0022-202X/82/7901-0025s$02.00/0
THE JOURNAL OF INVESTIGATIVE DERMATOLOGY, 79:25s–30s, 1982

Vol. 79, Supplement 1
Printed in U.S.A.

Genetic Heterogeneity of Collagens

ROBERT E. BURGESON, PH.D.

Department of Pediatrics, Division of Medical Genetics, Harbor-UCLA Medical Center, School of Medicine, 1124 W. Carson St., C-3, Torrance, California 90502 U.S.A.

The term collagen has recently been expanded to include at least 7 genetically distinct structural elements of mammalian connective tissues. It is assumed that differences in the primary structure specify the interactions of these different collagens with one another and with noncollagen connective tissue elements. The specific biomechanical properties of individual connective tissues result from the relative content of the different collagen types.

Four major classes of collagens can be defined on the basis of compositional and structural characteristics. The first class includes collagens Type I, II, and III which form classically described compact banded film structures resulting from crosslink stabilized side-by-side interactions of the triple helical structural domains. Type IV, or basement membrane collagen, comprises the second class. These molecules form open fiber structures by disulfide stabilized end-to-end interactions of the nonhelical amino and carboxyterminal structural domains. The third class of collagens contains the Type V collagens and the molecules containing the E and F chains found in cartilage. These molecules may interact by a combination of side-by-side and end-to-end aggregation. Still a fourth class is suggested by recent descriptions of several collagens which appear to contain extensive regions of unstable triple helix within the triple helical structural domain. It has been suggested that these collagens may serve as links between collagen and noncollagen structural elements.

These concepts suggest that the specificity of interactions directed by the different collagen types result from (1) the extent of removal of the nonhelical collagen domains and (2) the integrity of the triple helical structure within the triple helical domain. These 2 properties of the different collagens are directly specified by genetically determined amino acid sequence differences.

GENETIC HETEROGENEITY OF COLLAGENS

The connective tissue structures of the vertebrate body are extremely diverse morphologically and functionally. The dramatic differences in their properties have multiple examples. When one compares the properties of cartilage and bone, one sees in cartilage a structure uniquely suited to a relatively friction-free and self-lubricating articular surface. The resiliency of cartilage allows for absorption and dispersion of physical forces which might cause undue wear and fracture of a more rigid structure. In contrast, the apposition of mineral upon the organic matrix of bone results in a structure whose biomechanical properties are consistent with the inextensibility and incompressability required of these rigid skeletal elements.

The morphology of the organic matrices of bone and cartilage are unique. The fibrous elements of cartilage have a narrow fiber diameter and occur in random orientations, enmeshing extensively hydrated proteoglycan aggregates. These thin fiber bundles interact in specific ways with each other, with the proteoglycan subunits, with the chondrocyte cell surfaces, and with other structural proteins and tissue constituents. Demineralized bone, on the other hand, contains fibrous elements having much larger fiber diameters. These broad fibers interact one with the other to form parallel arrays of fibers coursing through the bone matrix with a general orientation that is parallel to the long axis of the bone. The molecular constituents of these fibers align in a side-by-side manner to a greater extent than the analogous fiber subunits of cartilage. Additionally, the resulting large diameter fibers interact with other fibers, mineral, osteocytes, and bone remodeling enzyme systems in unique ways to produce an osseous tissue with biomechanical properties unlike cartilage or other connective tissues.

Despite these functional and structural differences, the primary structural protein of both these tissues is collagen. Isolation of the presumed collagen subunit by sequential extraction resulted in the characterization of a molecule capable of self-assembly into fibers under defined conditions. The observation of an ubiquitous self-assembling fiber subunit posed a dilemma which is yet unresolved: that is, how can a single molecular entity determine the complex and unique interactions required for the development and stabilization of tissues with functions and structures as diverse as cartilage and bone?

Within the last decade, two developments within the field of connective tissue biochemistry have provided partial solutions to this dilemma. The first of these is the realization that all collagens are secreted as molecules which are substantially larger than those which were initially characterized. This initial product of translation contains two major structural domains. The first of these is the well-characterized triple-helical collagen domain. The second is the globular domain recognized to occur both at the amino and carboxy terminal ends of the triple-helical domain. The second development is our understanding that the word collagen refers not to a single molecular species, but to a family of closely related structural macromolecules. Each of these different types of collagen have unique amino acid sequences specifying their two main structural domains.

