THE ANATOMICAL RECORD 268:302–316 (2002)

Evolution of Collagens

JEAN-YVES EXPOSITO,* CAROLINE CLUZEL, ROBERT GARRONE, AND CLAIRE LETHIAS
Institut de Biologie et Chimie des Protéines, Université Claude Bernard, Lyon, France

ABSTRACT

The extracellular matrix is often defined as the substance that gives multicellular organisms (from plants to vertebrates) their structural integrity, and is intimately involved in their development. Although the general functions of extracellular matrices are comparable, their compositions are quite distinct. One of the specific components of metazoan extracellular matrices is collagen, which is present in organisms ranging from sponges to humans. By comparing data obtained in diploblastic, protostomic, and deuterostomic animals, we have attempted to trace the evolution of collagens and collagen-like proteins. Moreover, the collagen story is closely involved with the emergence and evolution of metazoa. The collagen triple helix is one of numerous modules that arose during the metazoan radiation which permit the formation of large multimodular proteins. One of the advantages of this module is its involvement in oligomerization, in which it acts as a structural organizer that is not only relatively resistant to proteases but also permits the creation of multivalent supramolecular networks. Anat Rec 268:302–316, 2002.

Key words: collagen; extracellular matrix; evolution

In multicellular animals, a complex of macromolecules acts as a glue to strengthen the link between cells, thereby providing structural integrity in tissues and, ultimately, in animals. This glue is the extracellular matrix (ECM), which is a highly organized network that results from the association of glycoproteins, proteoglycans, and glycosaminoglycans. In addition to its physical function, the ECM acts as a substrate for cell growth and is involved in different cell behaviors such as migration, differentiation, adhesion, and spreading. For each tissue, the ECM has a specific composition that results from its intimate association with surrounding cells. Cells govern the production of ECM, but interactions between ECM components and membrane receptors activate specific cellular signaling pathways involved in the control of gene expression.

While the functional biological importance of animal ECM in developmental processes and in tissue formation is well documented, plant ECM is often thought of as a substance involved only in providing mechanical strength. Although plant ECM has a composition quite different from that of animals, recent advances have shown that it also plays an important role during development (Anderson et al., 2001; Kohorn, 2001).

Bacteria are perceived as unicellular life forms, but in nature they often occur attached to surfaces within a structured biofilm ecosystem (Davey and O'Toole, 2000; Sutherland, 2001). Biofilms are composed mostly of water, and in addition to microbial cells they contain a complex of secreted polymers, nutrients, metabolites, and products from cell lysis and the surrounding environment. Microbial biofilms can be thought of as interactive organisms, and present certain ecological advantages, such as the development of fruiting bodies and spores.

Therefore, from prokaryotes to higher vertebrates, an ECM or its equivalent can be defined, but the composition of the ECM, its relation with cells, and the corresponding cellular signaling phenomena are quite distinct.

Collagen Superfamily

One of the major components of animal ECM is collagen. In humans, 21 types of collagen (I–XXI) have been described (Myllyharju and Kivirikko, 2001; Koch et al., 2001; Fitzgerald and Bateman, 2001). All collagen molecules are made of three α chains, which can either be identical

*Correspondence to: Jean-Yves Exposito, Institut de Biologie et Chimie des Protéines, CNRS UMR 5086, Université Claude Bernard, 7 passage du Vercors, 69367 Lyon Cedex 07, France. Fax: +33-4-72-72-26-04. E-mail: jy.exposito@ibcp.fr

Received 14 December 2001; Accepted 28 February 2002
DOI 10.1002/ar.10162

(homotrimer) or result from a combination of two or three genetically distinct α chains (heterotrimer). Each α chain contains at least one collagenous domain consisting of a repeating Gly-Xaa-Yaa triplet, in which X and Y are often proline and hydroxyproline, respectively. During molecular assembly, the collagenous domains of the three α chains adopt a triple-helical structure. Other proteins, such as the complement subcomponent C1q, acetyl cholinesterase (Q subunit), EMILINS, ficolins, macrophage scavenger receptors, collectins, ectodysplasin, and the adipose-specific factor apM1, possess collagenous domains (Myllyharju and Kivirikko, 2001). For this reason, collagens have the potential to form supramolecular structures or to participate in the fibrillar assemblies of the ECM (fibrils, networks, and dimers), either alone or in association with other ECM components (van der Rest and Garrone, 1991).

Collagens can be divided into several subfamilies according to their primary structures and/or forms of supramolecular organization. Fibrillar collagens (types I–III, V, and XI) present the same general structure and participate in the formation of cross-striated fibrils. In mature fibrillar collagen molecules, the collagenous domain of each α chain contains a perfect series of Gly-Xaa-Yaa triplets. Other subfamilies correspond to the so-called nonfibrillar collagens, which contain interruptions in their collagenous domains. Nonfibrillar subfamilies are: 1) basement membrane or type IV collagen, which is a network-forming collagen; 2) type VI, which forms beaded filaments; 3) type VII, which forms anchoring fibrils involved in the attachment of the basement membrane to the underlying ECM; 4) short-chain collagens, types VIII, and X, which are hexagonal network-forming collagens; 5) fibril-associated collagens with interrupted triple helices, FACIT collagens, which include types IX, XII, and XIV, and the newly characterized types XX and XXI; 6) FACIT-related collagens, types XVI and XIX; 7) plasma membrane collagens, including types XIII and XVII; and 8) multiplexins (types XV and XVIII), for multiple triple-helical domains and interruptions. The functional importance of collagens has been shown in several human genetic diseases related to mutations in their coding genes or genes coding for proteins involved in their maturation. A summary of these human diseases was presented in the recent review by Myllyharju and Kivirikko (2001).

The purpose of the present review is to summarize what we have learned about collagen evolution from invertebrate collagen data, and to consider whether collagens are peculiar to metazoans. First, we present data obtained in the first multicellular animal, the sponge. This "primitive" diploblastic animal has the same adaptive features found in multicellular animals, i.e., a peripheral layer of cells connected via stable junctions and internal cells embedded within an ECM. Electron microscopic analysis of this organism reveals two categories of collagens: 1) cross-striated fibrils of 20–25 nm diameter, and 2) thin (10 nm) spongin filaments that attach sponges to their substrata, and also, in some sponges, act as a cement between the spicules that comprise the skeleton (Garrone, 1978). Spongin filaments can also totally replace the inorganic skeleton, such as in the well-known bath sponges.

Fibrillar Collagens

Among the vertebrate fibrillar collagens, type I represents the most abundant protein of the body. It has important functions, for example, in the mechanical properties of bones, tendons, and skin, as emphasized in human diseases involving mutations of the genes coding for the proα1(I) and proα2(I) chains (for recent reviews see Burrows, 1999; Giguère and Rousseau, 2000; Primorac et al., 2001). In vertebrates, all procollagen fibrillar α chains consist of a major or central uninterrupted triple helix made up of approximately 338 Gly-Xaa-Yaa triplets (Fig. 1). This region is flanked by two noncollagenous regions, the N-propeptide, which contains a minor triple helix at its C-terminus, and the C-propeptide. The N- and C-propeptides are connected to the major triple helix via two short extensions, the N- and C-telopeptides, whose extremities generally define the N- and C-proteinase cleavage sites, respectively. During the maturation of procollagen molecules, the N- and C-propeptides are generally removed by the action of these proteinases. This leads to the formation of collagen molecules consisting of the major triple helix flanked by the telopeptides. The next step is the supramolecular assembly of collagen fibrils. Depending on the tissues analyzed and the collagen types involved, fibrils have either a unimodal- or a multimodal-diameter distribution. An illustration of a fibrillar collagen molecule and the molecular arrangement within the fibrils is presented in Figure 2.

Genetic units at the origin of fibrillar collagen genes. The common structure of the fibrillar collagen α chains is also observed at the genomic level, where the major triple helix is encoded by exons of 54 bp, multiples of 54 bp, or multiples of 54 – 9 bp. All of these exons begin with an intact glycine codon and end with an intact Yaa codon. For this reason, Yamada et al. (1980) suggested that an ancestral fibrillar collagen gene arose from multiple duplications of a primordial unit, including an exon of 54 bp coding for six Gly-Xaa-Yaa triplets. Unequal crossing-over could explain the presence of 45 and 99 bp exons. By genomic and cDNA analyses we have been able to characterize most of the primary structure of a sponge fibrillar collagen chain, which has the same overall structure as vertebrate fibrillar collagen chains (Exposito and Garrone, 1990; Exposito et al., 1993). Some unique features of its primary structure are discussed below. At the gene level, the most interesting feature of this sponge collagen is the major triple-helical region. As in vertebrates, exons are of 54 bp, multiples of 54 bp (108, 162, and 216), or multiples of 54 – 9 bp (45, 99, 153, and two related exons of 144 and 414 bp), and they begin with an intact glycine codon and end with an intact Yaa codon. The exon structure of regions encoding the major triple helix of sponge and vertebrate types I–III α chains has been aligned, and by taking into account all exon/intron junctions it has been possible to build the putative structure of an ancestral fibrillar collagen gene (Exposito et al., 1993). As suggested by Yamada et al. (1980), an ancestral fibrillar collagen gene probably arose first by multiple duplications of a primordial exon of 54 bp. After an unequal crossing-over, a new modular unit was generated, including a 54 bp exon and a 45 bp exon. Multiple duplications of these two genetic units could explain the structure of a putative ancestral fibrillar gene (Fig. 3A). More recently, the exon-intron organizations of *COL5A1* and *COL11A2* were determined (Takahara et al., 1995; Vuoristo et al., 1995) (*COLxAy* refers to the human gene encoding the pro$\alpha y(x)$ collagen chain), and these match the

