Collagen family of proteins

MICHEL VAN DER REST*,†,[1] AND ROBERT GARRONE*,‡

*Institute of Biology and Chemistry of Proteins (CNRS-UPR 412), Villeurbanne, France; †École Normale Supérieure de Lyon, Lyon, France; and ‡Université Claude Bernard, Lyon, France

ABSTRACT **Collagen molecules are structural macromolecules of the extracellular matrix that include in their structure one or several domains that have a characteristic triple helical conformation. They have been classified by types that define distinct sets of polypeptide chains that can form homo- and heterotrimeric assemblies. All the collagen molecules participate in supramolecular aggregates that are stabilized in part by interactions between triple helical domains. Fourteen collagen types have been defined so far. They form a wide range of structures. Most notable are *1*) fibrils that are found in most connective tissues and are made by alloys of fibrillar collagens (types I, II, III, V, and XI) and *2*) sheets constituting basement membranes (type IV collagen), Descemet's membrane (type VIII collagen), worm cuticle, and organic exoskeleton of sponges. Other collagens, present in smaller quantities in tissues, play the role of connecting elements between these major structures and other tissue components. The fibril-associated collagens with interrupted triple helices (FACITs) (types IX, XII, and XIV) appear to connect fibrils to other matrix elements. Type VII collagen assemble into anchoring fibrils that bind epithelial basement membranes and entrap collagen fibrils from the underlying stroma to glue the two structures together. Type VI collagen forms thin-beaded filaments that may interact with fibrils and cells. —van der Rest, M.; Garrone, R. Collagen family of proteins. *FASEB J.* 5: 2814–2823; 1991.**

Key Words: collagens • extracellular matrix • gene families • connective tissues • protein structure • supramolecular aggregates

THE WORD COLLAGENE IS A FRENCH neologism from the 19th century meant to designate the constituent of connective tissues that produces glue (from the Greek, κολλα, glue, and γενος, birth). The English adaptation "collagen" was first used around 1865. The Oxford Dictionary (1893) defines collagen as "that constituent of connective tissue which yields gelatin on boiling." Much of the early work was indeed done on heat-denatured collagen. On the other hand, the presence of fibers in connective tissues had been known since the 19th century from the work of early histologists such as Henle and Ranvier. Only in the 1920s did the pioneering work of Nageotte reveal that acid-solubilized collagen could precipitate into a material, later shown by x-ray diffraction and electron microscopy to be collagen fibers (1).

The existence of a monomeric building unit for the collagen fiber was postulated and called tropocollagen by Gross (2). This unit protein, now referred to as a collagen molecule, was shown to be composed of three polypeptide chains, two identical α1 chains and one distinct α2 chain, assembled in a triple helix with a coiled coil conformation. The primary structure of this protein is made for the most part of repeating Gly-Xaa-Yaa triplets with a high content of imino acids, Pro in the Xaa position and OH-Pro in the Yaa position. An extraordinarily complex posttranslational processing was shown to be needed for the maturation of the translation product into the fibrillar material with high tensile strength characteristic of collagen-rich tissues like tendon or skin (3).

At the end of the 1960s, it was found that cartilage contained a collagen made of polypeptides with a primary structure distinct from the collagen molecule isolated from the other tissues studied thus far (primarily skin and tendons). It was called type II collagen and the molecules were shown to be homotrimers (4). They are very similar in size and properties to the type I collagen molecule studied originally. Another homologous homotrimeric molecule was soon isolated from fetal skin and reported as the type III collagen molecule. Contrary to types I and II collagen, it is disulfide bonded but otherwise it is very similar to the first two types. Type III collagen molecules were isolated from tissues that also contained type I collagen, and it was later shown that molecules from the two types could form heterotypic (mixed) fibrils (5). This led to a certain confusion in terminology as the term collagen was used to designate a substance and a histological structure, whereas the collagen types now designate proteins with different amino acid sequences that are sometimes involved in the same histological structures.

The nomenclature problem was further complicated by the discovery of type IV collagen. In the original purification scheme involving solubilization after limited digestion by pepsin, the major collagenous component isolated from basement membranes had a size and amino acid composition similar in many ways to those of the other collagen types described thus far (6). Further work showed that the type IV collagen molecules differ from the first three types much more than initially thought. This is consistent with the observation that this collagen takes part in a very different histological structure, forming a sheet made by a meshwork of filaments rather than by linear fibrils. Since then, other molecules have been called collagens and been given a type number (14 at the time of writing of this review). The generally accepted definition of a collagen molecule now is that it is a structural protein of the extracellular matrix (ECM)[2] which contains one or more domains having the conformation of a collagen triple helix.[3] This definition excludes several proteins such as the complement component C1q, acetylcholine esterase, the mannose-binding protein, a lung

[1]To whom correspondence should be addressed at: IBCP, Bâtiment 403, Université Claude Bernard, 43, Bd du 11 Novembre 1918, 69622 Villeurbanne CEDEX, France.

[2]Abbreviations: ECM, extracellular matrix; RDEB, recessive dystrophic epidermolysis bullosa; FACITs, fibril-associated collagens with interrupted triple helices.

[3]Throughout this review we will use the terms triple helix and triple helical to designate the conformation found in collagen triple helices. The reader is reminded that triple α helices are found in some proteins, sometimes also containing collagen-like triple helices (60).

0892-6638/91/0005-2814/$01.50.

surfactant or the scavenger receptor of the macrophage, and possibly several others that contain triple helical domains but do not participate in the assembly of the ECM.

Collagen molecules are multidomain proteins. In this respect they are no different from most proteins, and the collagen triple helix should be viewed as one among the other conformations found in proteins together with α helices and β sheets (7). Nature has used this particular conformation in quite different molecules, and to understand this diversity it is best to consider first the structure and role of triple helices.

PROPERTIES AND FUNCTIONS OF TRIPLE HELICES

The collagen triple helix is characterized by the intertwining of three helical polypeptides forming a coiled coil structure (**Fig. 1**). This structure is now well characterized. Each polypeptide forms a left-handed helix in which every third residue comes into the center of the superhelix, shifted by 30° from the preceding central residue of the same chain. This results in the formation of a right-handed superhelix. Steric reasons impose that the center of the helix be occupied only by glycyl residues. The amino-acid sequences of triple helical domains are thus characterized by the repetition of triplets Gly-Xaa-Yaa. Any other amino acid sequence would perturb the triple helical conformation (3). Many triple helical domains actually contain some irregularities in the triplet repeats and these imperfections are thought to alter the otherwise uniform conformation of triple helical domains (5). The triple helical conformation is stabilized by the presence of imino acids in the Xaa and Yaa positions. About 30% of these positions are occupied by prolyl and hydroxyprolyl residues, respectively.

The triple helical domains may thus serve as molecular rods that can physically separate globular domains in a protein. The triple helix has still another important and unique property. From its structure it can easily be visualized that the amino acids in the Xaa and Yaa positions have their side chains pointing outward of the helix, thus at the surface of the protein (Fig. 1). Two-thirds of these are neither prolines nor hydroxyprolines. This offers an exceptional potential for lateral interactions, particularly with other triple helices. It will be seen from the remainder of this discussion that these two properties of the triple helices—to be molecular rods and to have a high potential for lateral interactions—are crucial to understanding the functions of the various collagen molecules.

COLLAGENS PARTICIPATING IN QUARTER-STAGGERED FIBRILS

The presence of striated fibrils in ECM has been recognized from the very first electron microscopy observations of collagen-containing tissues (8). The interpretation of this striation as a manifestation of the alignment of staggered ~300-nm collagen molecules with gaps and overlaps (**Fig. 2**) was presented in 1962 by Hodge and Petruska (9). It is now clear that lateral interaction between homologous regions within the triple helical domains is the basis for fibril formation. The molecules are staggered by ~67 nm (a length called D in collagen structure jargon). Internal homologies within the collagen molecule have been noted and could be related to D units.