The secreted form of all collagens thus far well characterized contains both a helical region of approximately 3,000 angstroms in length and large nonhelical structural domains at either end of the molecule (Fig 1). Once the molecule enters the extracellular space, the nonhelical domain becomes susceptible to specific proteolytic enzymes which remove part or all of these regions. The extent and rate of removal is collagen type specific. The specificity of this proteolytic event is determined by both the collagen type-specific amino acid sequences and the accessibility of the molecule to the extracellular proteases. The product of these proteolytic events are the final collagen molecules which are the subunits for connective tissue matrix assembly. The final cleavage event may be directly involved in fiber formation.

The stability of the triple helix under physiological conditions also varies with the collagen type. The most common and best characterized collagens contain a triple-helical domain having amino acid sequences which include glycine in every third position and the presence of sufficient hydroxyproline to stabilize the collagen helix. Having met these requirements, these

Reprint requests to: Robert E. Burgeson, Ph.D., Department of Pediatrics, Division of Medical Genetics, Harbor-UCLA Medical Center, School of Medicine, 1124 W Carson St., C-3, Torrance, CA 90502.

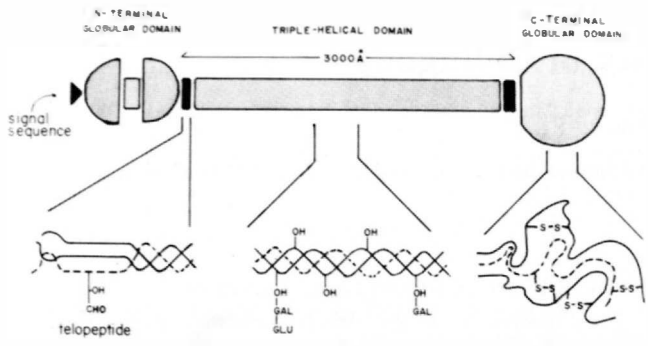

FIG 1. All well-characterized collagens have common features schematically illustrated here. Each has 3 major structural domains. The N-terminal domain has been shown to contain a short region of triple helix in some cases. The triple-helical domain is approximately 3,000 angstroms in length in the major structural collagens and in the pericellular collagens. It may be somewhat longer in Type IV collagen. The helical domains of Type IV collagen and the class 4 collagens contain regions of discontinuity which are susceptible to several proteases. The C terminal domain contains both interchain and intrachain disulfide bonds. The initial translation product of the collagen genes contain a short "signal" sequence at the end terminus, which directs the molecule into the secretory pathway, and which is removed intracellularly. At either end of the triple-helical domain are short regions of nontriple helix which survive extracellular removal of the globular domains in the class 1 collagens. This telopeptide is a major site of action for lysyl oxidase, and the product aldehyde participates in crosslinking which stabilizes the ultimate fiber form.

collagens contain a helical structural domain which is relatively rigid, nonspecific protease resistant, and capable of side-by-side interactions with analogous helical domains to produce fiber structures extending many times the length of the fiber subunit. Not all collagens have triple-helical domains which meet these rigid requirements for perfect helicity. Several newly characterized collagens have specific regions within this structural domain which are sensitive to noncollagenase proteolysis. In the case of Type IV collagen, this susceptibility appears to be due to a discontinuity in the helix resulting from the absence of glycine in the required positions. These interruptions in the triple helix cause focal disruptions in the linear rigidity of the structural domain. The precise functions of this increased flexibility are not yet determined, but it is likely that they serve to destabilize side-by-side helix region aggregation and provide specific sites for interaction of the triple helix, both with other collagens and with noncollagen structural elements.

In addition to specific discontinuities in the triple helix, the different collagens also undergo varying degrees of posttranslational modification along the helical structural domain. There are substantial differences in the degree of lysyl-hydroxylation and the addition of the neutral hexoses, glucose and galactose, to these modified lysine residues. Again, no function has yet been assigned to the carbohydrates within the helical structural domain, but they appear to correlate with the fiber length or the ultimate fiber diameter.