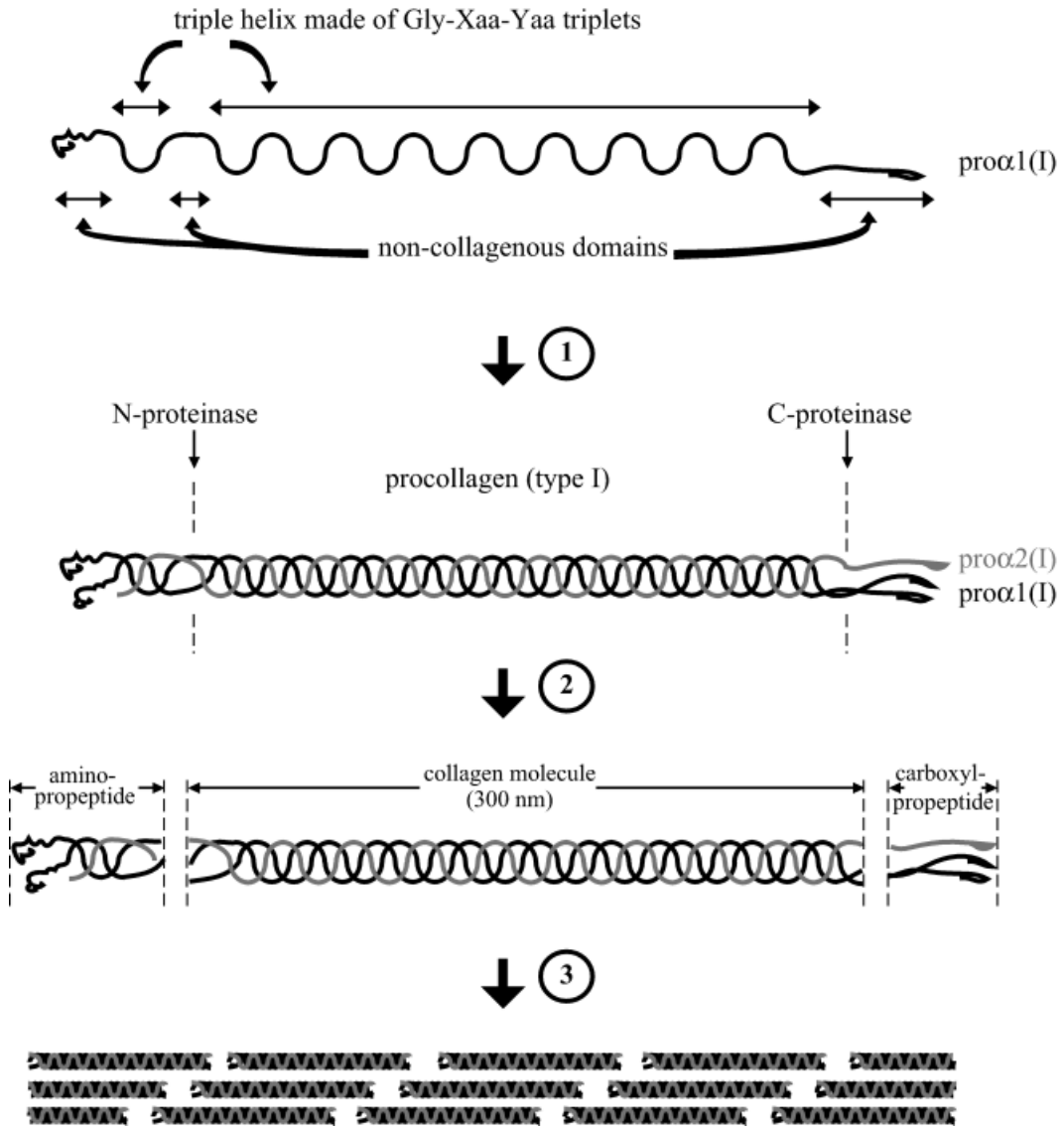

Fig. 1. Fibrillar collagens. Type I collagen is used to illustrate the different steps leading to the formation of fibrils from collagen chains. Step 1 consists of the formation of the procollagen molecule. Its maturation into a collagen molecule corresponds to step 2. Collagen molecules are involved in the supramolecular organization of collagen fibrils (step 3).

structure proposed for an ancestral fibrillar collagen gene. The relative conservation of the exon-intron organization of the region encoding the major triple helix of metazoan fibrillar α chains argues for a strong selection pressure preserving both the size and the perfect Gly-Xaa-Yaa repeat necessary for the formation of fibrils.

Adaptability features and diversity of fibrillar α chains. Although the structure of the sponge fibrillar collagen chain is classical, its primary structure reveals the presence of two successive glycine substitutions in the Gly-Xaa-Yaa repeat and one Gly-Xaa-Yaa-Zaa imperfection (Exposito and Garrone, 1990). In humans, glycine substitutions generate syndromes characterized by moderate to severe effects (Kuivaniemi et al., 1997). Glycine substitutions have also been detected in the fibrillar collagen chains of worms and abalone (Sicot et al., 1997; Yoneda et al., 1999). Because of the low temperature at which these animals live, however, it may be that the presence of these imperfections in the triple helix does not affect molecular stability and fibrillar integrity. During the last decade, several studies have characterized the primary structures of fibrillar collagen chains in hydra (Deutzmann et al., 2000), worms (Sicot et al., 1997, 2000), abalone (Yoneda et al., 1999), and sea urchin (Exposito et al., 1992a, b). These data confirm that the length of the major triple helix is conserved in both invertebrates and vertebrates, despite the presence of some imperfections and low levels of sequence identity. The major findings in this domain resulted from comparative studies of data obtained in coastal or deep-sea marine worms (Sicot et al.,

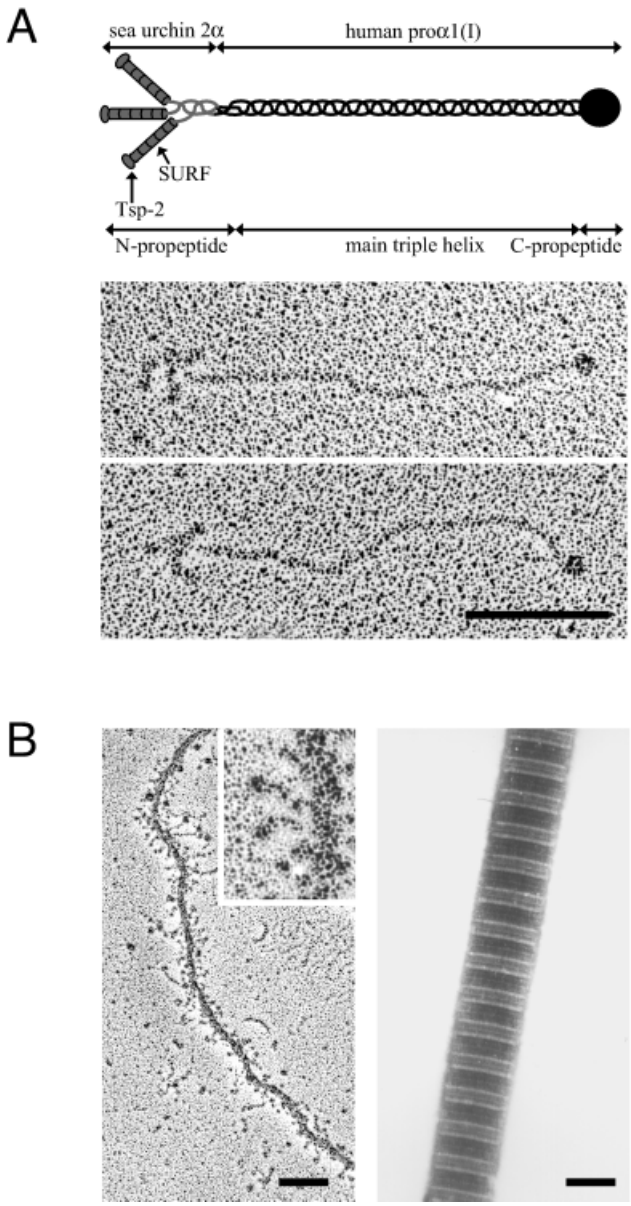

Fig. 2. Electron microscopy of sea urchin molecules and fibrils. **A:** Hybrid human-sea urchin recombinant fibrillar molecules visualized after rotary shadowing. A schematic representation of this recombinant molecule is shown above the rotary shadowing images. **B:** Thin embryo fibril observed after rotary shadowing. Insert: Magnification of a region of the fibril. Sea urchin adult fibril visualized after negative staining. Bar: 100 nm.

2000). These show that the α chain of a deep-sea thermophilic worm contains more proline residues and stabilizing triplets than are found in the temperate deep-sea or coastal worms. This is a good example of adaptation to the ambient surrounding temperature. In contrast, the central triple helix of hydra fibrillar α chains is relatively poor in proline residues (15% compared to 23% in vertebrate α chains), and this feature of hydra collagen may be correlated with the flexible properties of its ECM (Deutzmann et al., 2000). In addition, several studies have indicated that Gly-Xaa-Gly and Gly-Gly-Yaa triplets are more abundant in invertebrate and lower vertebrate α chains than in their mammalian counterparts (Su et al., 1991; Exposito et al., 1992a; Yoneda et al., 1999; Saito et al., 2001). These triplets have been described as destabilizing triplets (Shah et al., 1997), and are also found in vertebrate collagens. Hence, in mammalian type III collagen, 5.6% of triplets include two glycine residues compared to 1–1.2% in type I or II collagens. Shah et al. (1997) indicated that these unstable Gly-Gly-containing triplets might confer unstable or flexible properties to some regions of type III collagen molecules. It is worth noting that in type I–III collagens, the cleavage site for mammalian collagenase is near a flexible region characterized by a low amino acid content (Fields, 1991). The best example of Gly-Gly-containing triplets is probably that of rainbow trout type I collagen (Saito et al., 2001). As indicated by Saito et al. (2001), the three α chains making up type I collagen are rich in Gly-Gly pairs (11% for the α3(I) chain), and this situation may be correlated with the low thermal stability of type I collagen from this poikilothermic animal. Hence, depending on the environment and the ability to regulate body temperature, distinct selective features have been used during evolution.

Concerning the C-propeptide domain, invertebrate data confirm that this region remains the most conserved within the fibrillar collagens; this may be correlated with its function in the recognition and assembly of α chains into trimeric molecules. Several phylogenic studies using this domain have been performed, and the major result is that vertebrate fibrillar collagen chains can be divided into two groups: the A group, including types I–III and the proα2(V) chains; and the B group, including the proα1(V), proα1(XI), and proα2(XI) chains (Sicot et al., 1997; Saito et al., 2001; Välkkilä et al., 2001), and probably also the proα3(V) chain. This subdivision reflects the divergence of the exon-intron organization observed for the genes encoding these two groups (Takahara et al., 1995; Vuoristo et al., 1995).

Invertebrate data confirm that the N-propeptide domain is the most variable region within the fibrillar collagen chains. In vertebrates, three N-propeptide structures have been defined, while a fourth structure has been characterized in sea urchin (Exposito et al., 1992b). All possess a minor triple helix, which in some cases is interrupted. The noncollagenous parts of these domains are built of modules that, in sea urchin, consist of a 140 amino acid long motif (sea urchin fibrillar module (SURF), which appears to be specific to the echinoderm phylum (Cluzel et al., 2001). Previously, we observed that all α chains involved in the formation of thin fibrils present a short interruption in the minor triple helix (Cluzel et al., 2000). Our hypothesis is that this region plays an important function in the regulation of fibril diameter, although other parameters (such as association with proteoglycans, the sequence of the main triple helix, retention of the N-propeptide, and composition of heterotypic fibrils) also govern fibril diameter (Birk and Silver, 1984; Vogel and Trotter, 1987; Linsenmayer et al., 1993). In sponges, all fibrils present a uniform diameter of 20–25 nm, a situation observed in sea urchins during embryogenesis (Lethias et al., 1997). Figure 2B shows a sea urchin fibril harboring, at its surface, periodically-spaced N-propeptide extensions. These N-propeptide extensions have also been observed in chimeric sea urchin/human fibrillar molecules