The banding pattern of collagen fibrils is very similar from tissue to tissue. This reflects the presence of D-staggered collagen molecules. There are, however, important differences in the three-dimensional packing of the molecules. These differences have been detected by x-ray diffraction analysis (10) and by careful chemical analyses of the naturally occurring lysine-derived cross-links that stabilize the fibrillar aggregates (11). There are also important variations in the diameters of the fibrils. These differences can be explained, at least in part, by the existence of several distinct but structurally very homologous collagen molecules involved in fibril formation. These collagens are often referred to as the fibrillar collagens (**Table 1**) and all their genes are clearly derived from a single ancestral gene (12). Although these molecules have all been shown to be able to form fibrils by themselves, they appear to participate in heterotypic fibrils in vivo, i.e., fibrils made of more than one collagen type (13–16). Collagen fibrils may thus be looked at as molecular alloys.

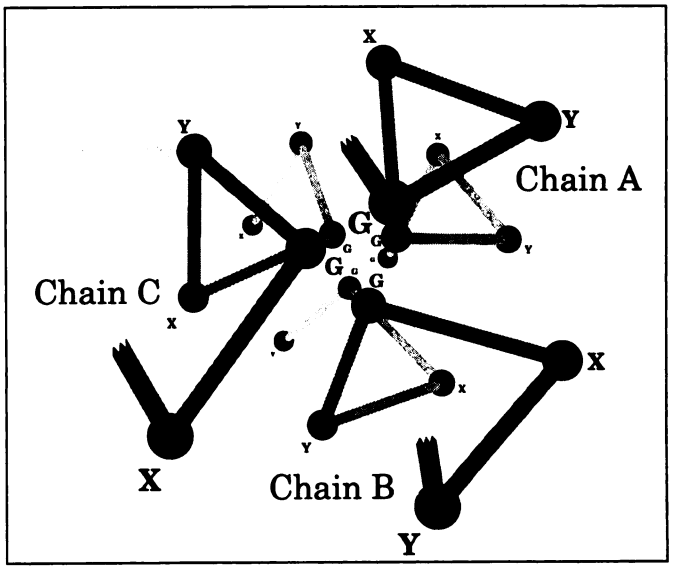

Figure 1. Schematic representation of a cross-section through a collagen triple helix. The balls represent the α carbons of the aminoacyl residues and the sticks represent the peptide bonds. The side chains of the residues are omitted. It should be noted that all the X and Y residues are at the surface and will have their side chains pointing outward of the helix. Each polypeptide forms a left handed helix and assemble with the other two chains in a right handed superhelix. (Reprinted from ref 7 with permission.)

The homotrimeric type II collagen molecule, $[\alpha 1(II)]_3$[4] was discovered in cartilage by Miller and Matukas in 1969 (4) and is the major collagenous constituent of this tissue. Most other connective tissues contain type I collagen molecules, $[\alpha 1(I)]_2\alpha 2(I)$, as their major structural elements. After type III collagen, $[\alpha 1(III)]_3$, was discovered, it was first suggested that this disulfide-bonded molecule would form thin fibrils of its own while type I collagen would form thicker fibrils (17). The demonstration of covalent lysine-derived cross-links between type I and type III collagen molecules

[4]The collagen nomenclature is based on the following rules. The individual polypeptide chains are called α chains and are numbered in arabic numerals for a given collagen type, which is itself indicated by Roman numerals in parentheses. It is assumed that chains of a given type assemble to form trimeric molecules. Several stoichiometries can be observed within a single collagen type. The numbering of collagen types and chains is arbitrary and reflects mostly the historical sequence of their biochemical characterization.

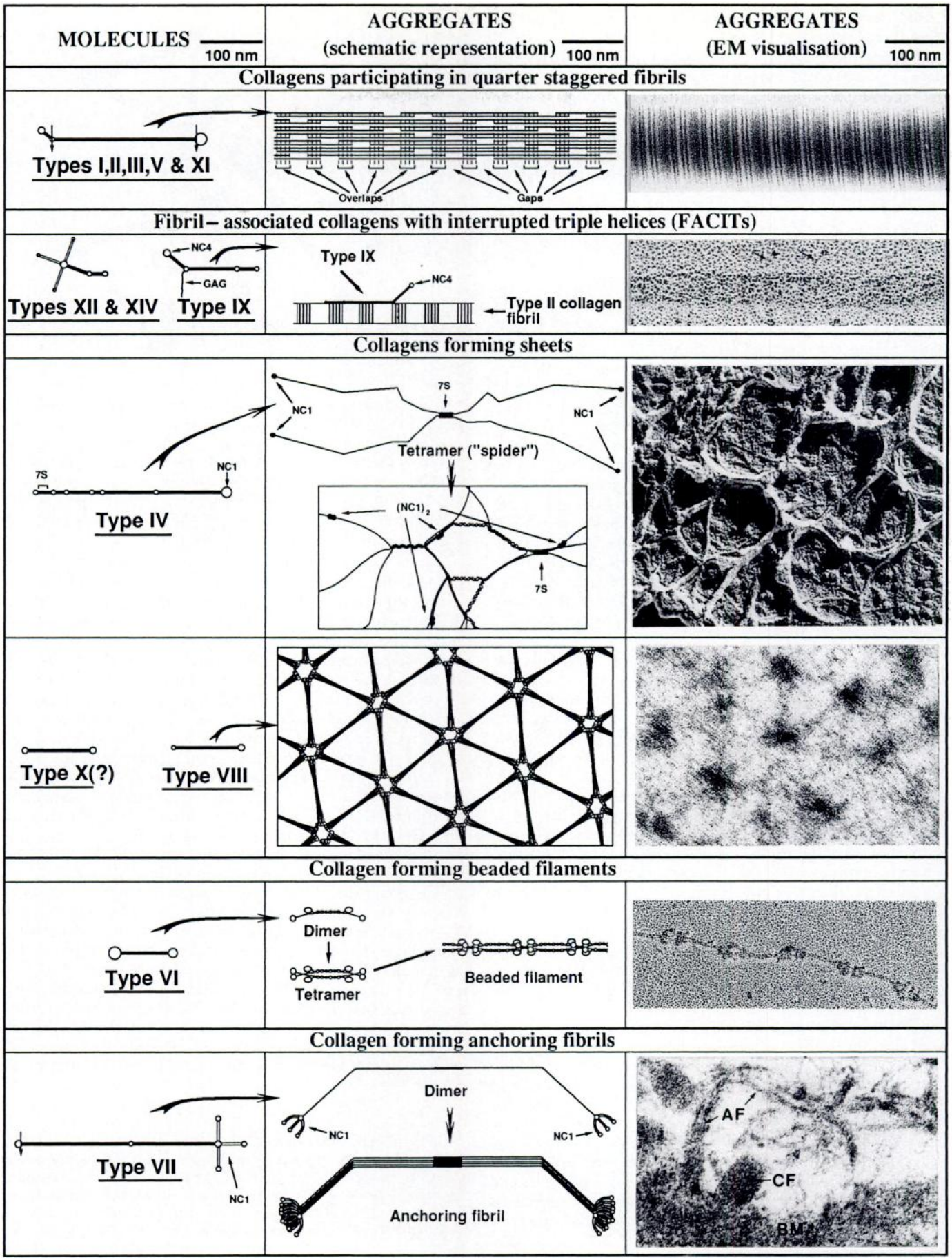
MOLECULES
100 nm
AGGREGATES
(schematic representation)
100 nm
AGGREGATES
(EM visualisation)
100 nm
Collagens participating in quarter staggered fibrils
Types I,II,III,V & XI
Overlaps
Gaps
Fibril– associated collagens with interrupted triple helices (FACITs)
Types XII & XIV
NC4
GAG
Type IX
Type IX
NC4
Type II collagen fibril
Collagens forming sheets
7S
NC1
NC1
Tetramer ("spider")
Type IV
7S
NC1
(NC1)2
7S
Type X(?)
Type VIII
Collagen forming beaded filaments
Type VI
Dimer
Tetramer
Beaded filament
Collagen forming anchoring fibrils
Type VII
NC1
Dimer
NC1
NC1
Anchoring fibril
AF
CF
BM