In summary, the differences between the different collagens in vertebrate connective tissues are far more extensive than simple point differences in the amino acid sequence. The fiber subunits of each collagen type differs in the "perfection" of its helical domain, the extent of posttranslational modification of this structural domain, and in the extent to which it retains the nonhelical structural domain. We presently believe that these differences specify the extent to which each of the collagens is capable of side-by-side interaction to form extended linear fiber systems, and determines the ability of the subunits to interact end-to-end through interactions of the nonhelical domains to form the more open lattices throught to be required for the structure of basement membranes and the exocytoskeleton.

THE GENETICALLY DISTINCT VERTEBRATE COLLAGENS

We are not yet ready to place an upper limit on the number of genetically distinct collagens. For the purposes of this discussion, I have chosen to subdivide the collagens into four classes based on our current knowledge of their characteristics. The first class consists of the major structural elements. The main components of this class are collagens Type I, II, and III. The Type I trimer also belongs within this class. The principal characteristics of these collagens is their ability to form and maintain extensive side-by-side interactions forming the long, linear aggregates once thought to be characteristic of all collagenous structures. The second class of collagens contains the molecules which form the supportive matrix of basement membranes. Currently, this class is represented only by Type IV collagen. The distinguishing characteristic of this class of collagens is their ability to form a relatively less compact and more randomly oriented fiber system. The third collagen class encompasses Type V collagen and the minor collagens found in cartilage. On the basis of their immunolocalization to pericellular regions, these molecules are often referred to as pericellular collagens. The fourth class of collagens contains a number of molecules which thus far are only poorly characterized. The most consistent characteristic of this class of molecules is their apparent discontinuous helical structural domain, as evidenced by their susceptibility to a variety of proteases. The enzyme susceptibility of regions of this type of helix suggests that these may be sites of greater flexibility, or sites of increased interaction with other collagen and noncollagen connective tissue elements.

CLASS 1. THE MAJOR STRUCTURAL COLLAGENS

This first class contains the most commonly occurring collagens which together contribute the greatest amount to the mass of all connective tissues. Because these collagens are the most abundant, they are also the best characterized [1]. Included in this class are Type I, Type I trimer [2], the major [3] and minor Type II [4] collagens, and Type III [5] collagen. The alpha chain subunit components of each of these collagens and their molecular arrangement are indicated in Table I. The principal characteristic of the collagens of this class is their ability to form and maintain extensive side-by-side interactions forming long linear aggregates. The amino acid compositions of the collagens of this class are listed in Table II. Each of these collagens contains one-third glycine, occurring in every third position of a linear amino acid sequence (see reference 6 for review). The imino acids, proline and hydroxyproline, each account for another 10% of the total composition. The positions and contents of glycine and hydroxyproline specify and stabilize the collagen helix of each alpha subunit and allow association of these subunits into a triple-helical conformation. The continuity of this conformation along the 3,000 angstrom molecular length provides relative inflexibility and resistance to noncollagen specific proteases. The single exception to this generalization is the presence of a tryptic cleavage site in Type III collagen very near the site of action of vertebrate collagenase [7].

The posttranslational hydroxylation of proline and lysine, as well as the glycosylation of hydroxylysine with glucose and galactose varies between these different collagen classes, as well as with age, the tissue source, and the species [8]. Generally, the level of prolylhydroxylation approximates 50% in this collagen class. The levels of hydroxylsine and glycosylated hydroxylysine are substantially lower in these collagens than found in other vertebrate collagens. These collagens also have a characteristic absence of extensive intramolecular disulfide bonding. Type III collagen alone contains covalent bonds between cysteine residues at the carboxy terminal end of the triple-helical domain [9]. Disulfide bonding between molecules is also often seen in preparations of Type III collagen, but these

may result from disulfide exchange occurring during extraction and purification, and may not be of physiological significance.