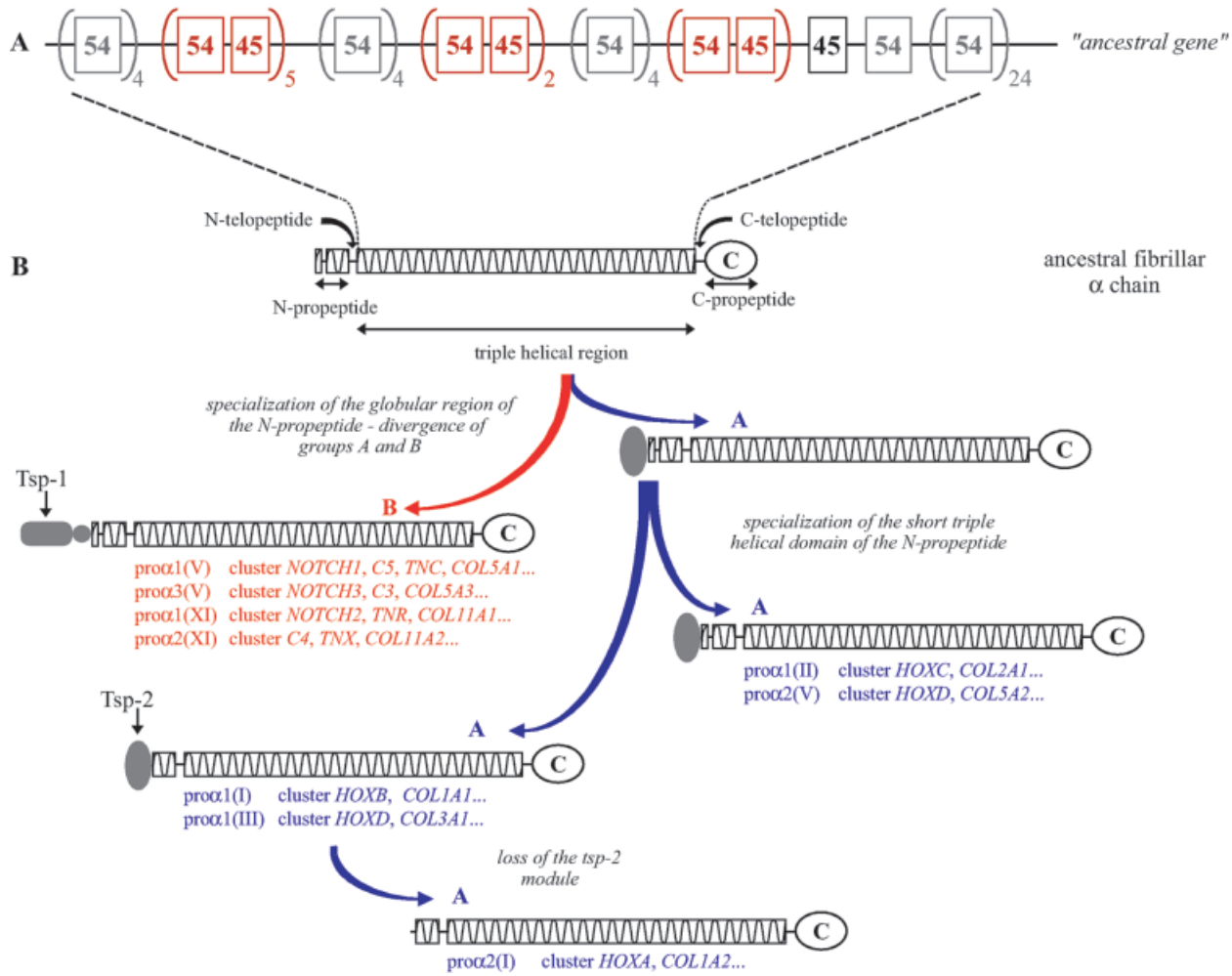

Fig. 3. Evolution of fibrillar collagens. **A:** Hypothetical structure of an ancestral gene encoding the major triple helix of a fibrillar α chain. This gene is constructed of two genetic units: one containing an exon of 54 bp, and one with two exons (one of 54 bp and one of 45 bp). **B:** Hypothetical steps leading to the formation of vertebrate fibrillar collagens from an ancestral α chain. Vertebrate fibrillar α chains previously were subdivided into two groups (Sicot et al., 1997). The A group includes types I–III and the proα2(V) chains, whereas the proα1(V), proα3(V), proα1(XI), and proα2(XI) chains represent the B group. Genes encoding members of the A group are linked to human homologous genomic segments, including the *Hox* cluster (Bailey et al., 1997). Genes encoding members of the B group are linked to the human homologous gene cluster segments. including *NOTCH*, tenascin (*TN*), and complement (**C**) genes (Smith et al., 1999). The human genome can be accessed via the BLAST site server (http://www.ncbi.nlm.nih.gov/BLAST/).

(Cluzel et al., 2000) (Fig. 2A). A peculiar situation has been observed in hydra (Deutzmann et al., 2000), wherein the N-propeptide of the fibrillar chain that has been characterized is related to that of the vertebrate proα2(I) chain, but is not removed during procollagen maturation as in vertebrate A group collagens. Moreover, its N-telopeptide (eight amino acids) is shorter than those of vertebrate fibrillar α chains. Homotrimeric molecules of this chain assemble as a network of thin fibrils rather than regular striated fibrils. As noted by Deutzmann et al. (2000), in the human-type VII-Ehlers-Danlos syndrome, the lack of processing of the N-propeptide in type I molecules alters the formation of fibrils, which become thin and irregular, resulting in joint and skin abnormalities. Hydra ECM is a highly elastic matrix, and the structural features of hydra fibrillar collagens partially mimic the physical state of the matrix of this Ehlers-Danlos syndrome (Sarras and Deutzmann, 2001). As indicated above, the low level of proline residues in the main triple helix of hydra collagen also supports the concept of a more flexible molecule.

What Is the Structure of an Ancestral Fibrillar α Chain?

Several studies have indicated that invertebrate collagens are related to vertebrate types V/XI collagens (Miura and Kimura, 1985; Exposito and Garrone, 1990; Tillet et al., 1996). Interestingly, the proposed structure of the region encoding the triple helix of an ancestral α chain is closely related to that encoding the type V and XI chains. In the most primitive animals, sponges, fibrils have uniform diameters of 25 nm, and like the α chains involved in the formation of thin fibrils, the minor triple helix of its N-propeptide presents a short interruption in the Gly-

Xaa-Yaa repeat. For these reasons, we have previously suggested that an ancestral chain might have contained the major triple helix, a C-propeptide, and an interrupted minor triple helix corresponding to the N-propeptide (Exposito et al., 2000). As presented in Figure 3B, different putative steps may have led to the formation of vertebrate α chains. In respect to this model, the initial structural function of the fibrillar collagen chain was to form thin fibrils, as observed with collagen in sponge, the first multicellular organism, and with type V and XI collagens in vertebrates. Later on, the incorporation of type I-like collagen into fibrils may have led to the formation of a large spectrum of fibril shapes.

In the future, the availability of new data will improve our knowledge of fibrillar collagen evolution. One example of such evolution is the sea urchin, in which fibrils are thin and unimodal in the embryo, and thicker and multimodal in adults. In addition to the maturation of the N-propeptide in the 2α chain in adults, we have preliminary evidence to suggest that these fibrils are built of four distinct α chains: the two previously biochemically characterized 1α and 2α chains (Trotter and Koob, 1994), and two new chains (Cluzel et al., unpublished data). Hence, heterotypic fibrils are already present in an invertebrate deuterostome. The final question relates to the lack of fibrillar collagens in some invertebrates, such as *Drosophila melanogaster* and *Caenorhabditis elegans.* Hynes and Zhao (2000) suggested that flies do not need the mechanical resistance provided by fibrillar collagens because they have a chitinous exoskeleton. The same hypothesis can be proposed for the cuticular exoskeleton of nematodes.

Spongin: A Collagen Family at the Origin of Vertebrate Nonfibrillar Collagens

In the fresh-water sponge *Ephydatia mülleri,* we have characterized two spongin collagens by cDNA cloning (Exposito et al., 1990, 1991) (Qin and Garrone, unpublished data). A schematic structure of these sponge collagens is presented in Figure 4. The two spongins (EmC8 and EmC13) possess two uninterrupted collagenous domains COL2 and COL1 of 66 and 171 amino acids, respectively. At the C-terminus, the noncollagenous NC1 domain is 155–156 amino acids long, including nine cysteine residues. Between the two COL domains, a short NC2 domain of 13 residues is present. These four domains are 83–100% identical in the two chains. In contrast, the N-terminal NC3 region is more distinct, both in size and in terms of sequence identity (37%). By Southern-blot analysis of the sponge genome, we have previously shown that at least 10 genes encode spongin-related proteins that are highly conserved in sequence (Exposito et al., 1991). Recent studies of a marine sponge have permitted the characterization of two related proteins whose sequences and structures are less conserved. One of these consists of a NC1 spongin-like domain, and the second possesses a similar C-terminal NC1 region as well as a collagenous domain (Schröder et al., 2000; Krasko et al., 2000).

Two striking features have been observed in spongins. First, the spongin NC2 domain is closely related to the comparable region of nematode cuticular collagens (Fig. 4). The greatest similarity was observed with group 3 cuticular collagens (Johnstone, 2000), with three perfectly aligned cysteine residues, while the other amino acids are often proline residues. This similarity between spongin and cuticular collagens is also observed at the C-terminus of the NC3 domain (cysteine residues). Moreover, these collagens are encoded by a large family of genes, and might appear as exocollagens because of their location. The second feature is the similarity observed between the spongin NC1 region and the NC1 domain of type IV collagen (Fig. 4). While the sequence identity is poor between these collagens, except for short stretches, the NC1 region of spongin can be divided, like type IV NC1 (Pihlajaniemi et al., 1985), into two subdomains: NC1-A and NC1-B, presenting 27% identity.

What is the significance of this similarity between spongin and type IV collagens? Basement membranes have been described in all multicellular organisms except the first, the sponge. However, a basement membrane-like structure has been observed in a group of marine sponges, Homoscleromorpha. A good marker of basement membranes is type IV collagen, which forms a network-structure via its amino- and carboxyl-terminal domains, and whose supramolecular organization provides the scaffold for other basement membrane components. In one Homoscleromorpha sponge, *Pseudocorticium jarrei,* two type IV collagen chains have been characterized by cDNA cloning (Boute et al., 1996). This group of sponges is considered to be a more evolved branch of the phylum, and, unlike the others, it possesses spermatozoa with acrosomes. In another diploblastic phylum, Cnidarian, the subepithelial zones of the mesoglae resemble vertebrate basal laminae, and a type IV collagen chain has been characterized in hydra (Fowler et al., 2000; Sarras and Deutzmann, 2001). In vertebrates, six type IV chains have been characterized and the corresponding genes are arranged pairwise (*COL4A1-COL4A2, COL4A3-COL4A4,* and *COL4A5-COL4A6*) in a head-to-head fashion on three different chromosomes (Hudson et al., 1993). This compares with the two ecdysozoan protostomes (drosophila and nematode) in which two type IV collagen genes have been characterized. The drosophila type IV collagen genes are in a head-to-head arrangement (Yasothornsrikul et al., 1997), as they are in vertebrates, whereas the two nematode type IV genes are located on two different chromosomes (Guo and Kramer, 1989). From these data, the common ancestor of protostomes and deuterostomes, known as Urbilateria (DeRobertis and Sasai, 1996), might possess at least two type IV collagen chains, and their related genes might be associated in a head-to-head fashion. The strong conservation of basement membrane collagens in all metazoan phyla emphasizes their essential function. It will be interesting to look for type IV collagen in sponges other than those belonging to the Homoscleromorpha group to determine if the appearance of type IV collagen preceded the formation of basement membranes. The same question can be addressed for another basement membrane collagen, type XVIII, which has been characterized in nematode and drosophila (Hynes and Zhao, 2000; Ackley et al., 2001).

In view of the absence of basement membranes in almost all of the sponges, and the similarities between spongin, nematode cuticular, and basement membrane type IV collagens, it is possible that the spongin family reflects two lines of evolution. One line might have been exocollagens (such as spongins) attaching sponges to their substrata (such as worm cuticles, mussel byssus threads, and the egg capsule of Selacians). The second might have been internalization of such collagens, leading to the differentiation of basement membrane collagens. Finally, verte-

