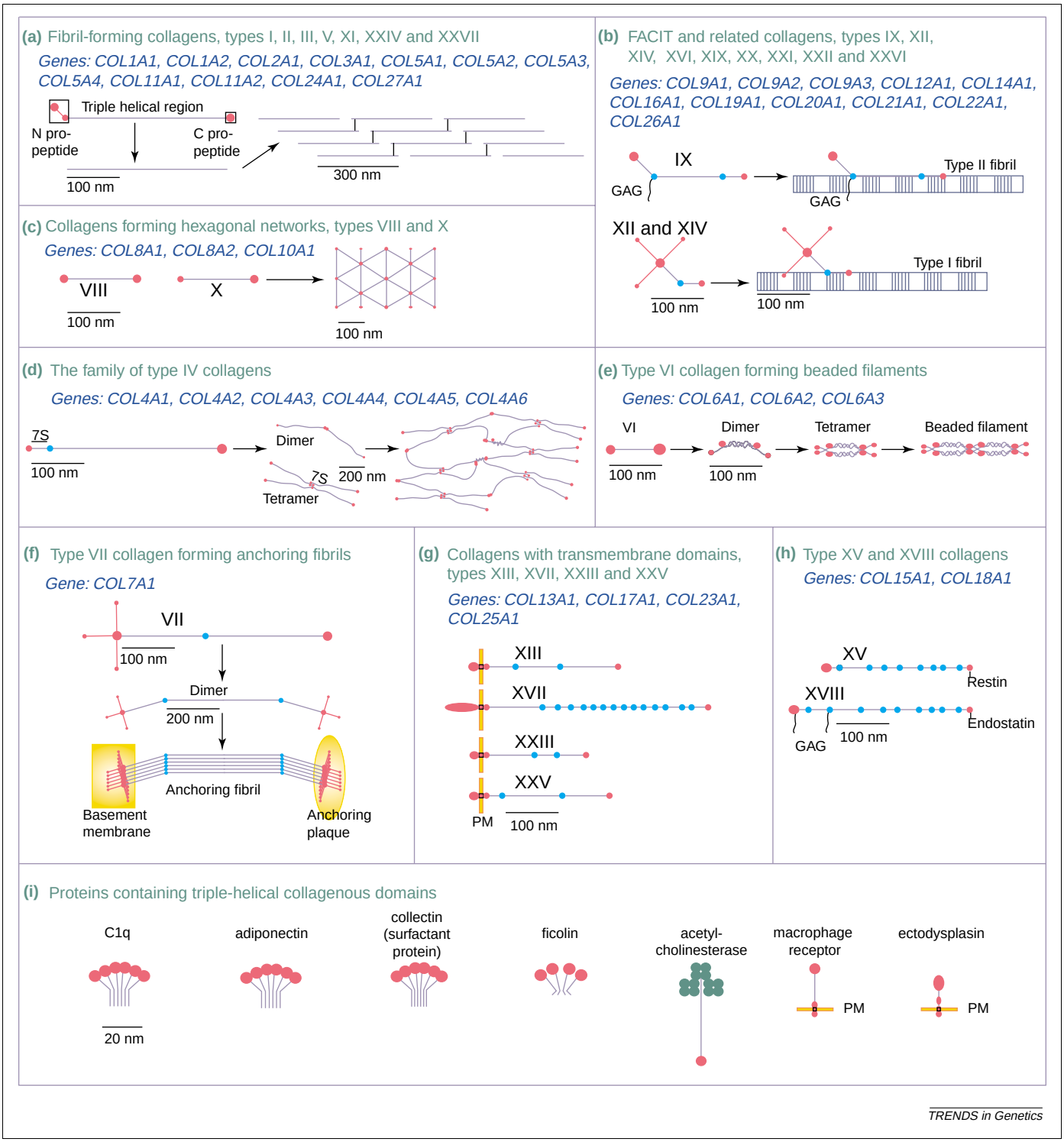

Figure 1. Members of the collagen superfamily and their known supramolecular assemblies. The collagen superfamily can be divided into nine families on the basis of the supramolecular assemblies and other features of its members: **(a)** fibril-forming collagens; **(b)** fibril-associated collagens with interrupted triple helices (FACITs) located on the surface of fibrils, and structurally related collagens; **(c)** collagens forming hexagonal networks; **(d)** the family of type IV collagens located in basement membranes; **(e)** type VI collagen, which forms beaded filaments; **(f)** type VII collagen, which forms anchoring fibrils for basement membranes; **(g)** collagens with transmembrane domains; and **(h)** the family of type XV and XVIII collagens. The supramolecular assemblies of families **(g)** and **(h)** are unknown and are therefore not shown in this figure. The polypeptide chains found in the 27 collagen types are coded by 42 genes in total (shown in blue), each molecule consisting of three polypeptide chains that can be either identical or different. An additional highly heterogenous group **(i)** within the superfamily comprises proteins that possess collagenous domains but have not been defined as collagens. Some of the group **(i)** proteins could also be defined as collagens, although some of the collagens might also belong to this group because there are no distinct criteria for distinguishing between a collagen and a protein containing a collagen domain(s) [1]. The collagen domains are shown in purple, the N and C-terminal non-collagenous domains are in dark pink, and the non-collagenous domains interrupting the triple helix in light blue, short interruptions of a few amino acids are not shown. For acetylcholinesterase, the catalytic domain (shown in green) and the tail structure are products of separate genes. Modified and up-dated from Refs [1,3]. Abbreviation: PM, plasma membrane.