The major structural collagens occur throughout vertebrate tissues. Immunolocalization studies suggest that these collagens are microlocalized separately from other classes of collagens. These collagens most often occur together in the same region of a tissue, the relative amount of one structural collagen to another being characteristic of the tissue type and developmental stage. Types I and III collagens generally occur together in varying relative amounts [10]. Studies of the relative occurrence of Type III collagen and type I collagen throughout development [8], and in certain pathological conditions [11], suggest that increased relative amounts of Type III collagen correlate with smaller fiber diameters and increased tissue extensibility [12]. The specific molecular interactions of Types I and III collagens are not known. There is no direct data supporting or contradicting the concept that two different collagens are subunits of the same fiber. However, indirect evidence obtained from studies of fibers formed *in vitro* from mixtures of Types I and III collagens suggests that the resulting fiber diameter is inversely proportional to the relative concentration of Type III in the initial collagen mixture [13]. The distribution of Type II collagen is largely restricted to cartilaginous structures. It most often occurs alone, regionally separated from other types of collagen. The intervertebral disc and fibrocartilage are exceptional in that these tissues contain both Types I and II collagens [14].

Although each of the collagens of this class is initially synthesized and secreted as a larger structure having extensive nonhelical structural domains in addition to the triple helix, only extremely short regions at the amino and carboxy terminal ends of the fiber subunits survive proteolytic excision of the nontriple-helical domains [15]. The relative rate of excision of these nonhelical regions varies between members of this collagen class. Biosynthetic studies indicate that this conversion occurs rapidly for Types I and II collagens, and at a retarded rate for Type III. Investigations of the disease condition in cattle known as dermatosparaxis indicate that this mutant lacks the specific enzyme for removal of the amino terminal structural domain from Type I collagen [16,17]. Ultrastructural studies of the collagen fibers from these mutant tissues indicate that the abnormal presence of the bulky, nonhelical domain at the amino terminus interferes with the compact side-by-side alignment of the resulting fibers. If Type I and Type III collagens can be incorporated into the same growing fiber as suggested above, and if the nonhelical structural domains interfere with lateral growth or stabilization of the fiber, then it is likely that the rate of excision of the Type III collagen nonhelical structural domain may play a significant role in limiting the rate of fiber growth, or the ultimate fiber diameter [18].

The structural characteristics of the class 1 collagens briefly summarized above indicate that collagens capable of extensive lateral association have several common characteristics. First, they contain stable triple helix throughout their 3,000 angstrom length. Secondly, the ultimate fiber diameter is inversely proportional to their content of neutral hexose, which may interfere stearicly with lateral aggregation. Thirdly, the product fibers do not contain extensive nontriple-helical collagen regions. And lastly, the relative amount of Type III collagen reduces the

TABLE I. *The α-chain compositions of vertebrate collagens*

Class 1—Major structural collagens	
Type I	$[\alpha1(I)]_2\alpha2$
I trimer	$[\alpha1(I)]_3$
II (major)	$[\alpha1(IIM)]_3$
II (minor)	$[\alpha1(IIm)]_3$
III	$[\alpha1(III)]_3$
Class 2—Basement membrane collagens	
Type IV	$[\alpha1(IV)]_2\alpha2(IV)$
Class 3—Pericellular collagens	
Type V	$[\alpha1(V)]_2\alpha2(V)$ "AB_2" $[\alpha1(V)]_3$, $[\alpha3(V)]_3$ $\alpha1(V)\alpha2(V)\alpha3(V)$
Type VI? (minor cartilage collagens)	$[E]_3$ $[F]_3$
Class 4—Collagens with discontinuous triple helix	
EC	$[EC]_3$
HMW, LMW (Reese & Mayne)	Disulfide bonded high molecular weight aggregates
Acidic and basic collagenous fragments (Furuto & Miller)	

TABLE II. *Compositional characteristics of different collagen classes: Residues/1,000 residues*