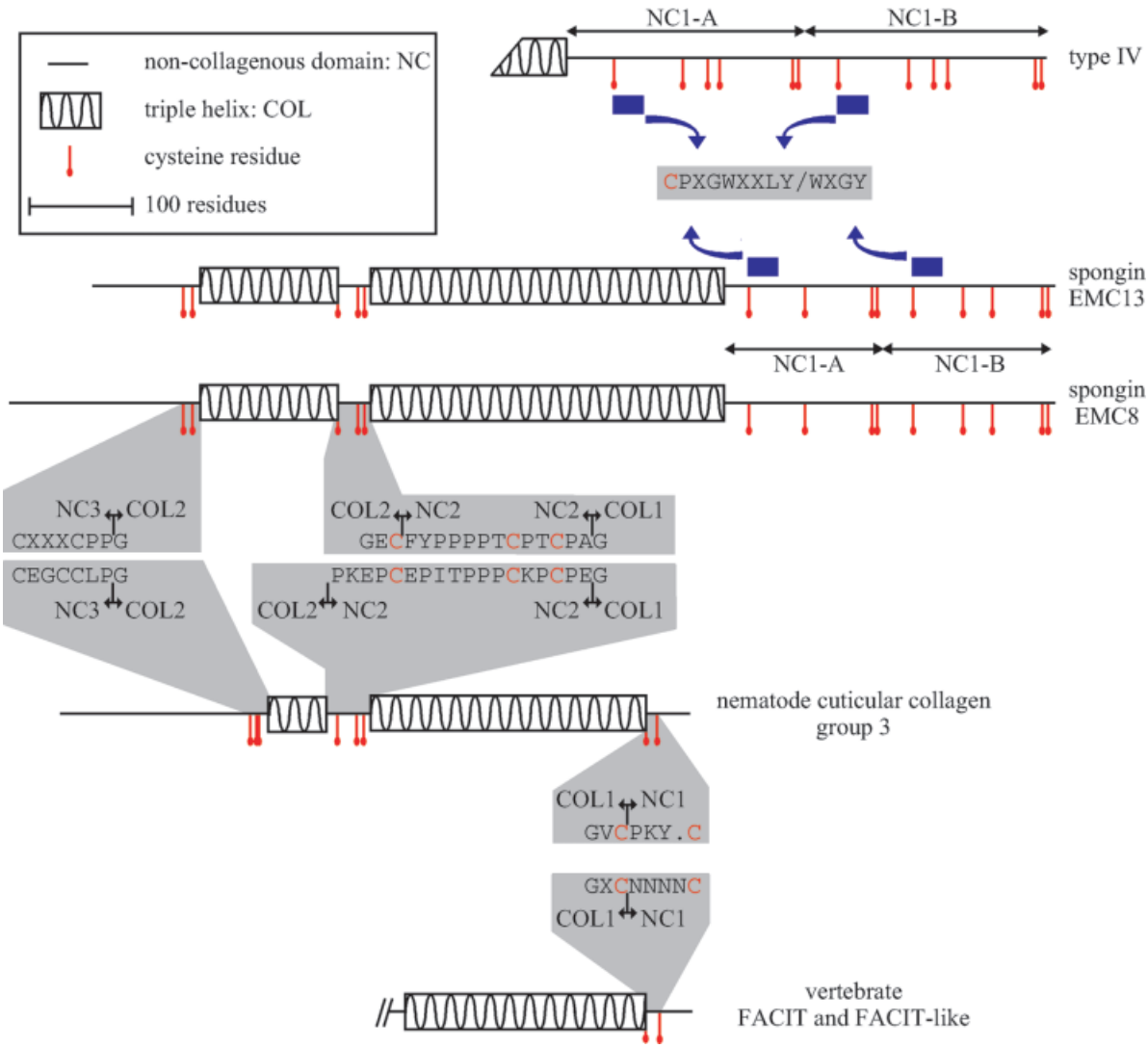

Fig. 4. Features common to spongins, nematode cuticular collagens, type IV, and FACIT collagens. Cysteine residues are in red. NC1-A and NC1-B represent the two similar subdomains of spongins and type IV collagens. The different collagen chains are illustrated to scale, with only the C-terminus shown for type IV and FACIT collagens. For nematodes, we used the sequence of one member of group 3 (c53b4.5) (Johnstone, 2000). The conserved sequence observed in both spongins and type IV collagen is represented by a blue box, and the consensus sequence is indicated.

brate FACIT and FACIT-related collagens appear to be evolutionarily related to nematode cuticular collagens. All consist of several short collagenous domains, with similar C-terminal noncollagenous (NC1) domains as well as conserved cysteine residues at the COL1–NC1 junctions.

Other Invertebrate Collagens

Several collagens seem to be specific to invertebrates. As indicated by Engel (1997), these invertebrate collagens are a good example of the functional diversity arising from the combination of triple helical domains and noncollagenous blocks. Hence, collagen from mussel byssus consists of a short collagenous region (128 nm long) flanked on each side by an elastic region ending with a histidine-rich domain. Coyne et al. (1997) suggested that the histidine-cluster may support zinc-mediated cross-linking of byssal collagens. The elastic region may be involved in the extensibility of the byssus thread. In comparison with the 300-nm-long triple helix of fibrillar collagen molecules, we note the short collagenous domain of hydra mini-collagens (14 nm) and the longest characterized collagenous region of annelid cuticle collagen (2,400 nm). The vestimentiferan tube worm *Riftia pachyptila* and the annelid *Alvinella pompejana* live around deep-sea hydrothermal vents. Their cuticle collagens are long (1,500 nm for *Riftia pachyptila and 2,400 nm for Alvinella pompejana*), but they clearly present a distinct amino acid content (Gaill et al., 1991). Their denaturation temperatures are comparable (37–40°C). While the cuticle collagen from *Riftia pachyptila* is poor in 4-hydroxyproline (21/1,000 vs. 178/1,000 for *Alvinella* cuticle), it contains an unusually high threonine content (20%). Mann et al. (1996) have shown that about 40% of the Y positions in the triple helix are occupied by glycosylated threonine residues, and that this

glycosylated amino-acid is the major contributor to triple helix stabilization of *Riftia* cuticular collagen. Partial amino acid sequence analysis of this cuticular collagen has revealed that its triple helix contains a few imperfections.

Invertebrate data have increased our knowledge about fibrillar collagen evolution, and to a lesser extent about basement membrane and FACIT collagens, although several collagens appear to be peculiar to invertebrates. Another interesting group, discussed in the two next sections, corresponds to the collagen-like proteins that have been found in both animals and prokaryotes.

Eukaryotic Collagen-Like Proteins

Soluble defense collagens. As indicated above, several vertebrate proteins possess collagenous domains but have not been included in the collagen family, because they are not involved in ECM structure. In this group of proteins are the collectins, which include a carbohydrate recognition domain or C-type (calcium-dependent) lectin connected to a collagenous region via an alpha helically-coiled neck region. This lectin family includes the mannan-binding proteins (MBPs), pulmonary surfactant proteins (SP-A and SP-D), collectin liver 1 (CL-L1), collectin placenta 1 (CL-P1), conglutinin, and collectin-43 (Hakansson and Reid, 2000; Ohtani et al., 2001). The collectin collagenous domain forms a triple helix and undergoes post-translational modifications such as hydroxylation of proline residues (for MBP see Ma et al. (1997)). Moreover, multimers (usually six) of trimers associate via the collagenous domains.

Several lines of evidence indicate that collectins are involved in innate immunity. Thus, SP-A and SP-D play important functions in the host defense of the lung, which is continually exposed to pathogens and allergenic agents. Such collectins can aggregate infectious agents via their carbohydrate recognition sites, and promote phagocytosis by macrophages. By binding to glycosylated allergenic proteins, they can prevent allergens from binding to specific IgE antibodies, thereby suppressing the release of inflammatory mediators from mast cells and basophils. MBP can bind to carbohydrate structures on the surface of microorganisms, acting as a direct opsonin and promoting phagocytosis via collectin receptors. MBP can also bind to microorganisms and mediate their lysis by activation of complement via the mannan-binding-lectin pathway (Zhang et al., 1999). This pathway is independent of complement component C1 and involves the formation of MBL-serine protease (MASP-1 and MASP-2) complexes, which activate complement factors C2 and C4. Recently, another group of collagen-like proteins, the ficolins, have been shown to activate the complement mannan-binding-lectin pathway (Matsushita et al., 2000). Ficolins can also act as opsonins and phagocytosis enhancers. They are structurally similar to collectins, with the replacement of the carbohydrate recognition domain by a fibrinogen-like region. Collectins and ficolins are involved in host defense, and for this reason they have been called soluble defense collagens. Two other members of the soluble defense collagen family are complement C1q and adiponectin. Both proteins consist of a carboxyl-terminal so-called C1q domain linked to an amino-terminal collagenous domain (Kishore and Reid, 1999; Yokota et al., 2000).

C1q-containing collagens. The C1q domain has been identified in several collagens and collagen-like proteins (Kishore and Reid, 1999). The latter include hibernian proteins found in the blood of mammalian hibernators (Kondo and Kondo, 1992), and EMILIN-1 and EMILIN-2, two ECM proteins widely distributed in elastic tissues (Colombatti et al., 2000). The C1q-containing collagens include the short-chain collagen types VIII and X, and the inner-ear structural protein of the bluegill sunfish, *Lepomis macrochirus* (Davis et al., 1995). In type X collagen, the C1q module forms a stable and compact trimer in the absence of the triple-helical domain (Dublet et al., 1999), indicating that the collagenous domain is not essential for trimerization. Similarly, in collectins, trimerization can occur in the absence of a collagen-like triple helix. In collectins, the α-helical coiled-coil neck region is responsible for trimerization, which in turn might promote the formation of the collagen-like triple helical region (Kishore et al., 1997; Kishore and Reid, 1999).

Membrane-associated type II collagen-like proteins. Another group of collagen-like proteins, the scavenger receptor-class A family (Shirai et al., 1999; Elshourbagy et al., 2000), is involved in host defense. These receptors include three isoforms of the macrophage scavenger receptor (SR-A) and the macrophage receptor with a collagenous structure (MARCO). Charged residues, including a lysine cluster in the collagenous domain of scavenger receptors, have been identified as a ligand-binding domain for a broad range of negatively charged macromolecules, such as modified low-density lipoproteins. These residues are also involved in the recognition of damaged or apoptotic cells and pathogenic microorganisms (Doi et al., 1993; Andersson and Freeman, 1998). Hence, these receptors are involved in the pathological deposition of cholesterol in foam cells of atherosclerotic lesions. Scavenger receptors contain a short cytoplasmic region at their amino-termini, and after the transmembrane region present a coiled-coil domain and a collagen-like domain made up of 23 Gly-Xaa-Yaa repeats (SR-A). MARCO lacks the coiled-coil domain and presents a large collagenous region made up of 90 Gly-Xaa-Yaa triplets. As shown in Figure 5, SR-A and MARCO are closely related to CL-P1, a transmembrane collectin, except that the C-lectin domain is replaced by the scavenger receptor cysteine-rich region. All belong to a group of membrane-associated type II proteins. This group includes another collagen-like protein, ectodysplasin, a transmembrane protein believed to have an important function in epithelial-mesenchymal interactions during the development of affected epithelial appendages, such as ectodermal dysplasias (Ezer et al., 1999). Two other members are collagen types XIII and XVII. For type XIII collagen, a short sequence adjacent to the transmembrane region in the extracellular environment is required for its trimerization. This region is also observed in other type II transmembrane collagens, i.e., type XVII collagen, MARCO, and ectodysplasin-A1 (Snellman et al., 2000).