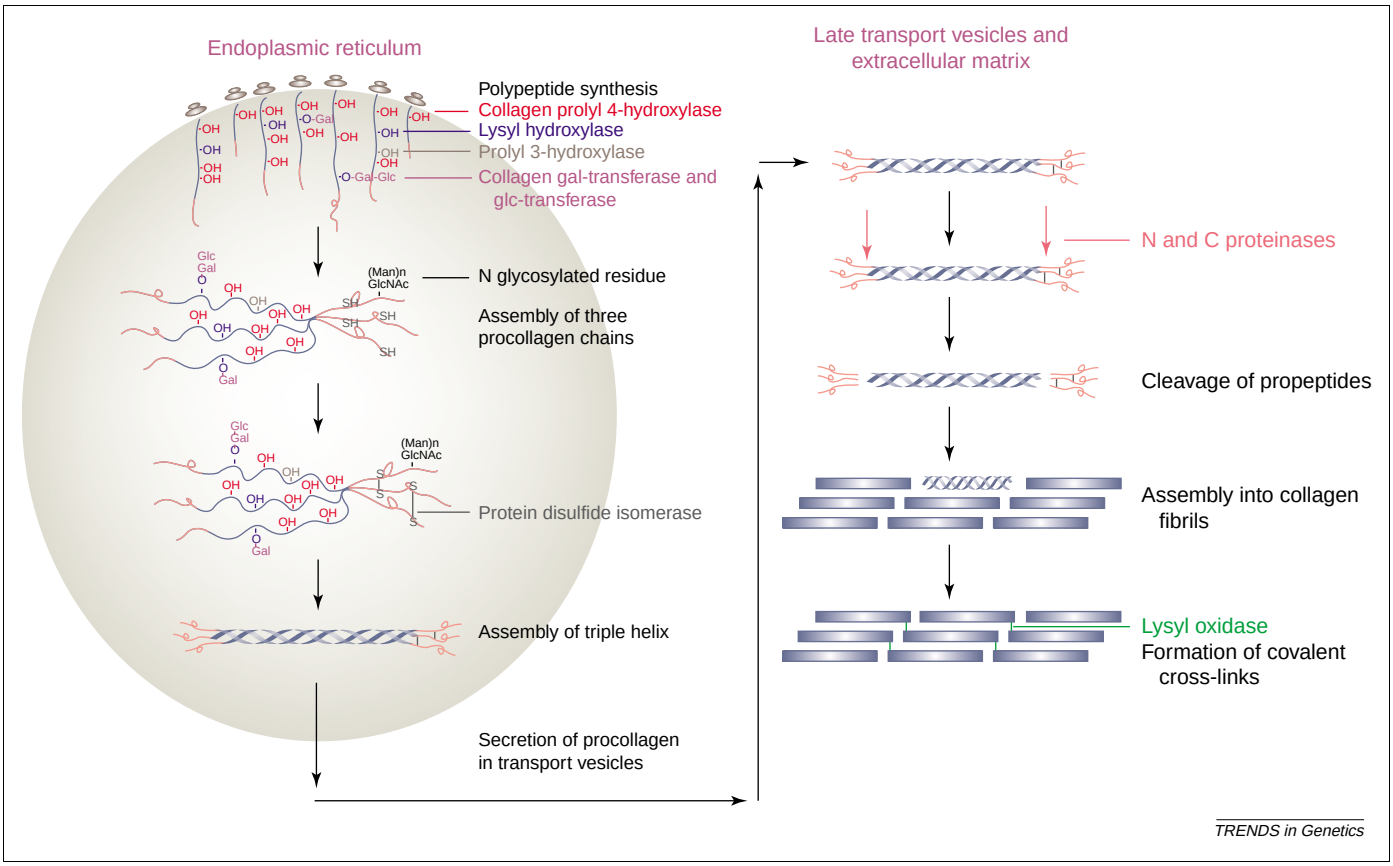

Figure 2. The main steps in the synthesis of a fibril-forming collagen. The polypeptide chains are synthesized on membrane-bound ribosomes and secreted into the lumen of the endoplasmic reticulum, where the main steps in biosynthesis are: (i) cleavage of the signal peptides (not shown); (ii) hydroxylation of certain proline and lysine residues to 4-hydroxyproline, 3-hydroxyproline and hydroxylysine; (iii) glycosylation of some of the hydroxylysine residues to galactosylhydroxylysine and glucosylgalactosylhydroxylysine; (iv) glycosylation of certain asparagine residues in the C propeptides, or both the N and C propeptides, by reactions similar to those in many other proteins; (v) association of three C propeptides directed by specific recognition sequences; and (vi) formation of intramolecular and intermolecular disulfide bonds. A nucleus for the assembly of the triple helix is formed in the C-terminal region after the C propeptides have become associated and ~100 proline residues have been hydroxylated to 4-hydroxyproline in each of the chains, and the triple helix is then propagated towards the N-terminus in a zipper-like fashion. The procollagen molecules are transported from the endoplasmic reticulum through the Golgi stacks. They begin to aggregate laterally during transport to form secretory vesicles. The subsequent steps are cleavage of the N and C propeptides, spontaneous self-assembly of the resulting collagen molecules into fibrils, and formation of covalent crosslinks initiated by oxidation of the ε amino group in certain lysine and hydroxylysine residues into reactive aldehyde derivatives [1–4].

in progress to elucidate differences in the expression patterns and functions of the various isoenzymes of the collagen modifying enzymes.

Mutations in human collagens and their modifying enzymes and corresponding mouse models

About 1100 mutations have been reported in just six of the 42 collagen genes, *COL1A1, COL1A2, COL2A1, COL3A1, COL4A5* and *COL7A1*, which encode the two kinds of polypeptides of type I collagen, the polypeptides of types II, III and VII and the α5 chain of type IV [1,3,38–44]. The number of identified mutations will probably increase because many of those found recently might not have been regarded as worth publishing separately. About 200 other mutations have been reported in 17 additional genes [1,3,38,39,41–47], whereas no data are available on mutations in 19 of the 42 genes. The mutations reported by mid-2000 have been reviewed previously [3]. The main development since then is that several mutations in one additional gene, *COL8A2*, have been reported in two forms of corneal endothelial dystrophy, one of which is among the most common indications for corneal transplantation [46]. In addition, new mutations have been reported in the 22 genes in which mutations were known in 2000 [38–45,47].

Most of the mutations that have been reported are single-base substitutions that convert the codon of the obligate glycine to that of a bulkier residue and either prevent the folding of the triple helix beyond this point or cause an interruption in the helix. Because the triple helix of most collagens is propagated from the C-terminus to the N-terminus (Figure 2), there is a tendency for a glycine substitution that is closer to the C-terminal end of the triple-helical domain to produce a more severe phenotype than a similar substitution that is closer to the N-terminal end [1,3]. There are numerous exceptions to this rule, however, which are probably explained by the fact that the triple helix has regions of high and low stability as determined by the amino acids present in the X and Y positions [48]. Mutations in different regions, therefore, have different effects. A continuous triple helix seems to be particularly important for the fibril-forming collagens, whereas some collagens that normally contain several interruptions in their triple helices can tolerate an additional interruption with mild or no consequences. Other mutations change the codon of an X or Y-position residue to that of another amino acid or to a translation

stop codon. Further mutations lead to abnormal RNA splicing or are gene deletions, insertions, duplications or complex rearrangements. Most of the amino acid substitutions in the X and Y positions produce milder phenotypes than mutations of the obligate glycines, and some X and Y-position substitutions are probably non-pathogenic polymorphisms [3].