TABLE 1. *Fibrillar collagens*

Type	Chains	Molecules	Representative tissues
		Associated with type I	
I	α1(I), α2(I)	$[\alpha1(I)]_2$ α2(I)	Skin, bone, tendon, dentin, etc.
		$[\alpha1(I)]_3$	Dentin, skin (minor form)
III	α1(III)	$[\alpha1(III)]_3$	Skin, vessels
V	α1(V),[a] α2(V), α3(V)	$[\alpha1(V)]_3$	Hamster lung cell cultures
		$[\alpha1(V)]_2$ α2(V)	Fetal membranes, skin, bone
		α1(V) α2(V) α3(V)	Placenta, synovial membranes
		Associated with type II	
II	α1(II)	$[\alpha1(II)]_3$	Hyaline cartilage, vitreous body
XI[b]	α1(XI), α2(XI), α3(XI)[c]	α1(XI) α2(XI) α3(XI)	Hyaline cartilage

[a] A chain similar in its triple helix, but different in its propeptides has been described and called α1(V)' or α4(V). [b] Often called 1α2α3α. [c] α3(XI) is probably identical to α1(II), except for posttranslational modifications.

was the first evidence that the two molecules may actually be part of the same aggregates (13). It is only recently that thorough immunoelectron microscopy studies have demonstrated that in a tissuelike skin all the fibrils are actually made of type I and type III collagens (15).

Two other fibrillar collagen types were described initially by the group of R. Burgeson and the late D. Hollister (5). They have been called type V and type XI collagens. Most studies of these molecules have been done after solubilization with pepsin. This enzyme cleaves in the nontriple helical regions where the covalent cross-links responsible for collagen insolubility are found. The triple helix itself is resistant to all proteolytic cleavages except collagenases. The resulting triple helical domains of types V and XI collagens have properties similar to those of collagen types I, II, and III. Type V collagen has been extracted from tissues containing type I collagen and type XI collagen from tissues containing type II collagen, leading to a dichotomous view of

Figure 2. Molecular structures and supramolecular assemblies of collagens. This figure combines schematic scale representations and electron microscope micrographs of molecules and aggregates of various collagen types. Schematic representations of collagen molecules are on the left panel with the molecules oriented with their NH_2 termini to the left. Triple helical (COL) domains are drawn as thick black lines and nontriple helical (NC) domains are represented by double lines or empty circles. The larger circles at the extremities of the molecules correspond to the globules visible on rotary shadowing electron micrographs of isolated intact collagen (or procollagen) molecules. Vertical arrows (drawn on the molecules depicted on the upper and lower panels) indicate the sites of action of the processing proteinases. GAG points to the glycosaminoglycan chain attached to the α2 chain of type IX collagen molecule; 7S is the domain of antiparallel interaction of the type IV collagen triple helices to form a tetramer (spider). The middle panel contains drawings of collagen aggregates and assemblies. The right panel illustrates actual collagen aggregates as observed by electron microscopy. *Types I, II, III, V, and XI* collagen molecules (left panel) have the same overall molecular structure. The indicated processing sites have been proved only for collagen types I, II, and III (see text). They participate in fibril-forming lateral aggregates. These fibrils are heterotypic. In this aggregation process, the molecules have the same orientation but are staggered by 67 nm (middle panel), yielding regions of maximal density (overlaps) and regions with holes (gaps). This alternate pattern is clearly visualized by transmission electron microscopy. The superposition of similarly charged residues reproduce periodically along a fibril, giving rise to a characteristic striated pattern (right panel). These residues interact rather specifically with the reagents (metal salts) used to obtain positive staining of collagen fibrils. The cross-striated fibril shows a periodic pattern (which can vary according to the reagents used) with a constant 67-nm periodicity (courtesy S. Franc). *Type IX and type XII* collagen molecules contain interrupted triple helical domains and large NH_2-terminal domains (left panel). Type IX collagen binds to the surface of a type II collagen fibril (middle panel). The relative orientation of the molecule has not been defined yet and is represented here antiparallel to type II collagen. The globular NC4 domain and the triple helical COL3 domain project out of the fibril. Actual assemblies of type IX collagen (arrows) around a type II collagen fibril from cartilage are visualized on an electron micrograph after rotary shadowing (right panel, courtesy A. Barge and F. Ruggiero). *Type IV* collagen molecules, containing several interruptions in the triple helix (left panel) assemble into tetramers (spiders) via their 7S region (middle panel, upper drawing). In basement membranes, tetramers interact by their globular domains NC1, then forming double NC1 globules, and by their arms (triple helical domains), which can wrap around each other (middle panel, lower drawing). The result of this association is a complex, branching network. It is illustrated on the right panel which shows an electron micrograph of a basement membrane secreted by an embryonic chick corneal epithelium in culture, prepared by quick-freezing, deep-etching, and shadow-casting (courtesy A. Barge). *Type VIII* collagen molecules, dumbbell-shaped (left panel), are thought to interact laterally and by their extremities to form regular hexagonal lattices in the basement membrane of the corneal endothelium (Descemet's membrane), as drawn in the middle panel and as shown on an electron micrograph (right panel). *Type VI* collagen molecule has a short triple helical domain with prominent globular ends (left panel). It forms dimers by antiparallel association and tetramers by lateral association of dimers. Association of tetramers by their extremities forms beaded filaments (middle panel) easily recognizable after rotary shadowing (right panel, courtesy C. Lethias). *Type VII* collagen molecules have a COOH-terminal domain (NC1) with three fingers (left panel). After processing of the NH_2-terminal globular domain, they form dimers by antiparallel, partial overlapping of their amino-terminal domain. Several dimers aggregate laterally in register to form anchoring fibrils ended by a large bunch of carboxyl-terminal domains (middle panel). These COOH-terminal domains interact with basement membranes (BM) and anchoring fibrils (AF) form loops entrapping collagen fibrils (CF), as shown on the transmission electron micrograph in the right panel. In this case, the anchoring fibril forms an almost complete loop and the collagen fibril is shown in cross section.

two symmetric systems, one around type I collagen associated with type III and type V collagens and the other around type II collagen associated with type XI collagen.

The situation is certainly much more complex. First, several molecular assemblies—$[\alpha 1(V)]_3$, $[\alpha 1(V)]_2\alpha 2(V)$, and $\alpha 1(V)\alpha 2(V)\alpha 3(V)$—have been demonstrated for type V collagen molecules (5) whereas only heterotrimers $\alpha 1(XI)\alpha 2(XI)$ $\alpha 3(XI)$ have been described for type XI collagen (18). The $\alpha 3(XI)$ chain is most probably a product from the same gene, COL2A1, as is the $\alpha 1(II)$ chain. Several pieces of evidence indicate that the $\alpha 1(XI)$ chain, in the absence of the other type XI chains, is expressed in some tissues, such as bone (19) and placenta (20), containing type I collagen as their major collagen type. The existence of hybrid molecules made of mixed type V and type XI chains has been proposed (19). A fourth chain of type V collagen [$\alpha 1'(V)$] has been reported (5). It may actually represent a type XI chain in a different molecular assembly. Recent sequence data have demonstrated a remarkable degree of primary structure similarity between the $\alpha 1(V)$, $\alpha 3(V)$, $\alpha 1(XI)$, and $\alpha 2(XI)$ collagen chains ((20, 21), and van der Rest, M., unpublished results). It thus appears that the delineation between type V and type XI collagens is not as clear as the nomenclature might imply.

All the fibrillar procollagens studied so far have the same overall structure (Fig. 2). They comprise a main triple helical domain of about 337 Gly-Xaa-Yaa repeats. At the carboxyl end, a disulfide-bonded globule (the C-propeptide) serves as recognition site for trimer assembly and also prevents premature fibril assembly. It is separated from the main triple helix by a short linear domain of about 30 residues which contain the C-proteinase cleavage site. At the amino end, a second short (33-54 residues) triple helical domain is separated from the main triple helix by a nontriple helical domain which contains the N-proteinase cleavage site, at least in types I, II, and III procollagens. This short triple helix is bordered at the amino end by a highly variable globular region. This region is absent in the pro $\alpha 2(I)$ chain and is up to 383 residues in the pro $\alpha 1(XI)$ chain (22).