Collagen tail of acetyl cholinesterase. Another collagen-like protein observed in vertebrates is the Q or collagen-tail subunit of acetyl cholinesterase. The collagen-tail subunit forms trimeric molecules, with each chain containing two collagenous regions of 58 and 5 Gly-Xaa-Yaa triplets, respectively, separated by a short non-collagenous sequence (Krejci et al., 1997). This molecule is

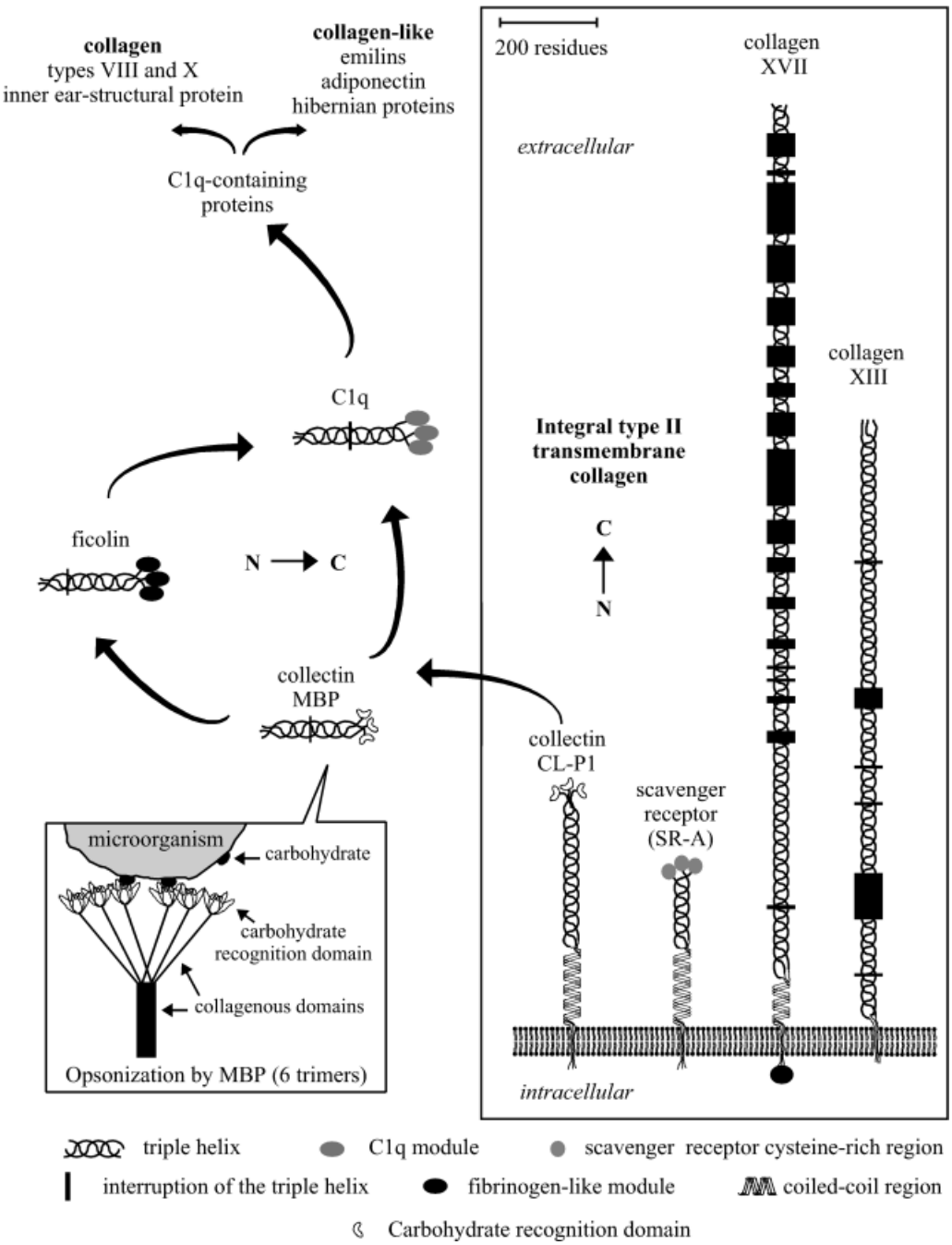

Fig. 5. Organization common to defense, short-chain, and transmembrane collagens. On the left, we have sketched the integral type II transmembrane collagens. Type XIII and XVII collagens present multiple interruptions in their triple helix. The scavenger receptor (SR-A) contains a coiled-coil region similar to type XVII collagen, while in collectin CL-P1, a carbohydrate recognition domain replaces the scavenger receptor cysteine-rich SR-A domain. From the transmembrane collectin we can progress to soluble collectin, which shares a common organization with ficolins and C1q. Finally, we have all the collagens and collagen-like proteins possessing a common C-terminus with C1q, i.e., a collagenous region and a C1q module. In this illustration, only the collagenous domain and these interruptions are represented to scale. Molecular orientation is indicated (N→C).

involved in the formation of a quaternary structure consisting of a trimeric collagen tail and one to three tetramers of the catalytic subunit of acetyl cholinesterase. This form represents the major component of acetyl cholinesterase in the neuromuscular junction of higher vertebrates. The exons coding for the triple helix of this Q subunit are of 27, 36, 45, 63, and 81 bp, and the introns are in phase zero.

General comments. One of the hallmarks of collagen-related proteins is that they all form multimeric aggre-

gates. Hence, the C1q molecule is composed of 18 polypeptide chains. As indicated by Engel and Kammerer (2000), oligomerization not only increases structural stabilization, but also leads to physiological advantages by grouping together functional sites. It is noteworthy that the C1q module is structurally related to tumor necrosis factors (Shapiro and Scherer, 1998), and that most tumor necrosis factor ligands are known to be active as trimers. Another feature of the collagen-like proteins is that the exons encoding their triple-helical regions begin with a split glycine codon, and the introns are in phase 1. This is also the case for type XVII collagen, whereas for type XIII almost all of the exons begin with an intact glycine codon. Altogether, these data indicate a common origin for the defense collagens, the C1q-containing proteins including types VIII and X collagens, type XVII collagen and ectodysplasin. The relation between these collagens and collagen-like proteins is presented in Figure 5. Because of its structural similarity with other transmembrane collagens, type XIII collagen is also included in Figure 5.

Although some components of innate immunity have been characterized in sponges (for review see Müller et al., 1999), no defense collagens have yet been characterized in invertebrates. Ascidians (Urochordata) can represent an intermediary position between invertebrates and vertebrates. In this phylum, several components of a primitive lectin-dependent complement system have been characterized, including ficolins and mannose-binding like proteins. It should be noted that the collagenous region of ascidian ficolins is shorter than that of vertebrate ficolins (Kenjo et al., 2001), and that the collagenous region of MBL is absent in the Ascidian MBL-like proteins characterized so far (Sekine et al., 2001).

Prokaryote Collagen-Like Proteins

Proteins possessing Gly-X-Y repeats have also been identified in prokaryotes. A collagen-like surface antigen has been shown to induce platelet aggregation in *Streptococcus* sanguis (Erickson and Herzberg, 1987). SclA is also a collagen-like surface protein that might contribute to the virulence of the gram-positive pathogen *Streptococcus* pyogenes (Rasmussen et al., 2000). The collagenous region is 171 residues long and might act as a stalk on which the hypervariable region of the protein is exposed. Short collagenous stretches have also been characterized in hyaluronidases of *Streptococcus pyogenes*. This enzyme degrades hyaluronic acid, a major component of the extracellular matrix. Degradation of hyaluronic acid is thought to help host tissue invasion and dissemination of *S. pyogenes.* The collagenous domain of these hyaluronidases is not necessary for this enzyme to function, but might play a role in human disease by inducing the production of antibodies which may cross-react with host tissues (Hynes et al., 1995). Another collagen-like protein (Saimiri transforming protein (STP)) has been identified in *Herpesvirus* saimiri, a gamma-2 herpesvirus. The collagenous domain is made up of 18 Gly-Xaa-Yaa repeats and is a determinant of the degree of *Herpesvirus* saimiri transforming activity (Choi et al., 2000). This virus family has been divided into three subgroups. Two of these subgroups (A and C) transform primary common marmoset lymphocytes to interleukin-2-independent growth, whereas subgroup B does not. In subgroup B, the collagenous region is absent from the STP protein (STP-B). Addition of the collagenous stretch in STP-B permits its oligomerization and restores transforming activity. Hence, while some vertebrate collagen-like proteins are involved in host defense, bacteria also use collagen-related proteins during infection.

Triple Helix as a Module, and Metazoan ECM Evolution

How old is collagen? According to fossil analyses, the traditional view is that animals originated 565 million years ago during the Vendian period of the latest Neoproterozoic era. This was followed by a "big-bang" episode during the Vendian-Cambrian period, during which time all the major phyla were established (Bowring et al., 1993; Smith et al., 1999; Philippe et al., 2000). However, according to sequence analyses, animals are more ancient and arose about 1,000–1,200 million years ago (Wray et al., 1996). Indeed, as indicated by Knoll and Carroll (1999), Cambrian fossils are not the oldest animal records. Impressions, molds, and casts of "animals" have been observed in rock dating back more than 600 million years ago. One suggestion is that animal evolution was triggered by environmental changes (Knoll and Carroll, 1999), among which was an increase in oxygen in the late proteozomic (Canfield and Teske, 1996). Knoll and Carroll (1999) suggested that the increased oxygen removed an environmental barrier to the evolution of large, metabolically active animals. Increased oxygen provided not only sufficient oxygen for animal respiration, but also became a key factor in the synthesis of collagen (Towe, 1970; Canfield and Teske, 1996) through its involvement in the post-translational hydroxylation of proline and lysine residues.

It is therefore likely that the hypothetical ancestor of metazoans contained collagen. But what is the relevance of collagenous domains in prokaryotes? The first possibility is that the collagen sequence predated the divergence of prokaryotes and eukaryotes. The second possibility is that collagen sequences were horizontally transferred from animals to prokaryotes. Whatever the answer, different steps and functions must be considered in the evolution of collagens, which can be considered as a hallmark of metazoans. From the first multicellular animal (the sponge) to humans, only two collagen types appear to have been conserved, i.e., the fibrillar and the basement membrane collagens. The latter have been identified only in one class of sponge. Sponges are anatomically simple, with a body surrounded by a layer of pinacoderm cells and internal tissue, the mesohyl, consisting of cells and a structured matrix rich in collagen fibrils. Hence, in comparison with bacterial biofilms, sponges not only have cells in contact with the environment, they also have cells embedded in a structured extracellular matrix (Garrone, 1999).

Functions of collagen. In sponges, collagens serve several functions. Fine fibrils made of fibrillar collagen are involved in construction of the mesohyl, while spongins help comprise the skeleton and permit the adhesion of the sponge to its substratum. As indicated by Müller (1997), sponge mesohyl should not be considered as an inert scaffold but as a dynamic and complex network of molecules that regulates the behavior of cells. The collagen fibrils not only provide the structure of the ECM, but also mediate cell–matrix interactions via membrane receptors, a situation observed in vertebrates. Integrins, for example, have

been identified in sponges (Pancer et al., 1997). During the process of evolution, the increase in diversity of fibrillar collagen chains, their different forms of maturation, and interactions with other ECM components permitted the formation of fibrils with distinct diameters and functions. Hence, vertebrate fibrils are involved in body protection in the dermis, in providing protection against mechanical strength in tendons, and in assuring the physical properties of cornea. Another interesting feature is the presence of collagen fibrils in the vertebrate skeleton, which is rich in calcium phosphate and nonfibrillar collagens (IX and X). In sponges, spongins are also incorporated into the skeleton with or without an inorganic compound made of calcium carbonate. Between sponges and vertebrates, few animals possess a combination of collagen and calcium. Among these, we note the collagenous axial skeleton of *Veretillum cynomorium* (Ledger and Franc, 1978). Hence, the collagen-based skeleton has not been successful during evolution, except in vertebrates (Garrone, 1999).

Spongins also act as exocollagens which attach the sponge to its substratum, a function which seems necessary since the pinacoderm cells are loosely arranged. During evolution, the exocollagens had a protective role, such as in the cuticle of nematodes and annelids. The protective effect of collagen can be related to its relative resistance to a broad range of proteases. Moreover, although collagen molecules are trimeric, they are involved alone or with other collagenous and noncollagenous components in the formation of polymers which not only increase their resistance to proteases, but promote the organization of the ECM. Spongins are evolutionarily related to basement membrane collagens, and type IV collagens have been described in sponges. Type IV collagens, which form the primary scaffold of basement membranes, are involved in the attachment of epithelial and endothelial layers. Similar roles can be assigned to type VII collagen in anchoring fibrils (structures within the epidermal basement membrane involved in attachment to the dermis), hemidesmosomal type XVII collagen (involved in the attachment of keratinocytes to the basal lamina), and transmembrane collagen type XIII (which is associated with cell-adhesion sites in tissues and seems to be involved in linking muscle fibers with basement membranes) (Kvist et al., 2001). Adhesion, therefore, appears to be a common function of collagens during evolution.