Mutations that lead to the production of a structurally altered polypeptide chain that is still able to associate with other chains usually cause more severe phenotypes than mutations that prevent trimer formation and heterozygous null alleles. Trimer formation from a mutant chain and normal chains can interfere with either the folding of the triple helix or the formation of the supramolecular assemblies. In the former case, the unfolded trimers containing both mutant and normal chains will first accumulate within cells and subsequently be degraded. If the chains form triple-helical molecules, the mutant molecules can have kinks or other abnormalities, which can reduce or delay the formation of supramolecular assemblies or alter their structure and function [1,3]. Most of the diseases caused by collagen mutations are dominantly inherited but there are also many examples of recessive inheritance [38–47].

The vast majority of the known collagen mutations have been identified in relatively rare heritable diseases (Table 2), including: (i) osteogenesis imperfecta, which is characterized by bone fragility, but also involves other tissues that are rich in type I collagen; (ii) various subtypes of the Ehlers-Danlos syndrome (EDS), a heterogeneous group of diseases characterized by joint hypermobility, skin changes, occasional skeletal deformities and rupture of the hollow organs; (iii) various chondrodysplasias, varying in severity from perinatal lethality to a very mild disease and non-syndromic hearing loss; (iv) autosomally inherited and X-linked forms of Alport syndrome, a disease characterized by haematuria and progressing to end-stage renal failure; (v) Bethlem myopathy and Ullrich muscular dystrophy; (vi) two forms of epidermolysis bullosa, a blistering skin disease; (vii) two forms of corneal endothelial dystrophy; and (viii) Knobloch syndrome, a disease characterized by high myopia, vitroretinal detachment and occipital encephalocele [1,3,38–47].

All the mutations in the two type I collagen genes that cause EDS type VII (Table 2) prevent cleavage of the N propeptide, EDS type VIIA results from mutations in the *COL1A1* gene and tends to be phenotypically more severe than type VIIB, which results from mutations in *COL1A2* [3,41]. One *COL1A1* mutation (not shown in Table 2) that replaces the codon of an X-position arginine with one of cysteine has also been reported in EDS I/II, thus indicating an overlap of phenotypes caused by mutations in different collagen types [3]. Overlaps are also seen in the case of cartilage collagens, where mutations in the main cartilage collagen, type II, usually produce more severe phenotypes than corresponding mutations in the minor cartilage collagens (i.e. types IX and XI) [3,42]. In the case of type VI collagen, Ullrich dystrophy is caused in most cases by recessive mutations whereas the milder Bethlem myopathy is caused by dominant mutations [3,45].

A few collagen mutations have also been identified in common diseases, such as osteoporosis, arterial aneurysms and the two most common musculoskeletal diseases, osteoarthrosis and intervertebral disc disease (Table 2). Of particular interest are two recently characterized mutations in the *COL9A2* and *COL9A3* genes that change a codon of an X-position glutamine or Y-position arginine, respectively, to a tryptophan codon [49,50]. The mutation in the *COL9A2* gene causes intervertebral disc disease in many families in a dominantly inherited fashion whereas the *COL9A3* mutation is a strong predisposing factor for this disease. It is probable that many additional collagen mutations will be found that either cause or act as predisposing factors for these and other common diseases.

All the 42 collagen genes identified in humans are probably present in the mouse genome and many mouse lines (Table 2) are now available that harbour a mutation in a collagen gene [1,3,51–56]. Such mouse models are particularly useful for defining the significance and function of proteins such as collagens, which are large, insoluble and difficult to study functionally. In addition, mouse models are useful for analyzing the consequences of mutations in various genes of the superfamily, for identifying additional diseases caused by these mutations and even for testing potential therapies. Indeed, many mouse models have reproduced the phenotypes of various human diseases (Table 2) and in several cases have even led to the identification of additional human diseases that are caused by mutations in collagen genes [1,3].

Homozygous mutations have been reported in only three out of ~20 genes encoding various isoenzymes of the human collagen-modifying enzymes (Table 2), namely in the genes for lysyl hydroxylase-1 and a procollagen N proteinase (also known as ADAMTS-2), involved in two subtypes of EDS [1,3,18,41], and for lysyl hydroxylase-2 in Bruck syndrome, a disease characterized by osteoporosis, joint contractures, fragile bones and short stature [57]. The lysyl hydroxylase mutations prevent the formation of stable hydroxylysine-derived crosslinks, whereas the N proteinase mutations lead to an accumulation of partially processed molecules containing the N propeptide.

In addition to these human enzyme mutations, homozygous knockout mouse models have been generated [1,3,58–61] for genes encoding the catalytic subunit of two of the three known collagen prolyl 4-hydroxylase isoenzymes, one lysyl hydroxylase, one procollagen N proteinase, two procollagen C proteinases and the first described isoenzyme of lysyl oxidase (Table 2). Homozygous inactivation of many of these genes causes embryonic or perinatal lethality, demonstrating a crucial role for these enzymes in collagen synthesis and development.

Collagens and their modifying enzymes in *Drosophila*

Drosophila melanogaster has three conserved genes encoding basement membrane collagens, two α chains of type IV collagen and a homologue of type XV and XVIII collagens [62]. An additional gene encodes pericardin, a protein in which the collagen domain shows some

Table 2. Mutations in human collagens, their modifying enzymes and the corresponding mouse models[a,b]

Name Human Gene	**Human disease**	**Name[c]** *Mouse Gene*	**Mouse model**
COL1A1, COL1A2	OI; osteoporosis; EDS type VIIA and EDS type VIIB [40,41]	*Col1a1*	TG, T
		Col1a2	N
COL2A1	Several chondrodysplasias; osteoarthrosis [42]	*Col2a1*	TG, KO, N
COL3A1	EDS type IV; arterial aneurysms [41]	*Col3a1*	KO
COL4A3, COL4A4	Autosomal forms of Alport syndrome [43]	*Col4a3*	KO
COL4A5	X-linked forms of Alport syndrome [43]		None
COL4A5 and *COL4A6*	Alport syndrome with diffuse oesophageal leiomyomatosis		None
COL5A1, COL5A2	EDS type I; EDS type II [41]	*Col5a2*	T
COL6A1, COL6A2, COL6A3	Bethlem myopathy; Ullrich muscular dystrophy [45]	*Col6a1*	KO
COL7A1	Dystrophic forms of EB [44]	*Col7a1*	KO
COL8A2	Two forms of corneal endothelial dystrophy [46]		None
COL9A1, COL9A2, COL9A3	Multiple epiphyseal dysplasia; osteoarthrosis; intervertebral disc disease [42,49,50]	*Col9a1*	TG, KO
COL10A1	Schmid metaphyseal chondrodysplasia [42]	*Col10a1*	TG, KO
COL11A1, COL11A2	Several mild chondrodysplasias; nonsyndromic hearing loss; osteoarthrosis [42]	*Col11a2*	KO, T
COL12A1	Not identified; disruption of matrix structure of periodontal ligaments and skin in mouse	*Col12a1*	TG [52]
COL13A1	Not identified; embryonic lethality or progressive muscular atrophy in mouse	*Col13a1*	TG, T [53,54]
COL15A1	Not identified; skeletal myopathy and cardiovascular defects in mouse	*Col15a1*	KO [55]
COL17A1	Two forms of EB [44]		None
COL18A1	Knobloch and pigment dispersion syndromes [47]	*Col18a1*	KO [56]
COL19A1	Not identified; abnormal muscle layer in the oesophagus in mouse[d]	*Col19a1*	KO[d]
Modifying enzyme		**Modifying enzyme**	
P4H-α(I)	Not identified; embryonic lethality in mouse[e]	P4H-α(I)	KO[e]
P4H-α(II)	Not identified; KO mice are viable[f]	P4H-α(II)	KO[f]
LH-1	Type VI EDS [41]		None
LH-2	Bruck syndrome [57]		None
LH-3	Not identified; embryonic lethality and lack of type IV collagen in basement membranes in mouse[g]	LH-3	KO[g]
ADAMTS-2	type VIIC EDS [41].	ADAMTS-2	KO [58]
BMP-1	Not identified; perinatal lethality and failure of ventral body wall closure in mouse	BMP-1	KO
Tolloid-like-1	Not identified; embryonic lethality and cardiac failure in mouse	Tolloid-like-1	KO [59]
Lox	Not identified; perinatal lethality and aortic aneurysms and cardiovascular dysfunction in mouse	Lox	KO [60,61]