The structural similarities between fibrillar collagens is reflected at the gene level (12). They show indeed a remarkable conservation of the intron-exon structures. In particular, the predominance of exons of 54 bp or of multiples of 54 bp is striking. The high conservation of the gene structure, including the C-propeptide region, can be traced from mammals to the most evolutionarily distant species such as sponges (23) and sea urchins (24). Recent data from J. Y. Exposito in our laboratory have revealed that the intron-exon structure of the gene of a sponge fibrillar collagen can be aligned with precision to vertebrate counterparts (ref 23 and unpublished results).

The proteolytic processing of the procollagens is well described for types I, II, and III procollagens, which are processed by the N- and C-proteinases into the functional fibril-forming collagen molecules (25, 26). The released N-propeptide of type I collagen has been shown to have an inhibitory effect on procollagen mRNA translation. For type III collagen, the processing at the NH_2 terminus is slower and handled by a distinct enzyme (25). For type V and type XI collagens, a significant globular region remains uncleaved (5, 18). It has been visualized in type V collagen by rotary shadowing electron microscopy. Fessler and Fessler (see chapter 3 in ref 5) have studied in considerable detail the biosynthesis of chicken type V collagen and have proposed a model for type V collagen processing involving several proteolytic steps. The direct correlation of this model with the cDNA-derived sequences remains to be established.

The functional reason for the existence of this molecular diversity in the fibrillar collagen molecules is still very much a matter of speculation. The higher O-glycosylation of hydroxylysyl residues in type II collagen compared with type I collagen may be the molecular basis for the larger volume occupied by collagen molecules in cartilage fibrils (10). In turn, this higher level of posttranslational modifications may be due, at least in part, to a slower folding of the triple helix of type II collagen. Lysyl hydroxylase and the glycosyl transferases, like prolyl hydroxylase, are active only on nontriple helical chains, and the level of hydroxylation and glycosylation is thus a function of the rate of helix formation (26).

Experimental evidence with reconstituted fibrils has shown that the presence of type III and type V collagen molecules has a direct effect on fibril diameter (27). Immunolocalization experiments have shown that the epitopes recognized by anti-type V collagen antibodies are masked and that unmasking can be obtained by swelling of the fibrils (14). Similarly, epitopes recognized by anti-type XI collagen antibodies could be detected only in disrupted cartilage fibrils at the point of disruption (16). This indicates that these collagens are located within heterotypic fibrils and might actually constitute the core of fibrils. Furthermore, the highly variable structure of the N-propeptide region of fibrillar collagens suggests that a major biological reason for the diversity of these molecules resides in this amino-terminal domain (28).

Recent data on primitive fibrillar collagens in sponges (23) and sea urchins (24) have revealed that the structure of invertebrate fibrillar collagens, including their C-propeptides, are highly homologous to vertebrate collagens. Homology analyses show that the sponge fibrillar collagen is significantly more similar to vertebrate type XI collagen than to any other vertebrate fibrillar collagen. Because the fibrillar sponge collagen is characterized by extremely thin (20 nm) fibrils, it is tempting to speculate that type V and type XI collagens also form thin fibrils, serving as cores for the larger fibrils after apposition of type I or type II collagen molecules.

FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED TRIPLE HELICES (FACITs)

Types IX, XII, and XIV collagens belong to the subgroup of molecules called FACITs by Olsen (29). They are nonfibrillar collagens (**Table 2**) as they do not form by themselves quarter-staggered fibrils. The structure of these molecules can be divided into three main functional regions. One region comprises one or two triple helical domains and serves for the interaction and adhesion of these molecules to the fibrils. A second region, comprising another triple helical domain, serves as a rigid arm that projects out of the fibril a third region, which does not include triple helices and may serve for interaction with other matrix elements or with cells. The various triple helical domains (COL domains) are separated (or interrupted) by short nontriple helical domains (NC domains).

Type IX collagen is the best-characterized molecule of this group (5). It is found in ECMs containing type II collagen as their main fibril-forming structure, such as hyaline cartilage and the vitreous body of the eye. It is made of three distinct polypeptide chains and is thus a heterotrimer $\alpha 1(IX)\alpha 2(IX)\alpha 3(IX)$. It comprises three triple helical domains (see Fig. 2). COL1 and COL2 appear to be involved in the interaction with the fibrils and COL3 serves as the arm sticking out of the fibril. Four NC domains intersperse the triple helices. Most of them are very short (12–30 amino

TABLE 2. *Nonfibrillar collagens*

Type	Chains	Molecules	Representative tissues
IV	α1(IV), α2(IV) α3(IV), α4(IV), α5(IV)	$[\alpha1(IV)]_2$ α2(IV) (?)	Basement membranes Glomerular basement membrane
VI	α1(VI), α2(VI), α3(VI)	α1(VI) α2(VI) α3(VI)	Vessels, skin, intervertebral disc
VII	α1(VII)	$[\alpha1(VII)]_3$	Dermoepidermal junction
VIII	α1(VIII), α2(VIII)	(?)	Descemet's membrane, endothelial cells
IX	α1(IX), α2(IX), α3(IX)	α1(IX) α2(IX) α3(IX)	Hyaline cartilage, vitreous humour
X	α1(X)	$[\alpha1(X)]_3$	Growth plate
XII	α1(XII)	$[\alpha1(XII)]_3$	Embryonic tendon and skin, periodontal ligament
XIII	α1(XIII)	(?)	Endothelial cells
XIV	α1(XIV)	$[\alpha1(XIV)]_3$	Fetal skin and tendon

acids) except for the NH_2-terminal NC4 domain of the α1(IX) chain, which is 243 residues long in cartilage. However, in some other tissues such as the primary stroma of the cornea, this domain is only a few amino acids long, because with the signal peptide it contains 25 residues. This is due to the use of an alternate promoter and first exon which are localized in an intronic sequence of the gene used for the cartilage transcript (between exons 6 and 7) (30).

The NC3 domain of the α2(IX) chain contains a unique five amino acid insertion (when compared to the α1(IX) chain) which defines an attachment site for a glycosaminoglycan side chain (5). The size of this side chain is highly variable. From nonexistent to small in cartilage, it is very large (31) in the avian vitreous humor where it might account for the total glycosaminoglycan content of the tissue. The possible role of this side chain in cartilage is still unknown.

The NC4 domain of the chicken cartilage α1(IX) chain has an estimated pI of 9.7 (10.55 for the human equivalent). It is therefore thought to interact with the acid proteoglycans found in the cartilage matrix (32). Its absence in some tissues suggests that type IX collagen may have other functions. Studies of the avian vitreous humor have shown that type IX collagen molecules are devoid of NC4 domains in this tissue (31). Here the major function of type IX collagen may therefore be the attachment of the large glycosaminoglycan side chain to the type II collagen fibrils. In cartilage and probably in other tissues as well, the interaction between type IX and type II collagens is stabilized by the formation of covalent cross-links between the two molecules (33, 34).

The role of type IX collagen in ECMs containing type II collagen is thus to serve as means of attachment of new functionalities to the fibrils. The need for such a function of course is not unique to cartilage. On this basis, Gordon et al. (35) have successfully searched for cDNA clones of homologs of type IX collagen in ECMs containing type I collagen as their major fibrillar constituent. A short cDNA was cloned and characterized. The sequence of this first partial cDNA clone revealed striking similarities with the COL1 domain of type IX collagen, namely, identically located cysteines and triple helix imperfections. The further characterization of the intact protein molecule, called type XII collagen, and of extended cDNAs revealed major differences between the two molecules (36). Type XII collagen is a homotrimer $[\alpha1(XII)]_3$ and contains only two triple helical domains, COL1 and COL2. The nontriple helical NH_2-terminal NC3 domain is very large (3×190 kDa) and gives the molecule a characteristic cross-shape that has been visualized by rotary shadowing experiments (see Fig. 2). A portion of this domain is homologous to the NC4 domain of the cartilage α1(IX) chain.