Cell surface receptors of vertebrate collagens include discoidin receptors (DDR1 and DDR2), integrins, and glycoprotein VI. These receptors trigger different signaling pathways upon collagen-binding. For example, after damage to the vessel wall, the indirect interaction of collagen with platelet receptors (glycoproteins Ib-V-VI and integrin αIIbβ3) via von Willebrand factors, and the direct interactions of collagens with glycoprotein VI and integrin $\alpha2\beta1$ induce the adhesion and aggregation of platelets, leading to thrombus formation (Savage et al., 1998; Nieswandt et al., 2001). Adhesion depends on the triple-helical conformation of collagen (Morton et al., 1994). It is noteworthy that snake venoms contain C-type lectins and metalloproteinase-disintegrins that selectively modulate platelet adhesion or activation through interactions with adhesion receptors or their ligands (for review see Andrews and Berndt, 2000). Another example of mimicry is the use of collagen receptors by bacteria. For example, *Staphylococcus aureus* bacteria produce adhesins that mediate their adherence to ECM proteins. Adhesins are microbial surface components that recognize adhesive matrix molecules. One of the collagen-binding adhesins (CAN) is necessary for *S. aureus* to attach to cartilage in vitro, and acts as a virulence factor (Patti et al., 1994).

Collagen as a multimodular protein. While collagen domains can be involved in different forms of interaction, the noncollagenous regions are also important, in two respects. First, noncollagenous regions are often necessary for the formation of the triple-helix-like domain in several defense collagens. These regions can also contain the sequences responsible for chain selection and assembly, such as the C-propeptide for fibrillar collagens and the NC1 domain for type IV collagens. Often, noncollagenous domains are involved in supramolecular assembly. Thus, in type IV collagen, two molecules are connected via their NC1 domains to form a dimer. Similarly, in type VII collagen, the C-terminal noncollagenous region is involved in the formation of an antiparallel dimer (Chen et al., 2001), while the von Willebrand factor-like N5 domain of the α3(VI) chain is important for type VI collagen microfibril formation (Fitzgerald et al., 2001). Second, noncollagenous regions include protein modules that, together with the collagenous regions, provide scope for generating increasing diversity from sponges to humans. Hence, collagens form one of the families of multimodular proteins (also called mosaic proteins) present in metazoans.

Two events, gene duplication and exon shuffling, may have contributed to the major increase in complexity during metazoan evolution. As indicated by Sankoff (2001), there are three consequences for duplicate genes: 1) higher expression levels of the gene product; 2) deletion of one copy; and 3) acquisition of a novel functional role or specialization of some aspect of previous function. Good examples of duplicate genes are the nematode cuticular collagens, which are encoded by approximately 154 genes (Johnstone, 2000). Gene duplications can also occur by duplication of genome segments. In humans, genes encoding the pro-α2(I), pro-α1(I), proα1(II), and proα1(III)-proα2(V) chains are closely linked to the Hox gene clusters A, B, C, and D, respectively (Bailey et al., 1997). This situation supports a linkage between duplication of collagen genes and Hox genes (Fig. 3B). The same type of association has been shown for the genes encoding the proα1(V), proα3(V), proα1(XI), and proα2(XI) chains in another homologous cluster seen in human chromosomes 1, 6, 9, and 19 (Smith et al., 1999) (human genome data base). These two groups of homologous clusters clearly agree with phylogenetic analysis showing a clear subdivision of the mammalian fibrillar α chains into two groups (Sicot et al., 1997; Exposito et al., 2000; Välkkilä et al., 2001).

Exon shuffling and metazoan evolution. Multimodular or mosaic proteins are composed of modules that are defined as evolutionarily mobile. By definition then, a module should be present in otherwise unrelated proteins and/or present in multiple copies in one protein (Doolittle, 1995). Mobile modules should also to be able to fold independently (Kolkman and Stemmer, 2001) and be encoded by one exon. This notion of exon shuffling was postulated by Gilbert (1978) and can occur by recombining introns and by retropositioning (Long, 2001). It has been suggested that exon shuffling is essential in the formation of mosaic proteins and, as indicated by Patthy (1999), this became significant with the formation of spliceosomal in-

trons, and increasingly so with the evolution of larger and less compacted genomes (larger spliceosomal introns) from protists to humans. Most mosaic proteins are extracellular or represent the extracellular regions of transmembrane proteins, while a large number of mobile modules contain cysteine residues forming intradomain disulfide bonds. This indicates the extreme importance of exon shuffling in metazoan evolution and in the formation of young proteins (Patthy, 1999), since these proteins are associated with the notion of multicellular animals. According to Patthy (1999), a good correlation can be made between the burst of the metazoan radiation and the rise of exon shuffling. Among mosaic proteins, the collagens have been some of the key elements needed for multicellularity and metazoan evolution. In the range of sponges to humans, we can note an increase in the diversity of collagens, and the formation of multimodular collagen proteins. In one sea urchin fibrillar collagen chain, 12 repetitions of a 4-cysteine module (SURF) are present in the amino-propeptide between a Tsp-2 mobile module and the short triple helix (Exposito et al., 1992b). The SURF module is encoded mainly by one exon, and the introns separating the SURF-coding exons are in phase 1 (Exposito et al., 1995). SURF modules are present in several extracellular collagenous and noncollagenous proteins in sea urchin (Cluzel et al., 2001) (Cluzel et al., unpublished data). All of these data match perfectly the definition of a mobile module. The Tsp-2 encoding exon is also in phase 1. A strong bias in favor of phase 1 introns has been indicated by Patthy (1996, 1999) for genes encoding young mosaic proteins.

Triple-helical sequence as a mobile module. Currently, several nonfibrillar collagen families have been defined in mammals, but unlike the case of the fibrillar collagens, it is difficult to suggest putative steps leading to their formation. In a previous study of sponges (Exposito et al., 1991), we suggested that two types of collagenous coding exons beginning with a glycine codon (i.e., one in phase 0 (sponge fibrillar collagen genes), and one in phase 1 (spongin collagen genes)) are the primitive building blocks of collagen genes, and that the convergence or divergence of these two genetic units arose early during evolution. As presented in Figure 3A, the putative genomic region coding for the main triple helix of an ancestral fibrillar collagen gene is made up of 54 and 45 bp exons in phase 0. Nevertheless, 36 and 63 bp exons in phase 0 have been characterized in the region encoding the minor triple helix and/or the C-terminus of the main triple helix. Type IV collagen and spongins have two features in common: an interrupted triple-helical domain and an evolutionarily related C-terminal NC1 domain. As in sponge genes, almost all exons encoding the C-terminal part of the triple helix are in phase 1, while in the 5′ part, exons encoding exclusively collagenous sequences are in phase 0 (Oohashi et al., 1995). These phase 0 exons are of the same length as those already identified in fibrillar collagen genes, except that some are 27 bp long.

For other collagen genes, three situations have been described for the exons encoding collagenous sequences. The first group includes genes possessing only phase 0 exons (i.e. the genes encoding collagen types VI, VII, and XIII), as well as the gene encoding the collagen tail of acetyl cholinesterase (Krejci et al., 1997). Although these genes are built of phase 0 exons, there is a predominance of 63 bp exons in the type VI gene (Saitta et al., 1991), 36 bp exons in the type VII gene (Christiano et al., 1994), and 27 bp exons in the type XIII gene (Kvist et al., 1999). The second group includes collagen genes composed of only phase 1 exons. This group contains all of the defense collagen genes, as well as type XVII collagen and *EMILIN* genes (Gatalica et al., 1997; Lawson et al., 1999; Doliana et al., 2000). The third group is composed mainly of phase 0 exons, which is represented by the genes encoding types XV and XVIII collagens. FACIT and FACIT-related genes are also included in this group, and only exons encoding the most C-terminal collagenous domain (COL1) are in phase 1 (Khaleduzzaman et al., 1997). Nematode cuticular collagens are related to these vertebrate collagens since they possess a comparable C-terminal noncollagenous junction as well as two highly conserved cysteine residues in the COL1-NC1 junction. As in the 3′ regions of FACIT genes, nematode genes are made of phase 1 exons. From these data, it is sometimes difficult to discern which fundamental exon sizes have been involved in the formation of collagen genes, but it is tempting to consider the Gly-Xaa-Yaa sequence, or multiples thereof, as a mobile module. In this way, it is easy to understand the advantage of such a module during evolution, which can lead to the formation of new genetic units made up of one exon of variable size but also multiples of 9 bp, or of two exons, as suggested for fibrillar collagen genes (Exposito et al., 1993). All of these units can be used to generate distinct triple helices.

Emergence of collagens during evolution. Only two types of vertebrate collagens have been described in diploblastic animals, and, despite the availability of protostome genomes (nematode and drosophila), type XVIII-like collagen is the only other vertebrate orthologue identified to date in invertebrates.

In vertebrates, type XVIII collagen is a component of endothelial and epithelial basement membranes, and has been associated with Knobloch syndrome, a disease characterized by neural tube closure defects and/or scalp and retinal degeneration (Sertie et al., 2000). Better known are the functions of the C-terminal noncollagenous domain, endostatin, which represents the preponderant molecular form in several tissues. Endostatin has been reported to inhibit endothelial cell proliferation and migration, angiogenesis, and tumor growth, and to activate endothelial cell apoptosis under some culture conditions (for review, see Zatterstrom et al., 2000). In nematodes, type XVIII collagen is observed in all basement membranes, but predominantly in neurons. It plays a functional role in cell and axon migration (Ackley et al., 2001). Although other vertebrate collagens are absent in two protostome animals, whether they are present in invertebrate deuterostomes or lower vertebrates is still unknown. It would be interesting to trace the appearance of hemidesmosomes and type XVII collagens. The forthcoming complete characterization of both the sea urchin (http:/sea-urchin.caltech.edu/genome) and zebrafish genomes will enable several questions concerning collagen evolution to be answered. Other insights will also come from a more extensive study of diploblastic genomes. Another challenge will be to further explore collagen structures that are still unknown in numerous animals. Altogether, these data will permit a better understanding of collagens and their evolution.

CONCLUSIONS

Clearly, the development of domains involved in oligomerization was a primordial step in the development of multifunctional proteins with the potential to form macromolecular networks. The collagenous domain should be considered not only as a module that permits oligomerization, but also as a rod that connects globular domains. As indicated by van der Rest and Garrone (1991), the triple-helical sequence has a strong potential for lateral interactions such as that defined for the interactions of type I and IV collagens with integrins $\alpha1\beta1$ and $\alpha2\beta1$ (Knight et al., 2000). In this respect, collagens and collagen-like proteins are unique and essential in the appearance and diversity of metazoa.

ACKNOWLEDGMENTS

We thank V. Miromov for the invitation to write this review. We are indebted to David J. Hulmes for critically reading the manuscript, and revising the English.

LITERATURE CITED

Ackley BD, Crew JR, Elamaa H, Pihlajaniemi T, Kuo CJ, Kramer JM. 2001. The NC1/endostatin domain of *Caenorhabditis elegans* type XVIII collagen affects cell migration and axon guidance. J Cell Biol 152:1219–1232.

Anderson CM, Wagner TA, Perret M, He ZH, He D, Kohorn BD. 2001. WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix. Plant Mol Biol 47:197–206.

Andersson L, Freeman MW. 1998. Functional changes in scavenger receptor binding conformation are induced by charge mutants spanning the entire collagen domain. J Biol Chem 273:19592–19601.

Andrews RK, Berndt MC. 2000. Snake venom modulators of platelet adhesion receptors and their ligands. Toxicon 38:775–791.

Bailey WJ, Kim J, Wagner GP, Ruddle FH. 1997. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol Biol Evol 14: 843–853.

Birk DE, Silver FH. 1984. Kinetic analysis of collagen fibrillogenesis: II. Corneal and scleral type I collagen. Coll Relat Res 4:265–277.

Boute N, Exposito JY, Boury-Esnault N, Vacelet J, Noro N, Miyazaki K, Yoshizato K, Garrone R. 1996. Type IV collagen in sponges, the missing link in basement membrane ubiquity. Biol Cell 88:37–44.

Bowring SA, Grotzinger JP, Isachsen CE, Knoll AH, Pelechaty SM, Kolosov P. 1993. Calibrating rates of early Cambrian evolution. Science 261:1293–1298.