[a]Abbreviations: OI, osteogenesis imperfecta; EDS, Ehlers-Danlos syndrome; EB, epidermolysis bullosa; P4H, collagen prolyl 4-hydroxylase; LH, lysyl hydroxylase; ADAMTS-2, N proteinase isoenzyme; BMP-1, C proteinase isoenzyme; Tolloid-like-1, C proteinase isoenzyme; Lox, lysyl oxidase-1; T, transgenic (i.e. expression of a mutant polypeptide); KO, knock-out; T, other targeted, (i.e. knock-in); N, naturally occurring mutations.
[b]References to most human mutations are found in [1,3,38,39] and to most mouse mutations in [1,3,51]. References given in the Table therefore indicate only some additional reviews or recent original articles.
[c]The mouse gene is shown only when a mouse model is available.
[d]H. Sumiyoshi *et al.*, unpublished, see Ref [3].
[e]T. Holster *et al.*, unpublished.
[f]The phenotype has not yet been analysed. O. Pakkanen *et al.*, unpublished.
[g]K. Rautavuoma *et al.*, unpublished.

similarity to type IV and which is involved in the morphogenesis and maintenance of the heart epithelium during dorsal ectoderm closure [63], but *Drosophila* has no fibril-forming collagen with a long triple helix [62].

One surprising aspect of the *Drosophila* collagen superfamily is the presence of ~20 genes encoding polypeptides of 480–550 residues with a similarity to the catalytic α subunits of the vertebrate collagen prolyl 4-hydroxylases [19,64]. Many of these show tissue-specific embryonic expression (e.g. in the salivary gland, mouth-part precursor, proventriculus or epidermis) [64]. Only one of the encoded polypeptides has been characterized in detail [65]. It is expressed only in larvae and the embryonic mouth-part precursor, but not in adults [64,65], and combines with PDI to form an active $\alpha_2\beta_2$ tetramer with properties similar to those of the vertebrate collagen prolyl 4-hydroxylases [65]. An additional gene encodes a HIF prolyl 4-hydroxylase indicating that *Drosophila* has a hypoxic response pathway similar to that in vertebrates [27]. The *Drosophila* collagen prolyl 4-hydroxylase family appears to be markedly larger than the corresponding vertebrate or nematode families, even though the number of collagen genes in *Drosophila* is much smaller. It is therefore possible that the functions of many of the *Drosophila* prolyl 4-hydroxylases might be to hydroxylate proline residues in proteins other than the collagens, and

thus detailed studies on these enzymes might help researchers to identify additional functions for the human prolyl 4-hydroxylases.

The collagen families and their modifying enzymes in *Caenorhabditis elegans*

Two major collagen families are present in *Caenorhabditis elegans* – the cuticle collagens and the basement membrane collagens. The *C. elegans* cuticle is an exoskeleton that is synthesized by the underlying hypodermis. The major proteins of the cuticle are small collagen-like polypeptides of ~30 kDa encoded by a multigene family of ~175 members [66–68] (J. Kramer, unpublished). These polypeptides typically have two collagen domains, a smaller N-terminal domain and a larger C-terminal domain, with 8–10 and 40–42 Gly-X-Y repeats, respectively. The repeats in the C-terminal domain usually have 1–4 small interruptions. These two domains are separated and flanked by three cysteine-containing, non-collagenous domains, and the cuticle collagen family can be divided into four main groups and several additional small groups on the basis of conserved cysteine patterns. The N-terminal non-collagenous domain varies in size and contains a cleavage site for a putative subtilisin-like protease that processes a procollagen precursor to a mature polypeptide [66–68]. The cuticle is synthesized five times during the life cycle, once in the embryo before hatching and subsequently at the end of each of the four larval stages before moulting. The cuticle collagen genes are expressed in a distinct temporal fashion, the pattern of which is repeated during each cuticle synthesis but not all the genes are expressed at the same time; some are early, some are intermediate and some are late with respect to the secretion of the new cuticle [67,68]. Sets of collagens that are temporally coexpressed have been shown in genetic studies to interact and to be capable of forming functionally distinct structures [69,70]. The cuticle collagens contain many interchain disulfide bonds that can be either intramolecular or intermolecular. These collagens also contain tyrosine and putative γ carboxyl glutamine-derived crosslinks that are not found in vertebrate collagens, the tyrosine crosslinks probably being in the form of di and trityrosines [66–68].

The *C. elegans* basement membrane collagens consist of homologues of vertebrate type IV, a heterotrimer of α1(IV) and α2(IV) chains, and a type XVIII homotrimer, $[\alpha1(XVIII)]_3$ [66,71].