Several pieces of evidence suggest that type XII collagen interacts with fibrils containing type I collagen. The precise nature of this interaction remains to be elucidated. The homology between the COL1 domains of type IX and type XII collagens suggests that this domain may play a role in this interaction. The presence of such a COL1 domain could therefore be an indication of a FACIT molecule.

We recently demonstrated the existence in skin and tendon of another homotrimeric molecule (type XIV collagen) with the characteristic FACIT COL1 domain (37). The characterization of the molecule at the cDNA and protein levels indicates that it is very similar but clearly distinct from type XII collagen (M. K. Gordon, P. Castagnola, B. Dublet, T. F. Linsenmayer, M. van der Rest, and B. R. Olsen, unpublished results).

COLLAGENS FORMING SHEETS

Besides the collagens involved in fibril formation and interactions, several other subgroups of nonfibrillar collagens have been discovered. Among these some are involved in the formation of sheets or protein membranes that surround tissues and organisms, such as basement membranes, Descemet's membrane, cuticle of worms, and organic skeleton of sponges. Any structure conservation for these collagens is far from being as evident as it is for fibrillar collagens. For example, the main constituent of the hexagonal lattice of Descemet's membrane (type VIII collagen) bears no clear similarity to basement membrane type IV collagen. On the other hand, it is closely related to type X collagen, which is synthesized by hypertrophic chondrocytes and whose function and mode of aggregation is still a matter of controversy. Because of this homology we have included type X collagen in this part of the discussion, but the reader should be aware that there is no evidence that type X collagen would form a sheet structure.

Basement membrane type IV collagen (or collagens)

Basement membranes are formed by a fine network of cords that are laced together and entrap large associated molecules

such as the laminin-nidogen complex and the heparan sulfate proteoglycan (38). This network is made of type IV collagen. The existence of two constituent chains—α1(IV) and α2(IV)—forming a [α1(IV)]$_2\alpha$2(IV) heterotrimer has been known for some time. Clear evidence for the presence of three additional chains in some basement membranes has been obtained through the partial characterization of two chains bearing the epitopes responsible for Goodpasture syndrome, an autoimmune disease affecting particularly the glomerular basement membrane in the kidney, and through the cloning of a cDNA encoding another basement membrane chain (39, 40). These three chains have been given the names α3(IV), α4(IV), and α5(IV), respectively, although there is no evidence that they can form heterotrimers with either α1(IV) or α2(IV). The molecular assemblies remain to be established. The α5(IV) chain has recently been shown to be defective in Alport syndrome, a heritable condition resulting in severe kidney disfunction (40).

The structure of collagen assembly in basement membranes was recently reviewed in these pages by Yurchenko (41) and will not be discussed here. It should be remembered that this assembly process requires lateral interactions between triple helices at two levels. In the 7S domain, four triple helical domains from four different molecules assemble head to tail to form the structure known as a spider whose legs are made of a long triple helical region made of several triple helical domains interrupted by short nontriple helical sequences. In the tissue the molecules interact laterally by this region in a staggered fashion. Finally, interactions of nontriple helical COOH-terminal NC1 domains to form dimers are responsible for tail-to-tail binding of type IV molecules (Fig. 2).

Type IV collagen has been characterized in several invertebrates and appears highly conserved during evolution. In particular, *Drosophila* type IV collagen has been extensively studied at the genomic and biochemical levels (42, 43). Contrary to the multiple small exons separated by large introns described for vertebrate type IV collagen genes, the *Drosophila* type IV collagen gene is composed of nine large exons separated by relatively small introns (43). In spite of this genomic difference, *Drosophila* and vertebrate type IV collagen chains share a similar general organization. However, the distribution of several cysteine residues along the chain of *Drosophila* type IV collagen led to the hypothesis that it could make both network and microfibrillar supramolecular assemblies (43).

Descemet's membrane type VIII collagen

Type VIII collagen was initially reported by Sage et al. (see chapter 6 in ref 5) in endothelial cell culture and first called EC collagen. The structure that was proposed was that of a large molecule with three different sizes (EC1, 177 kDa; EC2, 125 kDa; EC3, 100 kDa) appearing at different times of culture. These products are very sensitive to proteolysis and are readily cleaved into fragments of 50 kDa. A cassette model was proposed in which a parent polypeptide chain of 175–180 kDa participated in three triple helical domains separated by nontriple helical domains. The lower molecular weight forms would be derived from the parent molecule by proteolytic cleavage.

Another model was proposed by Benya and Padilla (44) in which type VIII collagen would be synthesized in the form of a smaller molecule with 61-kDa chains. This latter model was proved to be correct for the molecule found in Descemet's membrane. Full-size cDNAs for two homologous triple helical polypeptide chains, which have been called α1(VIII) and α2(VIII), have been identified by amino acid sequencing to the major collagenous constituent of Descemet's membrane (45, 46). It is still difficult, however, to fully reconcile the data obtained from cDNA analysis and the original data on the collagen synthesized by endothelial cells. There is a possibility that these may actually represent two distinct molecules. Because the correlation between EC chains and the main constituent of Descemet's membrane rests heavily on immunological evidence, particularly immunoprecipitation, this would imply that some epitopes would be shared between EC chains and Descemet's type VIII collagen. If the molecules are identical gene products it would be necessary to invoke very early cross-linking reactions between type VIII chains to form higher molecular weight molecules.

The structure derived from the cDNA-derived sequence of the α1(VIII) chain is that of a molecule with a single triple helical domain (454 amino acids/chain) containing eight imperfections of the Gly-Xaa-Gly-Xaa-Yaa type and bordered by two relatively short nontriple helical domains, 117 residues at the amino end and 173 at the carboxyl end (45). The α2(VIII) chain has a similar structure (46). Its triple helical domain is 457 residues long with identically located imperfections (two of these are of the Gly-Xaa-Yaa-Xaa-Yaa type). The COOH-terminal domain is only 167 residues long. The size of the NH_2-terminal domain is not known yet. It has been suggested that the stoichiometry of the chains might be [α1(VIII)]$_2\alpha$2(VIII) based on the chain ratio observed in the tissue but this will require further evidence to be firmly established.

Descemet's membrane, which separates the corneal endothelial cells from the stroma, consists of stacks of hexagonal collagen lattices made of type VIII collagen (47). This lattice is built by nodes interconnected by rodlike structures with a dumbbell appearance (Fig. 2). The structures observed at the nodes of the lattice appear quite large if the nontriple helical domains of type VIII collagen were the only constituents. It is therefore possible that other macromolecules contribute to the organization of this lattice.

As opposed to the highly fragmented genomic organization of other collagen genes, the triple helical domain and the COOH-terminal globular domain of the α1(VIII) chain are encoded in a single large exon (48).

Type X collagen

The amino acid sequence, gene structure, and molecular organization of type X collagen are extremely similar to those of type VIII collagen. This molecule is a homotrimer [α1(X)]$_3$. The triple helical domain is 460 residues long and is also marked by the presence of eight imperfections at locations similar to those of type VIII collagen. Three imperfections are of the Gly-Xaa-Yaa-Xaa-Yaa type. The COOH-terminal domain is 162 residues long and the NH_2-terminal domain is only 52 residues long (see chapter 8 in ref 5).

Type X collagen has a very restricted pattern of distribution, being synthesized primarily by hypertrophic chondrocytes during the process of endochondral ossification. It is cleaved at two sites by vertebrate collagenase. This enzyme initiates the degradation of collagen types I, II, and III by cleaving them at 3/4ths of their length. The resulting fragments then denature at physiological temperature and get degraded by general proteolysis. For type X collagen, however, the melting temperature of the major 32-kDa triple helical fragment is still 43°C, only 4°C below the T_m of the intact molecule. This fragment would thus remain triple helical at physiological temperature. This raises questions about the significance of this collagenase cleavage in the degradation of type X collagen.