Burrows NP. 1999. The molecular genetics of the Ehlers-Danlos syndrome. Clin Exp Dermatol 24:99–106.

Canfield DE, Teske A. 1996. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. Nature 382:127–132.

Chen M, Keene DR, Costa FK, Tahk SH, Woodley DT. 2001. The carboxyl terminus of type VII collagen mediates antiparallel dimer formation and constitutes a new antigenic epitope for epidermolysis *Bullosa acquisita* autoantibodies. J Biol Chem 276:21649–21655.

Choi JK, Ishido S, Jung JU. 2000. The collagen repeat sequence is a determinant of the degree of herpesvirus saimiri STP transforming activity. J Virol 74:8102–8110.

Christiano AM, Hoffman GG, Chung-Honet LC, Lee S, Cheng W, Uitto J, Greenspan DS. 1994. Structural organization of the human type VII collagen gene (*COL7A1*), composed of more exons than any previously characterized gene. Genomics 21:169–179.

Cluzel C, Lethias C, Garrone R, Exposito JY. 2000. Production in mammalian cells of chimeric human/sea urchin procollagen molecules displaying distinct versions of the minor triple helix. J Biochem (Tokyo) 128:957–963.

Cluzel C, Lethias C, Humbert F, Garrone R, Exposito JY. 2001. Characterization of fibrosurfin, an interfibrillar component of sea urchin catch connective tissues. J Biol Chem 276:18108–18114.

Colombatti A, Doliana R, Bot S, Canton A, Mongiat M, Mungiguerra G, Paron-Cilli S, Spessotto P. 2000. The EMILIN protein family. Matrix Biol 19:289–301.

Coyne KJ, Qin XX, Waite JH. 1997. Extensible collagen in mussel byssus: a natural block copolymer. Science 277:1830–1832.

Davey ME, O'Toole GA. 2000. Microbial biofilms: from ecology to molecular genetics. Microbiol Mol Biol Rev 64:847–867.

Davis JG, Oberholtzer JC, Burns FR, Greene MI. 1995. Molecular cloning and characterization of an inner ear-specific structural protein. Science 267:1031–1034.

DeRobertis EM, Sasai Y. 1996. A common plan for dorsoventral patterning in Bilateria. Nature 380:37–40.

Deutzmann R, Fowler S, Zhang X, Boone K, Dexter S, Boot-Handford RP, Rachel R, Sarras Jr MP. 2000. Molecular, biochemical and functional analysis of a novel and developmentally important fibrillar collagen (Hcol-I) in hydra. Development 127:4669–4680.

Doi T, Higashino K, Kurihara Y, Wada Y, Miyazaki T, Nakamura H, Uesugi S, Imanishi T, Kawabe Y, Itakura H, Yasaki H, Matsumoto A, Kodama T. 1993. Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors. J Biol Chem 268:2126–2133.

Doliana R, Canton A, Bucciotti F, Mongiat M, Bonaldo P, Colombatti A. 2000. Structure, chromosomal localization, and promoter analysis of the human Elastin Microfibril Interfase Located proteIN (*EMILIN*) gene. J Biol Chem 275:785–792.

Doolittle RF. 1995. The multiplicity of domains in proteins. Annu Rev Biochem 64:287–314.

Dublet B, Vernet T, van der Rest M. 1999. Schmid's metaphyseal chondrodysplasia mutations interfere with folding of the C-terminal domain of human collagen X expressed in *Escherichia coli.* J Biol Chem 274:18909–18915.

Elshourbagy NA, Li X, Terrett J, Vanhorn S, Gross MS, Adamou JE, Anderson KM, Webb CL, Lysko PG. 2000. Molecular characterization of a human scavenger receptor, human MARCO. Eur J Biochem 267:919–926.

Engel J. 1997. Versatile collagens in invertebrates. Science 277:1785–1786.

Engel J, Kammerer RA. 2000. What are oligomerization domains good for? Matrix Biol 19:283–288.

Erickson PR, Herzberg MC. 1987. A collagen-like immunodeterminant on the surface of *Streptococcus sanguis* induces platelet aggregation. J Immunol 138:3360–3366.

Exposito JY, Garrone R. 1990. Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. Proc Natl Acad Sci USA 87:6669–6673.

Exposito JY, Ouazana R, Garrone R. 1990. Cloning and sequencing of a Porifera partial cDNA coding for a short-chain collagen. Eur J Biochem 190:401–406.

Exposito JY, Le Guellec D, Lu Q, Garrone R. 1991. Short chain collagens in sponges are encoded by a family of closely related genes. J Biol Chem 266:21923–21928.

Exposito JY, D'Alessio M, Solursh M, Ramirez F. 1992a. Sea urchin homolog of vertebrate pro-$\alpha2$(I) collagen gene. J Biol Chem 267: 15559–15562.

Exposito JY, D'Alessio M, Ramirez F. 1992b. Novel amino-propeptide configuration in a fibrillar procollagen undergoing alternative splicing. J Biol Chem 267:17404–17408.

Exposito JY, van der Rest M, Garrone R. 1993. The complete intron/exon structure of *Ephydatia mülleri* fibrillar collagen gene suggests a mechanism for the evolution of an ancestral gene module. J Mol Evol 37:254–259.

Exposito JY, Boute N, Deleage G, Garrone R. 1995. Characterization of two genes coding for a similar four-cysteine motif of the amino-terminal propeptide of a sea urchin fibrillar collagen. Eur J Biochem 234:59–65.

Exposito JY, Cluzel C, Lethias C, Garrone R. 2000. Tracing the evolution of vertebrate fibrillar collagens from an ancestral α chain. Matrix Biol 19:275–279.

Ezer S, Bayes M, Elomaa O, Schlessinger D, Kere J. 1999. Ectodysplasin is a collagenous trimeric type II membrane protein with a tumor necrosis factor-like domain and co-localizes with cytoskeletal

structures at lateral and apical surfaces of cells. Hum Mol Genet 8:2079–2086.
Fields GB. 1991. A model for interstitial collagen catabolism by mammalian collagenases. J Theor Biol 153:585–602.
Fitzgerald J, Bateman JF. 2001. A new FACIT of the collagen family: *COL21A1*. FEBS Lett 505:275–280.
Fitzgerald J, Morgelin M, Selan C, Wiberg C, Keene DR, Lamande SR, Bateman JF. 2001. The N-terminal N5 subdomain of the α3(VI) chain is important for collagen VI microfibril formation. J Biol Chem 276:187–193.
Fowler SJ, Jose S, Zhang X, Deutzmann R, Sarras Jr MP, Boot-Handford RP. 2000. Characterization of hydra type IV collagen. Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. J Biol Chem 275: 39589–39599.
Gaill F, Wiedemann H, Mann K, Kuhn K, Timpl R, Engel J. 1991. Molecular characterization of cuticle and interstitial collagens from worms collected at deep sea hydrothermal vents. J Mol Biol 221: 209–223.
Garrone R. 1978. Phylogenesis of connective tissue. In: Robert L, editor. Morphological aspects and biosynthesis of sponge intercellular matrix. Basel: S. Karger. 250 p.
Garrone R. 1999. Evolution of metazoan collagens. Prog Mol Subcell Biol 21:119–139.
Gatalica B, Pulkkinen L, Li K, Kuokkanen K, Tyynänen M, McGrath JA, Uitto J. 1997. Cloning of the human type XVII collagen gene (*COL17A1*), and detection of novel mutations in generalized atrophic benign epidermolysis bullosa. Am J Hum Genet 60:352–365.
Giguère Y, Rousseau F. 2000. The genetics of osteoporosis: 'complexities and difficulties'. Clin Genet 57:161–169.
Gilbert W. 1978. Why genes in pieces? Nature 271:501.
Guo XD, Kramer JM. 1989. The two *Caenorhabditis elegans* basement membrane (type IV) collagen genes are located on separate chromosomes. J Biol Chem 264:17574–17582.
Hakansson K, Reid KB. 2000. Collectin structure: a review. Protein Sci 9:1607–1617.
Hudson BG, Reeders ST, Tryggvason K. 1993. Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. J Biol Chem 268:26033–26036.
Hynes RO, Zhao Q. 2000. The evolution of cell adhesion. J Cell Biol 150:F89–F95.
Hynes WL, Hancock L, Ferretti JJ. 1995. Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: evidence for a third hyaluronidase involved in extracellular enzymatic activity. Infect Immunol 63:3015–3020.
Johnstone IL. 2000. Cuticle collagen genes. Expression in *Caenorhabditis elegans*. Trends Genet 16:21–27.
Kenjo A, Takahashi M, Matsushita M, Endo Y, Nakata M, Mizuochi T, Fujita T. 2001. Cloning and characterization of novel ficolins from the solitary ascidian, *Halocynthia roretzi*. J Biol Chem 276: 19959–19965.
Khaleduzzaman M, Sumiyoshi H, Ueki Y, Inoguchi K, Ninomiya Y, Yoshioka H. 1997. Structure of the human type XIX collagen (*COL19A1*) gene, which suggests it has arisen from an ancestor gene of the FACIT family. Genomics 45:304–312.
Kishore U, Eggleton P, Reid KB. 1997. Modular organization of carbohydrate recognition domains in animal lectins. Matrix Biol 15: 583–592.
Kishore U, Reid KB. 1999. Modular organization of proteins containing C1q-like globular domain. Immunopharmacology 42:15–21.
Knight CG, Morton LF, Peachey AR, Tuckwell DS, Farndale RW, Barnes MJ. 2000. The collagen-binding A-domains of integrins α(1)β(1) and α(2)β(1) recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. J Biol Chem 275:35–40.
Knoll AH, Carroll SB. 1999. Early animal evolution: emerging views from comparative biology and geology. Science 284:2129–2137.
Koch M, Foley JE, Hahn R, Zhou P, Burgeson RE, Gerecke DR, Gordon MK. 2001. α1(XX) collagen, a new member of the collagen subfamily, fibril-associated collagens with interrupted triple helices. J Biol Chem 276:23120–23126.
Kohorn BD. 2001. WAKs: cell wall associated kinases. Curr Opin Cell Biol 13:529–533.
Kolkman JA, Stemmer WP. 2001. Directed evolution of proteins by exon shuffling. Nat Biotechnol 19:423–428.
Kondo N, Kondo J. 1992. Identification of novel blood proteins specific for mammalian hibernation. J Biol Chem 267:473–478.
Krasko A, Lorenz B, Batel R, Schröder HC, Müller IM, Müller WE. 2000. Expression of silicatein and collagen genes in the marine sponge *Suberites domuncula* is controlled by silicate and myotrophin. Eur J Biochem 267:4878–4887.
Krejci E, Thomine S, Boschetti N, Legay C, Sketelj J, Massoulie J. 1997. The mammalian gene of acetylcholinesterase-associated collagen. J Biol Chem 272:22840–22847.
Kuivaniemi H, Tromp G, Prockop DJ. 1997. Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. Hum Mutat 9:300–315.
Kvist AP, Latvanlehto A, Sund M, Horelli-Kuitunen N, Rehn M, Palotie A, Beier D, Pihlajaniemi T. 1999. Complete exon-intron organization and chromosomal location of the gene for mouse type XIII collagen (col13a1) and comparison with its human homologue. Matrix Biol 18:261–274.
Kvist AP, Latvanlehto A, Sund M, Eklund L, Vaisanen T, Hagg P, Sormunen R, Komulainen J, Fassler R, Pihlajaniemi T. 2001. Lack of cytosolic and transmembrane domains of type XIII collagen results in progressive myopathy. Am J Pathol 159:1581–1592.
Lawson PR, Perkins VC, Holmskov U, Reid KBM. 1999. Genomic organization of the mouse gene for lung surfactant protein D. Am J Respir Cell Mol Biol 20:953–963.
Ledger PW, Franc S. 1978. Calcification of the collagenous axial skeleton of *Veretillum cynomorium* pall. (Cnidaria: Pennatulacea). Cell Tissue Res 192:249–266.
Lethias C, Exposito JY, Garrone R. 1997. Collagen fibrillogenesis during sea urchin development: retention of SURF motifs from the N-propeptide of the 2α chain in mature fibrils. Eur J Biochem 245:434–440.
Linsenmayer TF, Gibney E, Igoe F, Gordon MK, Fitch JM, Fessler LI, Birk DE. 1993. Type V collagen: molecular structure and fibrillar organization of the chicken α1(V) NH_2-terminal domain, a putative regulator of corneal fibrillogenesis. J Cell Biol 121:1181–1189.
Long M. 2001. Evolution of novel genes. Curr Opin Genet Dev 11:673–680.
Ma Y, Shida H, Kawasaki T. 1997. Functional expression of human mannan-binding proteins (MBPs) in human hepatoma cell lines infected by recombinant vaccinia virus: post-translational modification, molecular assembly, and differentiation of serum and liver MBP. J Biochem 122:810–818.
Mann K, Mechling DE, Bachinger HP, Eckerskorn C, Gaill F, Timpl R. 1996. Glycosylated threonine but not 4-hydroxyproline dominates the triple helix stabilizing positions in the sequence of a hydrothermal vent worm cuticle collagen. J Mol Biol 261:255–266.
Matsushita M, Endo Y, Fujita T. 2000. Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease. J Immunol 164:2281–2284.
Miura S, Kimura S. 1985. Jellyfish mesogloea collagen. Characterization of molecules as α1α2α3 heterotrimers. J Biol Chem 260:15352–15356.
Morton LF, Peachey AR, Zijenah LS, Goodall AH, Humphries MJ, Barnes MJ. 1994. Conformation-dependent platelet adhesion to collagen involving integrin α2β1-mediated and other mechanisms: multiple α2β1-recognition sites in collagen type I. Biochem J 299: 791–797.
Müller WE. 1997. Origin of metazoan adhesion molecules and adhesion receptors as deduced from cDNA analyses in the marine sponge *Geodia cydonium*: a review. Cell Tissue Res 289:383–395.
Müller WE, Koziol C, Müller IM, Wiens M. 1999. Towards an understanding of the molecular basis of immune responses in sponges: the marine demosponge *Geodia cydonium* as a model. Microsc Res Tech 44:219–236.
Myllyharju J, Kivirikko KI. 2001. Collagens and collagen-related diseases. Ann Med 33:7–21.