The catalytic α subunits of collagen prolyl 4-hydroxylases are encoded by four genes, *phy-1–phy-4* [18,19,72–76]. PDI has two isoforms in *C. elegans*, PDI-1 and PDI-2, both have disulfide isomerase activity [68] and PDI-2 also serves as the β subunit in the prolyl 4-hydroxylase forms that are involved in the synthesis of cuticle collagens (Figure 3). PHY-1 and PHY-2 combine with PDI-2 to form a unique mixed tetramer (Figure 3) that catalyzes the synthesis of the cuticle collagens. Both PHY subunits also form a dimer with PDI-2 (Figure 3), although PHY-2 does this very ineffectively [75]. *Phy-3* is expressed in embryos, late larval stages and adult nematodes but expression in adults is restricted to the spermatheca (a specialized region of the gonad where oocytes are fertilized) [76]. PHY-3 forms an active enzyme only with PDI-1, but its molecular composition is unknown, whereas PHY-4 has not yet been characterized. A further gene, *egl-9* encodes a cytoplasmic HIF prolyl 4-hydroxylase, which is involved in the regulation of the response to hypoxia [26].

Lysyl hydroxylase has only one isoenzyme in *C. elegans* [30,77], which is essential for the synthesis of type IV collagen [77]. Additional enzymes include: subtilisin-like proteases that are required in the processing of the cuticle procollagens [68]; a thioredoxin-like enzyme needed for proper crosslinking [78]; a homologue of the vertebrate PDI-like protein ERp60 with both disulfide isomerase and transglutaminase-like crosslinking activity [79]; and two multidomain dual oxidases with oxidase and peroxidase activity that catalyze the formation of the tyrosine-derived crosslinks [80].

Mutations in *C. elegans* collagens and their modifying enzymes

Many mutations in the cuticle collagen genes affect the body morphology (Table 3), and most of the mutations that have been characterized so far result from a glycine codon being altered to that of another amino acid. Insertion of a mutant collagen into the cuticle often causes a more severe phenotype than null alleles or mutations that lead to a loss or a reduction in the amount of the collagen in the matrix [66–70]. Different mutations in a given gene can produce different phenotypes (Table 3). Homozygous null alleles of the *sqt-1* or *rol-6* gene that result in complete loss of the SQT-1 or ROL-6 collagen, for example, have mild consequences, leading to a defective tail structure or a mild dumpy phenotype, respectively [81]. Mutations in either of these two genes that remove a conserved cysteine in the C-terminal non-collagenous domain eliminate a disulfide bond that is necessary for the formation of a tyrosine-derived cross-link and cause a recessive left-handed roller phenotype, whereas mutations that affect the N propeptide cleavage site and lead to an accumulation of molecules that retain this propeptide produce a recessive dumpy and dominant right-handed roller phenotype [81]. Interestingly, a corresponding mutation that affects the N propeptide cleavage site in the *dpy-10* gene causes a dominant left-handed roller [81]. These mutations are similar to those found in the human *COL1A1* and *COL1A2* genes discussed previously, which prevent cleavage of the N propeptide from type I procollagen and lead to EDS types VIIA and VIIB, respectively. The data currently available indicate a high degree of complexity and redundancy between the *C. elegans* cuticle collagens, their interacting partners and the higher-order structures that they form [69].

Mutations in either of the two type IV basement membrane collagen genes are embryonically lethal, indicating that this collagen is essential for *C. elegans* embryogenesis [82], whereas mutations in the type XVIII collagen gene cause cell and axon migration defects and affect the organization and function of neuromuscular junctions [71,83]. Mutations in the lysyl hydroxylase gene lead to intracellular accumulation of type IV collagen, which is then absent from basement membranes [77], and are therefore also lethal at the embryonic stage (Table 3). This

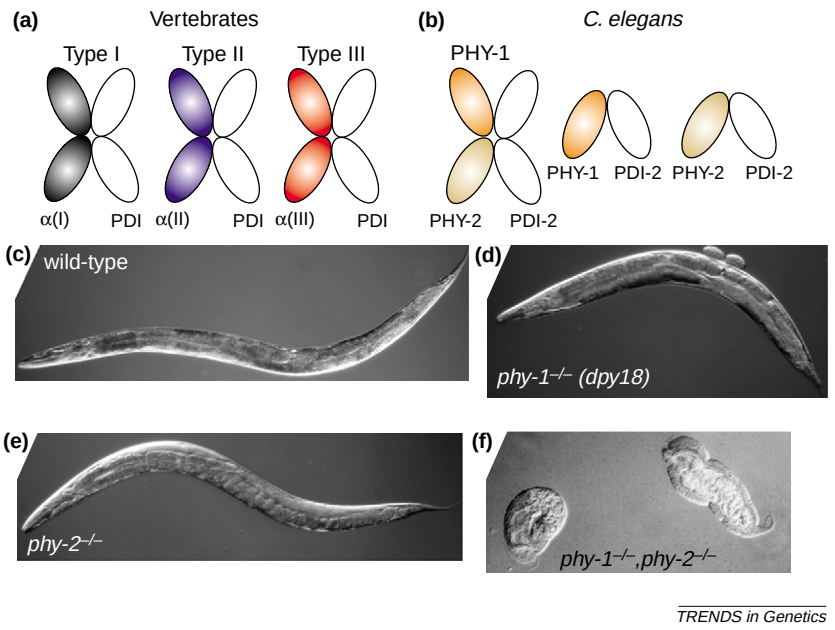

Figure 3. Schematic representation of the forms of collagen prolyl 4-hydroxylase characterized in vertebrates and *C. elegans*, and phenotypes resulting from inactivation of the *Caenohabditis elegans phy-1* and *phy-2* genes. **(a)** The three vertebrate isoenzymes have unique catalytic α subunits and the same β subunits [i.e. the protein disulfide isomerase (PDI) polypeptide] Refs [18,19,24,25]. **(b)** The catalytic α subunits of the prolyl 4-hydroxylase forms that catalyze the synthesis of the cuticle collagens in *C. elegans* are coded by two conserved genes, *phy-1* and *phy-2* [18,19,72–74]. The processed PHY-1 and PHY-2 polypeptides consist of 543 and 523 residues, respectively, PHY-1 being slightly longer than the human α subunits, which are 514–525 residues. PHY-1 and PHY-2 combine with a single β subunit, PDI-2, both in recombinant expression systems and *in vivo* to form a unique mixed tetramer PHY-1–PHY-2–$(PDI\text{-}2)_2$, whereas neither forms a tetramer in the absence of the other [75]. Both also form a PHY–PDI-2 dimer, although the PHY-2–PD1-2 dimer is formed only in small, almost non-detectable amounts [75]. Homozygous inactivation of either the *phy-1* or *phy-2* prevents formation of the mixed tetramer [75]. The null mutant nematodes can in part compensate for the lack of the mixed tetramer by increasing the formation of the corresponding PHY–PDI-2 dimer but the *phy-1* mutants do this ineffectively owing to the very small amount of the PHY-2–PDI-2 dimer that is formed [75] and, therefore, have a dumpy phenotype **(d)**, whereas the *phy-2* null mutants **(e)** are of the wild-type **(c)** [72–75]. The homozygous *phy-1,phy-2* double null [72,74] **(f)**, and the *pdi-2* null mutants (not shown) [74] lack all the forms of prolyl 4-hydroxylase involved in the synthesis of the cuticle collagens [75] and therefore either have a severe dumpy phenotype or are embryonically lethal. Antony Page is gratefully acknowledged for panels c–f. Panel f has been reproduced with permission from Ref. [74].

effect is similar to that seen in the homozygous knockout of the mouse lysyl hydroxylase-3 gene (Table 2).