Immunolocalization at the EM level has shown that it is associated with very fine fibrillar structures that are also labeled by anti-type II collagen antibodies and with poorly organized mat structures (49). This latter structure resembles the hexagonal lattice of Descemet's membrane. It would not be surprising that the close primary structure homology between type VIII and type X collagens would be reflected at the level of their ultrastructural organization.

Invertebrate sheet forming collagens

Some collagenous structures appear as external sheets or stacked layers forming protective or adhesive devices. They are generally encountered in invertebrates, although the egg case of sharks belongs to these collagenous exoskeletons. The cuticle of annelids is the best known of these collagens from biochemical studies (50). The body of those worms is surrounded by several layers (with often a plywood organization) of large, nonbanded collagen fibers. The unit molecules can reach up to 2400 nm in length and are the longest extended protein molecule known. They are insensitive to pepsin proteolysis and their amino acid composition reveals several peculiarities, i.e., a high ratio of hydroxyproline to proline and an absence of cysteine or lysine-derived cross-links.

The other collagens of this category are known from cDNA sequences and partial biochemical data. The cuticle of nematods, which is largely microfibrillar in structure, contains several collagenous proteins that vary according to the developmental stage of the animal (51). Basically, these molecules are composed of several triple helical domains separated by nontriple helical domains. The cuticle collagens of nematodes are encoded by a large (up to 150) gene family. Sponges secrete a microfibrillar, collagenous layer which serves to attach the animal to its substratum (basal spongin) and expands inside the body, forming the organic skeleton (spongin fibers). Several cDNA clones corresponding to these proteins have been sequenced (J.-Y. Exposito, D. Le Guellec, Q. Lu, R. Garrone, unpublished results). The encoded collagen chains have an overall organization very similar to nematode cuticle collagens (and in some respects to human type XIII collagen) and a carboxyl-terminal, nontriple helical domain presenting some homology to the same region of type IV collagen. These sponge collagens are encoded by a family of closely related genes (J.-Y. Exposito et al., unpublished results).

TYPE VI COLLAGEN, FORMING BEADED FILAMENTS

Type VI collagen is made of three polypeptide chains, α1(VI), α2(VI), and α3(VI), assembled into a heterotrimeric molecule (see chapter 4 in ref 5). The primary structure of the protein has been completely described through protein and cDNA sequence analysis (52, 53). The protein contains a proportionally very short unique triple helical domain (335 or 336 amino acid residues/chain). A single cysteine/chain is found inside this domain. It is involved in the stabilization of the supramolecular aggregate, whereas clusters of cysteine residues, probably forming intramolecular bonds, are found at the extremities of the triple helix. Like most other nonfibrillar collagens, it contains two imperfections of the triple helix. It is characterized by the presence of 11 arginine-glycine-aspartate (RGD) sequences for the three chains. Such sequences have been shown to play a central role in the interaction of matrix constituents, in particular fibronectin, with cell receptors of the integrin type.

The nontriple helical amino and carboxyl regions are large compared with the triple helix. In the α1(VI) and α2(VI) chains, they are 234-237 residues and 429-436 residues long, respectively. In the α3(VI) chain the predicted sizes of the intact human NH_2-terminal and COOH-terminal regions are 1804 and 803 residues long, respectively. These domains are characterized by the presence of 200 residue repeats showing significant similarities to the A domains of von Willebrand factor. The COOH-terminal region of the α3(VI) chain contains, in addition to two of these domains, three additional domains showing similarities to salivary proteins, to fibronectin type III repeats, and to Kunitz-type protease inhibitors, respectively. The functions of these domains remain to be elucidated.

The protein actually expressed may vary from tissue to tissue as there is clear evidence of alternate splicing and proteolytic cleavage (54). Furthermore, the chains do not appear to be synthesized in a stoichiometric ratio in all tissues (55). The full understanding of this complex molecule will therefore require much further experimental studies.

The mode of aggregation of type VI collagen molecules has been elegantly demonstrated by rotary shadowing electron microscopy analysis of native and partially denatured or digested molecules (5). As shown in Fig. 2, two molecules aggregate in a head-to-tail orientation to form a dimer. Two dimers assemble into tetramers, which in turn form linear aggregates described as beaded filaments with repeats of approximately 110 nm. This complex assembly process must involve interactions between triple helices, between globular domains, and between globular domains and triple helices. These assemblies are stabilized by disulfide bonds and not by lysine-derived cross-links like most other collagens. The beaded filaments often aggregate laterally into bundles that conserve the 110-nm periodicity and therefore appear striated.

The function of type VI collagen is still largely a matter of speculation. Type VI collagen has been observed in most ECMs. The network of filaments is found together with classical collagen fibrils but appears to have independent orientations. They may play a role as an interface between the main collagen fibril network and the cells.

Increased type VI collagen synthesis or deposition has been noted in several pathological conditions, such as several fibrotic diseases and osteoarthrosis, and most notably in the cultured fibroblasts of a patient with cutis laxa, a condition characterized by a loss of the elastic properties of the skin (5). The genes coding for the α1(VI) and α2(VI) chains have been mapped to the extremity of the long arm of chromosome 21 (map location 21q22.3) (52). This is precisely the minimal region of triplication required for the expression of the phenotype of Down's syndrome (mongolism) in partial trisomy 21. It is well known that patients with Down's syndrome have marked connective tissue abnormalities, such as hypermobile joints. Whether this relates to an hyperexpression of type VI collagen remains to be established.

TYPE VII COLLAGEN, FORMING ANCHORING FIBRILS

Contrary to type VI collagen whose structure is very well known but whose function is not so clear, type VII collagen has a well understood function but very few data are available yet on its amino acid sequence. Its overall structure and mode of aggregation have been well described (see chapter 5 in ref 5). It is thought to be a homotrimer. It contains a very large ~420-nm long discontinuous triple helical region that

yields two fragments after pepsin digestion, the NH_2-terminal P1 fragments (94 kDa/chain) and the COOH-terminal P2 fragment (76 kDa/chain). A very large nontriple helical NC1 domain (3 × 150 kDa) is observed at the COOH-terminal end of the molecule. This domain is made of three ~50 nm arms terminated by small globules as seen on rotary shadowing electron micrographs. A smaller 100 kDa globular NC2 domain is observed at the amino end.

Type VII collagen, which is synthesized by keratinocytes (56), is assembled first into antiparallel dimers that overlap by 60 nm. The NH_2-terminal NC2 domain appears to be cleaved during this process. The dimers then laterally aggregate in a nonstaggered fashion, most probably by triple helix-triple helix interactions, to become the main constituent of anchoring fibrils. The NC1 domain has been shown to bind basement membrane type IV collagen. These molecular properties nicely account for the properties of anchoring fibrils. These structures attach by their extremities to epithelial basement membranes and to anchoring plaques that are found close to these basement membranes and appear to contain the same constituents. The resulting network physically entraps interstitial collagen fibrils and contributes significantly to adherence of the basement membrane, and thus of the epithelium to the stroma (Fig. 2).

This function is clearly demonstrated by the clinical phenotype of recessive dystrophic epidermolysis bullosa (RDEB), a severe blistering disease of the skin and other external mucous membranes (57). The blisters are caused by the separation of the epithelial basement membrane from the stroma under extremely mild mechanical stress. Absence or abnormalities in type VII collagen and anchoring fibrils have been noted in several cases of RDEB both biochemically and histologically. As in most heritable disorders involving structural macromolecules, RDEB appears to be a very heterogeneous condition.

THE UPCOMING COLLAGENS

The repetitive primary structure of the triple helices and the resulting high similarities in DNA sequences of collagen-encoding genes have helped tremendously in cloning and identification of collagen cDNAs through cross-hybridization. It can be safely stated that the vast majority of cloning in the collagen field was done by cross-hybridization with previously isolated cDNA clone for some other collagen chain, in particular with the very first collagen cDNA (for the chicken α2(I) chain) that was isolated in the laboratory of H. Boedtker (58) and readily made available to the scientific community. This has resulted not only in the identification of clones encoding already identified chains but also in the discovery of sequences from yet undescribed molecules.