Nieswandt B, Brakebusch C, Bergmeier W, Schulte V, Bouvard D, Mokhtari-Nejad R, Lindhout T, Heemskerk JW, Zirngibl H, Fassler R. 2001. Glycoprotein VI but not α2β1 integrin is essential for platelet interaction with collagen. EMBO J 20:2120–2130.

Ohtani K, Suzuki Y, Eda S, Kawai T, Kase T, Keshi H, Sakai Y, Fukuoh A, Sakamoto T, Itabe H, Suzutani T, Ogasawara M, Yoshida I, Wakamiya N. 2001. The membrane type collectin CL-P1 is a scavenger receptor on vascular endothelial cells. J Biol Chem 276:44222–44228.

Oohashi T, Ueki Y, Sugimoto M, Ninomiya Y. 1995. Isolation and structure of the COL4A6 gene encoding human α6(IV) collagen chain and comparison with other type IV collagen genes. J Biol Chem 270:26863–26867.

Pancer Z, Kruse M, Müller I, Müller WE. 1997. On the origin of Metazoan adhesion receptors: cloning of integrin α subunit from the sponge *Geodia cydonium*. Mol Biol Evol 14:391–398.

Patthy L. 1996. Exon shuffling and other ways of module exchange. Matrix Biol 15:301–310.

Patthy L. 1999. Genome evolution and the evolution of exon-shuffling—a review. Gene 238:103–114.

Patti JM, Bremell T, Krajewska-Pietrasik D, Abdelnour A, Tarkowski A, Ryden C, Hook M. 1994. The *Staphylococcus aureus* collagen adhesin is a virulence determinant in experimental septic arthritis. Infect Immun 62:152–161.

Philippe H, Germot A, Moreira D. 2000. The new phylogeny of eukaryotes. Curr Opin Genet Dev 10:596–601.

Pihlajaniemi T, Tryggvason K, Myers JC, Kurkinen M, Lebo R, Cheung MC, Prockop DJ, Boyd CD. 1985. cDNA clones coding for the pro-α1(IV) chain of human type IV procollagen reveal an unusual homology of amino acid sequences in two halves of the carboxyl-terminal domain. J Biol Chem 260:7681–7687.

Primorac D, Rowe DW, Mottes M, Barisic I, Anticevic D, Mirandola S, Gomez Lira M, Kalajzic I, Kusec V, Glorieux FH. 2001. Osteogenesis imperfecta at the beginning of bone and joint decade. Croat Med J 42:393–415.

Rasmussen M, Eden A, Bjorck L. 2000. SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. Infect Immun 68:6370–6377.

Saito M, Takenouchi Y, Kunisaki N, Kimura S. 2001. Complete primary structure of rainbow trout type I collagen consisting of α1(I)α2(I)α3(I) heterotrimers. Eur J Biochem 268:2817–2827.

Saitta B, Wang YM, Renkart L, Zhang RZ, Pan TC, Timpl R, Chu ML. 1991. The exon organization of the triple-helical coding regions of the human α1(VI) and α2(VI) collagen genes is highly similar. Genomics 11:145–153.

Sankoff D. 2001. Gene and genome duplication. Curr Opin Genet Dev 11:681–684.

Sarras Jr MP, Deutzmann R. 2001. Hydra and Niccolo Paganini (1782–1840)—two peas in a pod? The molecular basis of extracellular matrix structure in the invertebrate, Hydra. BioEssays 23:716–724.

Savage B, Almus-Jacobs F, Ruggeri ZM. 1998. Specific synergy of multiple substrate-receptor interactions in platelet thrombus formation under flow. Cell 94:657–666.

Schröder HC, Krasko A, Batel R, Skorokhod A, Pahler S, Kruse M, Müller IM, Müller WE. 2000. Stimulation of protein (collagen) synthesis in sponge cells by a cardiac myotrophin-related molecule from *Suberites domuncula*. FASEB J 14:2022–2031.

Sekine H, Kenjo A, Azumi K, Ohi G, Takahashi M, Kasukawa R, Ichikawa N, Nakata M, Mizuochi T, Matsushita M, Endo Y, Fujita T. 2001. An ancient lectin-dependent complement system in an ascidian: novel lectin isolated from the plasma of the solitary ascidian, *Halocynthia roretzi*. J Immunol 167:4504–4510.

Sertie AL, Sossi V, Camargo AA, Zatz M, Brahe C, Passos-Bueno MR. 2000. Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). Hum Mol Genet 9:2051–2058.

Shah NK, Sharma M, Kirkpatrick A, Ramshaw JA, Brodsky B. 1997. Gly-Gly-containing triplets of low stability adjacent to a type III collagen epitope. Biochemistry 36:5878–5883.

Shapiro L, Scherer PE. 1998. The crystal structure of a complement-1q family protein suggests an evolutionary link to tumor necrosis factor. Curr Biol 8:335–338.

Shirai H, Murakami T, Yamada Y, Doi T, Hamakubo T, Kodama T. 1999. Structure and function of type I and II macrophage scavenger receptors. Mech Age Dev 111:107–121.

Sicot FX, Exposito JY, Masselot M, Garrone R, Deutsch J, Gaill F. 1997. Cloning of an annelid fibrillar collagen gene and phylogenetic analysis of vertebrate and invertebrate collagens. Eur J Biochem 246:50–58.

Sicot FX, Mesnage M, Masselot M, Exposito JY, Garrone R, Deutsch J, Gaill F. 2000. Molecular adaptation to an extreme environment: origin of the thermal stability of the pompeii worm collagen. J Mol Biol 302:811–820.

Smith NG, Knight R, Hurst LD. 1999. Vertebrate genome evolution: a slow shuffle or a big bang? Bioessays 21:697–703.

Snellman A, Tu H, Vaisanen T, Kvist AP, Huhtala P, Pihlajaniemi T. 2000. A short sequence in the N-terminal region is required for the trimerization of type XIII collagen and is conserved in other collagenous transmembrane proteins. EMBO J 19:5051–5059.

Su MW, Suzuki HR, Bieker JJ, Solursh M, Ramirez F. 1991. Expression of two nonallelic type II procollagen genes during *Xenopus laevis* embryogenesis is characterized by stage-specific production of alternatively spliced transcripts. J Cell Biol 115:565–575.

Sutherland IW. 2001. The biofilm matrix—an immobilized but dynamic microbial environment. Trends Microbiol 9:222–227.

Takahara K, Hoffman GG, Greenspan DS. 1995. Complete structural organization of the human α1(V) collagen gene (*COL5A1*): divergence from the conserved organization of other characterized fibrillar collagen genes. Genomics 29:588–597.

Tillet E, Franc JM, Franc S, Garrone R. 1996. The evolution of fibrillar collagens: a sea-pen collagen shares common features with vertebrate type V collagen. Comp Biochem Physiol B 113:239–246.

Towe KM. 1970. Oxygen-collagen priority and the early metazoan fossil record. Proc Natl Acad Sci USA 65:781–788.

Trotter JA, Koob TJ. 1994. Biochemical characterization of fibrillar collagen from the mutable spine ligament of the sea urchin *Eucidaris tribuloides*. Comp Biochem Physiol 107B:125–134.

Välkkilä M, Melkoniemi M, Kvist L, Kuivaniemi H, Tromp G, Ala-Kokko L. 2001. Genomic organization of the human *COL3A1 and COL5A2* genes: *COL5A2* has evolved differently than the other minor fibrillar collagen genes. Matrix Biol 20:357–366.

van der Rest M, Garrone R. 1991. Collagen family of proteins. FASEB J 5:2814–2823.

Vogel KG, Trotter JA. 1987. The effect of proteoglycans on the morphology of collagen fibrils formed in vitro. Coll Relat Res 7:105–114.

Vuoristo MM, Pihlajamaa T, Vandenberg P, Prockop DJ, Ala-Kokko L. 1995. The human *COL11A2* gene structure indicates that the gene has not evolved with the genes for the major fibrillar collagens. J Biol Chem 270:22873–22881.

Wray GA, Levinton JS, Shapiro LH. 1996. Molecular evidence for deep precambrian divergences among metazoan phyla. Science 274:568573.

Yamada Y, Avvedimento VE, Mudryj M, Ohkubo H, Vogeli G, Irani M, Pastan I, de Crombrugghe B. 1980. The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp. Cell 22:887–892.

Yasothornsrikul S, Davis WJ, Cramer G, Kimbrell DA, Dearolf CR. 1997. Viking: identification and characterization of a second type IV collagen in *Drosophila*. Gene 198:17–25.

Yokota T, Oritani K, Takahashi I, Ishikawa J, Matsuyama A, Ouchi N, Kihara S, Funahashi T, Tenner AJ, Tomiyama Y, Matsuzawa Y. 2000. Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages. Blood 96:1723–1732.

Yoneda C, Hirayama Y, Nakaya M, Matsubara Y, Irie S, Hatae K, Watabe S. 1999. The occurrence of two types of collagen proα-chain in the abalone *Haliotis discus* muscle. Eur J Biochem 261:714–721.

Zatterstrom UK, Felbor U, Fukai N, Olsen BR. 2000. Collagen XVIII/endostatin structure and functional role in angiogenesis. Cell Struct Funct 25:97–101.

Zhang Y, Suankratay C, Zhang XH, Lint TF, Gewurz H. 1999. Lysis via the lectin pathway of complement activation: minireview and lectin pathway enhancement of endotoxin-initiated hemolysis. Immunopharmacology 42:81–90.