Homozygous inactivation of either the *phy-1* or *phy-2* gene (Table 3) prevents assembly of the mixed prolyl 4-hydroxylase tetramer (Figure 3). The mutants can in part compensate for its absence by increased assembly of the corresponding PHY–PDI-2 dimer but the *phy-1* mutants do this only very ineffectively [75]. The *phy-1* $^{-/-}$ mutations therefore cause a dumpy phenotype [72–75], whereas *phy-2* $^{-/-}$ mutants have a wild-type phenotype [72,74,75]. The *phy-1* $^{-/-}$,*phy-2* $^{-/-}$ double null [72,74] and *pdi-2* $^{-/-}$ null [74] mutants lack all of the prolyl 4-hydroxylase forms needed for the synthesis of cuticle collagens [75] and are embryonically lethal (Table 3). Similar to vertebrates,

Table 3. Mutations in *Caenorhabditis elegans* collagens and their modifying enzymes

Polypeptide	Gene	Typical phenotype[a]
Cuticle collagen	*dpy-2*, *dpy-3*, *dpy-5*, *dpy-7*, *dpy-8*, *dpy-10*, *dpy-13*	Dumpy
	bli-1, *bli-2*	Blister
	rol-6	Roller or dumpy[b]
	sqt-1,*sqt-3*	Roller or dumpy[b]
	lon-3	Long
Collagen IV: α1(IV); α2(IV)	*emb-9*; *let-2*	Embryonically lethal
Collagen XVIII	*cle-1*	Defects in cell and axon migration and neuromuscular synapse function
P4H[c], PHY-1	*phy-1* (also known as *dpy-18*)	Dumpy
P4H[c], PHY-2	*phy-2*	Wild-type
P4H[c], PHY-1 and PHY-2	*phy-1* and *phy-2*	Severe dumpy or embryonically lethal[d]
P4H[c], PHY-3	*phy-3*	Wild-type
PDI-2[c]	*pdi-2*	Severe dumpy or embryonically lethal[d]
LH	*let-268*	Embryonically lethal
Subtilisin-like protease	*bli-4*	Embryonically lethal or blister
Thioredoxin	*dpy-11*	Dumpy
ERp60	*pdi-3*	Mild disruption of cuticle collagen localization
Duox 1; duox 2	F56C11.1; F53G12.3	Dumpy and blister

[a]Definitions: Dumpy, shortening in the length and thickening of the nematode; Blister, blistering of the cuticle; roller, helical twisting of the nematode body; Long, elongation of the nematode.
[b]Recessive dumpy/dominant or recessive roller.
[c]Abbreviations: Duox, dual oxidase; LH, lysyl hydroxylase; PDI, protein disulfide isomerase; P4H, prolyl 4-hydroxylase.
[d]As observed by RNA interference experiments.

homozygous inactivation of genes for most of the other collagen modifying enzymes also results in severe phenotypes (Table 3), indicating a crucial role for these enzymes in the synthesis of the various *C. elegans* collagens.

Conclusions

It is now well established that collagens and proteins with collagen domains form large superfamilies in many species. The number of family members is constantly growing but research to elucidate the specific properties and functions of the different members has a long way to go. The collagen-modifying enzymes also form large families with multiple isoenzymes, although research into their expression patterns and functions is still in its early stages. Collagen prolyl 4-hydroxylases play a key role in the synthesis of all collagens. An interesting new development has been the identification of a second prolyl 4-hydroxylase family that plays a key role in the regulation of the hypoxia-inducible transcription factor HIF, and it seems likely that other proteins might be found in which 4-hydroxyproline residues play a crucial role. An interesting aspect of the *Drosophila* collagen family is the presence of ~20 putative prolyl 4-hydroxylase isoenzymes, even though the number of collagens in *Drosophila* is much less than in vertebrates and *C. elegans*. Many of these enzymes might therefore be involved in the hydroxylation of proline residues in proteins other than the collagens, and studies on these enzymes might help identify additional functions for vertebrate prolyl 4-hydroxylases. Numerous human collagen mutations have been characterized, but these probably represent only a small fraction of all the existing mutations. It will be important to learn how critical are the roles of collagen mutations as direct causes or predisposing factors in common diseases. The existing mouse and *C. elegans* models for mutations in collagens and their modifying enzymes, and several additional mutants that are likely to be generated or identified, will provide important information on the functions of various members of the superfamily and the effects of mutations in them.