As discussed previously, type XII collagen was discovered in this fashion and was described going from the gene to the protein. Type XIII collagen has also been discovered at the cDNA level and extensively characterized at that level (59). The presumptive translation product, clearly distinct from any other known collagen molecule, would contain three triple helical domains. The analysis of cDNA sequences has revealed the existence of several alternate splicing sites. Although the synthesis of this chain has been demonstrated using antibodies raised against a peptide synthesized according to the cDNA sequence, no data are yet available regarding the protein itself.

It is clear that the list of collagens in vertebrates is not closed. Preliminary evidence from several laboratories already indicates that additional members of the family are being identified. Because the bulk of the collagen content of the vertebrates is accounted for by the early described molecules, the newcomers are likely to participate in specialized aggregates found either ubiquitously at low concentrations or focally at higher concentrations or expressed at specific times of development. Most of the collagen molecules described recently comprise large nontriple helical domains that appear to be important in the aggregation process. Whereas the roles of the triple helical domains appear relatively easy to predict, unraveling the molecular functions of the nontriple helical domains in the dynamic process of matrix assembly presents a challenge for the upcoming research in the field. FJ

We thank Drs. B. R. Olsen, R. Mayne, and F. Ramirez for providing manuscripts before publication and for helpful discussions. We also thank our colleagues who contributed the electron micrographs used for the assembly of Fig. 2.

Note added in proof: Parente et al. (61) have very recently reported the isolation and characterization of a cDNA for the human α1(VII) collagen chain. These data demonstrate that the orientation of the type VII collagen molecule is inverted from what had been previously reported and quoted in this review. The NC1 domain shown in Fig. 1 and responsible for interaction with basement membranes is thus the amino end of the molecule and the processed globular end is actually at the carboxyl end.

REFERENCES

1. Eastoe, J. E. (1967) Composition of collagen and allied proteins. In *Treatise on collagen* (Ramachandran, G. N., ed) pp. 1-72, Academic, London
2. Gross, J. (1956) The behavior of collagen units as a model in morphogenesis. *J. Biophys. Biochem. Cytol. Suppl.* **2,** 26
3. Ramachandran, G. N., and Reddi, A. H., eds (1976) *Biochemistry of collagen.* Plenum, New York
4. Miller, E. J., and Matukas, V. J. (1969) Chick cartilage collagen: a new type of α1 chain not present in bone or skin of the species. *Proc. Natl. Acad. Sci. USA* **64,** 1264-1268
5. Mayne, R., and Burgeson, R. E., eds (1987) Structure and function of collagen types. Academic, Orlando, Florida
6. Kefalides, N. A. (1973) Structure and biosynthesis of basement membranes. *Int. Rev. Connect. Tissue Res.* **6,** 63-104
7. van der Rest, M., and Garrone, R. (1990) Collagens as multidomain proteins. *Biochimie* **72,** 473-484
8. Schmitt, F. O. (1942) Electron microscope investigations of the structure of collagen. *J. Cell. Comp. Physiol.* **20,** 11-33
9. Hodge, A. J., and Petruska, J. A. (1963) Recent studies with the electron microscope on the ordered aggregates of the tropocollagen molecule. *Aspects of protein structure* (Ramachandran, G. N., ed) p. 289, Academic, New York
10. Brodsky, B., and Eikenberry, E. F. (1985) Supramolecular collagen assemblies. *Ann. N.Y. Acad. Sci.* **460,** 73-84
11. Yamauchi, M., Katz, E. P., Otsubo, K., Teraoka, K., and Mechanic, G. L. (1989) Cross-linking and stereospecific structure of collagen in mineralized and nonmineralized skeletal tissues. *Connect. Tissue Res.* **21,** 159-169
12. Vuorio, E., and De Crombrugghe, B. (1990) The family of collagen genes. *Annu. Rev. Biochem.* **59,** 837-872
13. Henkel, W., and Glanville, R. W. (1982) Covalent crosslinking between molecules of type I and type III collagen. *Eur. J. Biochem.* **122,** 205-213
14. Birk, D. E., Fitch, J. M., Babiarz, J. P., and Linsenmayer, T. F. (1988) Collagen type I and V are present in the same fibril in the avian corneal stroma. *J. Cell Biol.* **106,** 999-1008
15. Keene, D. R., Sakai, L. Y., Bächinger, H. P., and Burgeson, R. E. (1987) Type III collagen can be present on banded collagen fibrils regardless of fibril diameter. *J. Cell Biol.* **105,** 2393-2402
16. Mendler, M., Eich-Bender, S. G., Vaughan, L., Winterhalter, K. H., and Bruckner, P. (1989) Cartilage contains mixed fibrils of collagen types II, IX, and XI. *J. Cell Biol.* **108,** 191-197
17. Fleischmajer, R., Timpl, R., Tuderman, L., Raisher, L., Wiestner, M., Perlish, J. S., and Graves, P. N. (1981) Ultrastructural identification of extension aminopropeptides of type I and type III collagens in human skin. *Proc. Natl. Acad. Sci. USA* **78,** 7360-7364
18. Morris, N. P., and Bächinger, H. P. (1987) Type XI collagen is a heterotrimer with a composition (1α,2α,3α) retaining non-triple-