References

1 Kivirikko, K.I. and Prockop, D.J. (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 64, 403–443
2 Kadler, K. (1995) Extracellular matrix 1: fibril-forming collagens. *Protein Profile* 2, 491–619
3 Myllyharju, J. and Kivirikko, K.I. (2001) Collagens and collagen-related diseases. *Ann. Med.* 33, 7–21
4 Kielty, C.M. and Grant, M.E. (2002) The collagen family: structure, assembly, and organization in the extracellular matrix. In *Connective Tissue and Its Heritable Disorders. Molecular, Genetic, and Medical Aspects*, 2nd edn, (Royce, P.M. and Steinmann, B., eds), pp. 159–221, Wiley-Liss
5 Jenkins, C.L. and Raines, R.T. (2002) Insights on the conformational stability of collagen. *Nat. Prod. Rep.* 19, 49–59
6 Franzke, C.F. *et al.* (2003) Collagenous transmembrane proteins: collagen XVII as a prototype. *Matrix Biol.* 22, 299–309
7 Koch, M. *et al.* (2001) α1(XX) collagen, a new member of the collagen subfamily, fibril-associated collagens with interrupted triple helices. *J. Biol. Chem.* 276, 23120–23126
8 Fitzgerald, J. and Bateman, J.F. (2001) A new FACIT of the collagen family: COL21A1. *FEBS Lett.* 505, 275–280
9 Banyard, J. *et al.* (2003) Type XXIII collagen, a new transmembrane collagen identified in metastatic tumor cells. *J. Biol. Chem.* 278, 20989–20994
10 Koch, M. *et al.* (2003) Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone. *J. Biol. Chem.* 278, 43236–43244
11 Hashimoto, T. *et al.* (2002) CLAC: a novel Alzheimer amyloid plaque component derived from a transmembrane precursor, CLAC-P/collagen type XXV. *EMBO J.* 21, 1524–1534
12 Sato, K. *et al.* (2002) Type XXVI collagen, a new member of the collagen family, is specifically expressed in the testis and ovary. *J. Biol. Chem.* 277, 37678–37684
13 Pace, J.M. *et al.* (2003) Identification, characterization and expression analysis of a new fibrillar collagen gene, *COL27A1*. *Matrix Biol.* 22, 3–14
14 Boot-Handford, R.P. *et al.* (2003) A novel and highly conserved collagen [proα1(XXVII)] with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. *J. Biol. Chem.* 278, 31067–31077
15 Holmskov, U. *et al.* (2003) Collectins and ficolins: humoral lectins of the innate immune defense. *Annu. Rev. Immunol.* 21, 547–578
16 Borza, D-B. *et al.* (2001) The NC1 domain of collagen IV encodes a novel network composed of the α1, α2, α5, and α6 chains in smooth muscle basement membranes. *J. Biol. Chem.* 276, 28532–28540
17 Marneros, A.G. and Olsen, B.R. (2001) The role of collagen-derived proteolytic fragments in angiogenesis. *Matrix Biol.* 20, 337–345
18 Kivirikko, K.I. and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv. Enzymol. Relat. Areas Mol. Biol.* 72, 325–398
19 Myllyharju, J. (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* 22, 15–24
20 Prockop, D.J. *et al.* (1998) Two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signaling. *Matrix Biol.* 16, 399–408
21 Kagan, H.M. and Li, W. (2003) Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. *J. Cell. Biochem.* 88, 660–672
22 Hendershot, L.M. and Bulleid, N.J. (2000) Protein-specific chaperones: the role of hsp47 begins to gel. *Curr. Biol.* 10, R912–R915
23 Nagai, N. *et al.* (2000) Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. *J. Cell Biol.* 150, 1499–1506
24 Van Den Diepstraten, C. *et al.* (2003) Cloning of a novel prolyl 4-hydroxylase subunit expressed in the fibrous cap of human atherosclerotic plaque. *Circulation* 108, 508–511
25 Kukkola, L. *et al*. Identification and characterization of a third human, rat and mouse collagen prolyl 4-hydroxylase isoenzyme. *J. Biol. Chem.* (in press).
26 Epstein, A.C.R. *et al.* (2001) *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107, 43–54
27 Bruick, R.K. and McKnight, S.L. (2001) A conserved family of prolyl 4-hydroxylases that modify HIF. *Science* 294, 1337–1340
28 Ivan, M. *et al.* (2002) Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13459–13464
29 Hirsilä, M. *et al.* (2003) Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor HIF. *J. Biol. Chem.* 278, 30772–30780
30 Passoja, K. *et al.* (1998) Cloning and characterization of a third human lysyl hydroxylase isoform. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10482–10486
31 Valtavaara, M. *et al.* (1998) Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl hydroxylase 3). *J. Biol. Chem.* 273, 12881–12886
32 Mäki, J.M. *et al.* (2001) Cloning and characterization of a fifth human lysyl oxidase isoenzyme: the third member of the lysyl oxidase-related subfamily with four scavenger receptor cysteine-rich domains. *Matrix Biol.* 20, 493–496
33 Ito, H. *et al.* (2001) Molecular cloning and biological activity of a novel lysyl oxidase-related gene expressed in cartilage. *J. Biol. Chem.* 276, 24023–24029