helical domains. *J. Biol. Chem.* **262,** 11345-11350
19. Niyibizi, C., and Eyre, D. R. (1989) Identification of the cartilage α1(XI) chain in type V collagen from bovine bone. *FEBS Lett.* **242,** 314-318
20. Bernard, M. P., Yoshioka, H., Rodriguez, E., van der Rest, M., Kimura, T., Ninomiya, Y., Olsen, B. R., and Ramirez, F. (1988) Cloning and sequencing of pro-α1(XI) collagen cDNA demonstrates that type XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilagenous tissue. *J. Biol. Chem.* **263,** 17159-17166
21. Kimura, T., Cheah, K. S. E., Chan, S. D. H., Lui, V. C. H., Mattei, M.-G., van der Rest, M., Ono, K., Solomon, E., Ninomiya, Y., and Olsen, B. R. (1989) The human α2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and genomic DNA reveals characteristics of a fibrillar collagen with differences in genomic organization. *J. Biol. Chem.* **264,** 13910-13916
22. Yoshioka, H., and Ramirez, F. (1990) Pro-α1(XI) collagen. Structure of the amino-terminal propeptide and expression of the gene in tumor cell lines. *J. Biol. Chem.* **265,** 6423-6426
23. Exposito, J.-Y., and 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
24. D'Alessio, M., Ramirez, F., Suzuki, H. R., Solursh, M., and Gambino, R. (1990) Cloning of a fibrillar collagen gene expressed in the mesenchymal cells of the developing sea urchin embryo. *J. Biol. Chem.* **265,** 7050-7054
25. Peltonen, L., Halila, R., and Ryhänen, L. (1985) Enzymes converting procollagens to collagens. *J. Cell. Biochem.* **28,** 15-21
26. Prockop, D. J., Kivirikko, K. I., Tuderman, L., and Guzman, N. A. (1979) The biosynthesis of collagen and its disorders (first of two parts). *N. Engl. J. Med.* **301,** 13-23
27. Adachi, E., Hayashi, T., and Hashimoto, P. H. (1989) Immunoelectron microscopical evidence that type V collagen is a fibrillar collagen: importance for an aggregating capability of the preparation for reconstituting banding fibrils. *Matrix* **9,** 232-237
28. Fleischmajer, R., Perlish, J. S., and Timpl, R. (1985) Collagen fibrillogenesis in human skin. *Ann. N.Y. Acad. Sci.* **460,** 246-257
29. Gordon, M. K., and Olsen, B. R. (1990) The contribution of collagenous proteins to tissue-specific matrix assemblies. *Curr. Opin. Cell Biol.* **2,** 833-838
30. Muragaki, Y., Nishimura, I., Henney, A., Ninomiya, Y., and Olsen, B. R. (1990) The α1(IX) collagen gene gives rise to two different transcripts in both mouse embryonic and human fetal RNA. *Proc. Natl. Acad. Sci. USA* **87,** 2400-2404
31. Brewton, R. G., Wright, D. W., and Mayne, R. (1991) Structural and functional comparison of type IX collagen-proteoglycan from chicken cartilage and vitreous humor. *J. Biol. Chem.* **266,** 4752-4757
32. Vasios, G., Nishimura, I., Konomi, H., van der Rest, M., Ninomiya, Y., and Olsen, B. R. (1988) Type IX collagen-proteoglycan contains a large amino-terminal globular domain encoded by multiple exons. *J. Biol. Chem.* **263,** 2324-2329
33. Eyre, D. R., Apone, S., Wu, J.-J., Ericsson, L. H., and Walsh, K. A. (1987) Collagen type IX: evidence for covalent linkages to type II collagen in cartilage. *FEBS Lett.* **220,** 337-341
34. van der Rest, M., and Mayne, R. (1988) Type IX collagen from cartilage is covalently crosslinked to type II collagen. *J. Biol. Chem.* **263,** 1615-1618
35. Gordon, M. K., Gerecke, D. R., and Olsen, B. R. (1987) Type XII collagen: distinct extracellular matrix component discovered by cDNA cloning. *Proc. Natl. Acad. Sci. USA* **84,** 6040-6044
36. Dublet, B., Oh, S., Sugrue, S. P., Gordon, M. K., Gerecke, D. R., Olsen, B. R., and van der Rest, M. (1989) The structure of avian type XII collagen. α1(XII) chains contain 190-kDa non-triple helical amino-terminal domains and form homotrimeric molecules. *J. Biol. Chem.* **264,** 13150-13156
37. Dublet, B., and van der Rest, M. (1991) Type XIV collagen, a new homotrimeric molecule extracted from bovine skin and tendon, with a triple helical disulfide-bonded domain homologous to type IX and type XII collagens. *J. Biol. Chem.* **266,** 6853-6858
38. Timpl, R. (1989) Structure and biological activity of basement membrane proteins. *Eur. J. Biochem.* **180,** 487-502
39. Saus, J., Wieslander, J., Langeveld, J. P. M., Quinones, S., and Hudson, B. G. (1988) Identification of the Goodpasture antigen as the α3(IV) chain of collagen IV. *J. Biol. Chem.* **263,** 13374-13380
40. Hostikka, S. L., Eddy, R. L., Byers, M. G., Höyhtyä, M., Shows, T. B., and Tryggvason, K. (1990) Identification of a distinct type IV collagen α chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. *Proc. Natl. Acad. Sci. USA* **87,** 1606-1610
41. Yurchenco, P. D., and Schittny, J. C. (1990) Molecular architecture of basement membranes. *FASEB J.* **4,** 1577-1590
42. Cecchini, J. P., Knibiehler, B., Mirre, C., and Le Parco, Y. (1987) Evidence for a type-IV-related collagen in *Drosophila melanogaster.* Evolutionary constancy of the carboxy-terminal noncollagenous domain. *Eur. J. Biochem.* **165,** 587-593
43. Blumberg, B., MacKrell, A. J., and Fessler, J. H. (1988) *Drosophila* basement membrane procollagen α1(IV). II. Complete cDNA sequence, genomic structure, and general implications for supramolecular assemblies. *J. Biol. Chem.* **263,** 18328-18337
44. Benya, P. D., and Padilla, S. R. (1986) Isolation and characterization of type VIII collagen synthesized by cultured corneal endothelial cells. A conventional structure replaces the interrupted-helix model. *J. Biol. Chem.* **261,** 4160-4169
45. Yamaguchi, N., Benya, P. D., van der Rest, M., and Ninomiya, Y. (1989) The cloning and sequencing of α1(VIII) collagen cDNAs demonstrate that type VIII collagen is a short chain collagen and contains triple-helical and carboxyl-terminal non-triple-helical domains similar to those of type X collagen. *J. Biol. Chem.* **264,** 16022-16029
46. Muragaki, Y., Jacenko, O., Apte, S., Mattei, M.-G., Ninomiya, Y., and Olsen, B. R. (1991) The α2(VIII) collagen gene—a novel member of the short-chain collagen family located on the human chromosome 1. *J. Biol. Chem.* **266,** 7721-7727
47. Sawada, H., Konomi, H., and Hirosawa, K. (1990) Characterization of the collagen in the hexagonal lattice of Descemet's membrane: its relation to type VIII collagen. *J. Cell Biol.* **110,** 219-227
48. Muragaki, Y., Mattei, M.-G., Yamaguchi, N., Olsen, B. R., and Ninomiya, Y. (1991) The complete primary structure of the human α1(VIII) chain and assignment of its gene (COL8A1) to chromosome 3. *Eur. J. Biochem.* **197,** 615-622
49. Schmid, T. M., and Linsenmayer, T. F. (1990) Immunoelectron microscopy of type X collagen: supramolecular forms within embryonic chick cartilage. *Dev. Biol.* **138,** 53-62
50. Murray, L. W., and Tanzer, M. (1985) The collagens of the annelida. In *Biology of invertebrate and lower vertebrate collagens* (Bairati, A., and Garrone, R., eds) pp. 243-258, Plenum, New York
51. Cox, G. N. (1990) Molecular biology of the cuticle collagen gene families of *Caenorhabditis elegans* and *Haemonchus contortus. Acta Tropica* **47,** 269-281
52. Chu, M.-L., Zhang, R.-Z., Pan, T., Stokes, D., Conway, D., Kuo, H.-J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R., and Timpl, R. (1990) Mosaic structure of globular domains in the human type VI collagen α3 chain: similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J.* **9,** 385-393
53. Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R., and Colombatti, A. (1990) Structural and functional features of the α3 chain indicate a bridging role for chicken collagen VI in connective tissues. *Biochemistry* **29,** 1245-1254
54. Saitta, B., Stokes, D. G., Vissing, H., Timpl, R., and Chu, M.-L. (1990) Alternative splicing of the human α2(VI) collagen gene generates multiple mRNA transcripts which predict three protein variants with distinct carboxyl termini. *J. Biol. Chem.* **265,** 6473-6480
55. Kielty, C. M., Boot-Handford, R. P., Ayad, S. Shuttleworth, C. A., and Grant, M. E. (1990) Molecular composition of type VI collagen. Evidence for chain heterogeneity in mammalian tissues and cultured cells. *Biochem. J.* **272,** 787-795
56. Regauer, S., Seiler, G. R., Barrandon, Y., Easley, K. W., and Compton, C. C. (1990) Epithelial origin of cutaneous anchoring fibrils. *J. Cell Biol.* **111,** 2109-2115
57. Bruckner-Tuderman, L., Mitsuhashi, Y., Schnyder, U. W., and Bruckner, P. (1989) Anchoring fibrils and type VII collagen are absent from skin in severe recessive dystrophic epidermolysis bullosa. *J. Invest. Dermatol.* **93,** 3-9
58. Frischauf, A. M., Lehrach, H., Rosner, C., and Boedtker, H. (1978) Procollagen complementary DNA, a probe for messenger RNA purification and the number of type I collagen genes. *Biochemistry* **17,** 3243-3249
59. Pihlajaniemi, T., and Tamminen, M. (1990) The α1 chain of type XIII collagen consists of three collagenous and four noncollagenous domains, and its primary transcript undergoes complex alternative splicing. *J. Biol. Chem.* **265,** 16922-16928
60. Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P., and Krieger, M. (1990) Type I macrophage scavenger receptor contains α-helical and collagen-like coiled coils. *Nature (London)* **343,** 531-535
61. Parente, M. G., Chung, L. C., Ryynämen, J., Woodley, D., Winn, K. C., Bauer, E., Chu, M.-L., and Uitto, J. (1991) Human type VII collagen: cDNA cloning and chromosomal mapping of the gene. *Proc. Natl. Acad. Sci. USA* **88,** 6931-6935