34 Leighton, M. and Kadler, K.E. (2003) Paired basic/furin-like proprotein convertase cleavage of pro-BMP-1 in the trans-Golgi network. *J. Biol. Chem.* 278, 18478–18484
35 Colige, A. *et al.* (2002) Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying high homology with ADAMTS-2 and ADAMTS-3. *J. Biol. Chem.* 277, 5756–5766
36 Wang, C. *et al.* (2002) Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). *J. Biol. Chem.* 277, 18568–18573
37 Rautavuoma, K. *et al.* (2002) Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1–3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. *J. Biol. Chem.* 277, 23084–23091
38 Dalgleish, R. (1997) The human type I collagen mutation database. *Nucleic Acids Res.* 25, 181–187
39 Krawczak, M. and Cooper, D.N. (1997) The human gene mutation database. *Trends Genet.* 13, 121–122
40 Byers, P.H. and Cole, W.G. (2002) Osteogenesis imperfecta. In *Connective Tissue and Its Heritable Disorders. Molecular, Genetic, and Medical Aspects*, 2nd edn, (Royce, P.M. and Steinmann, B., eds), pp. 385–430, Wiley-Liss
41 Steinmann, B. *et al.* (2002) The Ehlers-Danlos syndrome. In *Connective Tissue and Its Heritable Disorders. Molecular, Genetic, and Medical Aspects*, 2nd edn, (Royce, P.M. and Steinmann, B., eds), pp. 431–523, Wiley-Liss
42 Horton, W.A. and Hecht, J.T. (2002) Chondrodysplasias: disorders of cartilage matrix proteins. In *Connective Tissue and Its Heritable Disorders. Molecular, Genetic, and Medical Aspects*, 2nd edn, (Royce, P.M. and Steinmann, B., eds), pp. 909–937, Wiley-Liss
43 Tryggvason, K. and Martin, P. (2002) Alport syndrome. In *Connective Tissue and Its Heritable Disorders. Molecular, Genetic, and Medical Aspects*, 2nd edn, (Royce, P.M. and Steinmann, B., eds), pp. 1069–1102, Wiley-Liss
44 Bruckner-Tudermann, L. (2002) Epidermolysis bullosa. In *Connective Tissue and Its Heritable Disorders. Molecular, Genetic, and Medical Aspects*, 2nd edn, (Royce, P.M. and Steinmann, B., eds), pp. 687–725, Wiley-Liss
45 Pan, T.C. *et al.* (2003) New molecular mechanism for Ullrich congenital muscular dystrophy: a heterozygous in-frame deletion in the *COL6A1* gene causes a severe phenotype. *Am. J. Hum. Genet.* 73, 355–369
46 Biswas, S. *et al.* (2001) Missense mutations in *COL8A2*, the gene encoding the α2 chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Hum. Mol. Genet.* 10, 2415–2423
47 Suzuki, O.T. *et al.* (2002) Molecular analysis of collagen XVIII reveals novel mutations, presence of a third isoform, and possible genetic heterogeneity in Knobloch syndrome. *Am. J. Hum. Genet.* 71, 1320–1329
48 Kramer, R.Z. *et al.* (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* 6, 454–457
49 Annunen, S. *et al.* (1999) An allele of *COL9A2* associated with intervertebral disc disease. *Science* 285, 409–412
50 Paassilta, P. *et al.* (2001) Identification of a novel common genetic risk factor for lumbar disk disease. *J.A.M.A.* 285, 1843–1849
51 Gustafsson, E. and Fässler, R. (2000) Insights into extracellular matrix functions from mutant mouse models. *Exp. Cell Res.* 261, 52–68
52 Reichenberger, E. *et al.* (2000) Collagen XII mutation disrupts matrix structure of periodontal ligament and skin. *J. Dent. Res.* 79, 1962–1968
53 Sund, M. *et al.* (2001) Abnormal adherence junctions in the heart and reduced angiogenesis in transgenic mice overexpressing mutant type XIII collagen. *EMBO J.* 20, 5153–5164
54 Kvist, A-P. *et al.* (2001) Lack of cytosolic and transmembrane domains of type XIII collagen results in progressive myopathy. *Am. J. Pathol.* 159, 1581–1592
55 Eklund, L. *et al.* (2001) Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1194–1199
56 Fukai, N. *et al.* (2002) Lack of collagen XVIII/endostatin results in eye abnormalities. *EMBO J.* 21, 1535–1544
57 Van der Slot, A.J. *et al.* (2003) Identification of *PLOD2* as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *J. Biol. Chem.* 278, 40967–40972
58 Li, S-W. *et al.* (2001) Transgenic mice with inactive alleles for procollagen N-proteinase (ADAMTS-2) develop fragile skin and male sterility. *Biochem. J.* 355, 271–278
59 Clark, T.G. *et al.* (1999) The mammalian Tolloid-like 1 gene, *Tll1*, is necessary for normal septation and positioning of the heart. *Development* 126, 2631–2642
60 Mäki, J.M. *et al.* (2002) Inactivation of the lysyl oxidase gene *Lox* leads to aortic aneurysms, cardiovascular dysfunction, and perinatal death in mice. *Circulation* 106, 2503–2509
61 Hornstra, I.K. *et al.* (2003) Lysyl oxidase is required for vascular and diaphragmatic development in mice. *J. Biol. Chem.* 278, 14387–14393
62 Hynes, R.O. and Zhao, Q. (2000) The evolution of cell adhesion. *J. Cell Biol.* 150, F89–F95
63 Chartier, A. *et al.* (2002) Pericardin, a *Drosophila* type IV collagen-like protein is involved in the morphogenesis and maintenance of the heart epithelium during dorsal ectoderm closure. *Development* 129, 3241–3253
64 Abrams, E.W. and Andrew, D.J. (2002) Prolyl 4-hydroxylase α-related proteins in *Drosophila melanogaster*: tissue-specific embryonic expression of the 99F8-9 cluster. *Mech. Dev.* 112, 165–171
65 Annunen, P. *et al.* (1999) Cloning of the α subunit of prolyl 4-hydroxylase from *Drosophila* and expression and characterization of the corresponding enzyme tetramer with some unique properties. *J. Biol. Chem.* 274, 6790–6796
66 Kramer, J.M. (1997) Extracellular matrix. In *C. elegans II* (Riddle, D.L. *et al.*, eds), pp. 471–500, Cold Spring Harbor Laboratory Press
67 Johnstone, I.L. (2000) Cuticle collagen genes. Expression in *Caenorhabditis elegans*. *Trends Genet.* 16, 21–27
68 Page, A.P. (2001) The nematode cuticle: synthesis, modification and mutants. In *Parasitic Nematodes* (Kennedy, M.W. and Harnett, W., eds), pp. 167–193, CABI Press
69 Thein, M.C. *et al.* (2003) *Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. *Dev. Dyn.* 226, 523–539
70 McMahon, L. *et al.* (2003) Two sets of interacting collagens form functionally distinct substructures within a *Caenorhabditis elegans* extracellular matrix. *Mol. Biol. Cell* 14, 1366–1378
71 Ackley, B.D. *et al.* (2001) The NC1/endostatin domain of *Caenorhabditis elegans* type XVIII collagen affects cell migration and axon guidance. *J. Cell Biol.* 152, 1219–1232
72 Friedman, L. *et al.* (2000) Prolyl 4-hydroxylase is required for viability and morphogenesis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4736–4741
73 Hill, K.L. *et al.* (2000) *dpy-18* encodes an α-subunit of prolyl 4-hydroxylase in *Caenorhabditis elegans*. *Genetics* 155, 1139–1148
74 Winter, A.D. and Page, A.P. (2000) Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode *Caenorhabditis elegans*. *Mol. Cell. Biol.* 20, 4084–4093
75 Myllyharju, J. *et al.* (2002) The exoskeleton collagens in *Caenorhabditis elegans* are modified by prolyl 4-hydroxylases with unique combinations of subunits. *J. Biol. Chem.* 277, 29187–29196
76 Riihimaa, P. *et al.* (2002) Egg shell collagen formation in *Caenorhabditis elegans* involves a novel prolyl 4-hydroxylase expressed in spermatheca and embryos and possessing many unique properties. *J. Biol. Chem.* 277, 18238–18243
77 Norman, K.R. and Moerman, D.G. (2000) The *let-268* locus of *Caenorhabditis elegans* encodes a procollagen lysyl hydroxylase that is essential for type IV collagen secretion. *Dev. Biol.* 227, 690–705
78 Ko, F.C. and Chow, K.L. (2002) A novel thioredoxin-like protein encoded by the *C. elegans dpy-11* gene is required for body and sensory organ morphogenesis. *Development* 129, 1185–1194
79 Eschenlauer, C.P. and Page, A.P. (2003) The *Caenorhabditis elegans* ERp60 homolog protein disulfide isomerase-3 has disulfide isomerase and transglutaminase-like cross-linking activity and is involved in the maintenance of body morphology. *J. Biol. Chem.* 278, 4227–4237
80 Edens, W.A. *et al.* (2001) Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peroxidase with homology to the phagocyte oxidase subunit gp91*phox*. *J. Cell Biol.* 154, 879–891

81 Yang, J. and Kramer, J.M. (1999) Proteolytic processing of *Caenorhabditis elegans* SQT-1 cuticle collagen is inhibited in right roller mutants whereas cross-linking is inhibited in left-roller mutants. *J. Biol. Chem.* 274, 32744–32749
82 Gupta, M.C. *et al.* (1997) Characterization of α1(IV) collagen mutations in *Caenorhabditis elegans* and the effects of α1 and α2(IV) mutations on type IV collagen distribution. *J. Cell Biol.* 137, 1185–1196
83 Ackley, B.D. *et al.* (2003) The basement membrane components nidogen and type XVIII collagen regulate organization of neuromuscular junctions in *Caenorhabditis elegans*. *J. Neurosci.* 23, 3577–3587