Amino acid	Class 1 Major structural collagens				Class 2 Basement membrane collagen	Class 3 Pericellular collagens				Class 4 Collagens with discontinuous triple helix		
	α1(I)	α2(I)	α1(II)	α1(III)	Type IV	α1(V) (B)	α2(V) (A)	α2(V) (C)	E	F	Furuto & Miller [38] Acidic Frx'n.	HMW [37]
3 OH-proline	1.1	1.2	1.1	0	12	2.9	2.5	2.2	1.1	2.3	nr.	0
4 OH-proline	114	105	96	126	140	109	109	92	95	91	86	95
Aspartic acid	46	45	36	42	51	50	51	42	46	50	83	53
Threonine	18	18	22	13	20	19	26	19	17	24	12	16
Serine	35	30	27	38	37	36	31	34	24	28	23	38
Glutamic acid	77	70	95	71	79	91	84	98	107	97	114	97
Proline	118	114	106	107	65	118	97	99	109	118	100	90
Glycine	330	331	334	352	328	344	341	332	330	322	293	323
Alanine	119	105	108	95	37	46	52	49	54	49	48	57
Cystine/2	0	0	0	2	1	0	0	1.3	0	0	31	2
Valine	19	35	19	15	28	25	24	29	27	18	22	23
Methionine	7	8	11	8	13	8	11	8.1	10	9	10	9
Isoleucine	8	16	11	13	29	19	16	20	15	16	20	27
Leucine	20	33	27	23	52	39	35	56	35	39	23	48
Tyrosine	2	3	1	3	2	2	2	2.4	2	3	23	5
Phenylalanine	13	11	13	8	29	12	14	9.2	11	11	17	8
Hydroxylysine	10	12	18	6	49	35	24	43	37	40	28	34
Lysine	27	20	20	29	9	20	18	15	18	15	11	21
Histidine	4	10	2	6	6	8	11	14	6	11	3	8
Arginine	49	51	52	48	26	45	50	42	42	47	53	46
Glucose	1	1	5	1	40	18	10	17	29	36	23	nr.
Galactose	1	2	10	2	41	30	14	24	28	34	28	nr.

ultimate fiber diameter or orientation, possibly reflecting the decreased rate of proteolytic removal of the nonhelical structural domains.

CLASS 2. THE BASEMENT MEMBRANE COLLAGENS

The most distinguishing characteristic of the collagens of this class is their inability to associate into compact banded fibers. This characteristically different fiber form results from structural differences in the subunit molecules reflecting amino acid compositional differences and a lack of proteolytic removal of the nonhelical structural domains.

Presently, two alpha chain subunits have been described as components of molecules of this collagen class [19]. The molecular composition of basement membrane collagen is in dispute [19,20], though the bulk of the evidence favors the prediction of a single molecule having two like subunits, known as $\alpha 1$(IV) chains, in triple-helical conformation with a second alpha chain described as $\alpha 2$(IV). Due to technical difficulties, it has not been possible to obtain unambiguous amino acid compositions of the individual Type IV chains within the triple-helical structural domain. Table II illustrates the amino acid composition of Type IV collagen isolated following solubilization of the basement membrane matrix by limited proteolysis. Comparisons of this composition with the compositions of the other collagens solubilized under identical conditions indicate substantial homology of Type IV collagen to other characterized collagens. Like the other collagens, Type IV collagen contains approximately one-third glycine and 20% total imino acid within its triple-helical structural domain. In contrast, nearly 85% of the lysines of basement membrane collagen are converted to hydroxylysine. Of these hydroxylysine residues, better than 80% are substituted with the disaccharide glucosyl-galactose. The total lysine plus hydroxylysine content is approximately 50% greater than that found in the class 1 collagens. The content of cysteine reported in pepsin-solubilized Type IV collagen is variable, but this value is substantially greater than determined for the class 1 collagens. In addition, Type IV collagen has a characteristically low content of alanine.

The triple-helical structural domain of Type IV collagen contains several regions in which glycine is not present in every third position [1]. The substitution of an amino acid with a bulky functional group regionally distorts the alignment of the three alpha chain subunits of the triple-helical domain. In addition to rendering this area of the triple helix sensitive to numerous proteases, the discontinuity of the triple helix allows increased regional flexibility, and as a consequence, may destabilize lateral associations of the Type IV collagen helical domain. One possible function of such discontinuities might be to allow greater interaction of the collagen helical region with noncollagen components of basement membrane. For example, fibronectin is thought to interact with collagen under a variety of physiological conditions to influence such diverse processes as cellular attachment, cellular architecture, and neoplasia. While the noncovalent binding of fibronectin to nontriple-helical collagen alpha chains has been well documented, the binding of fibronectin to fully triple-helical class 1 collagen molecules has been shown to be substantially less [21]. It is quite possible that the specific interactions between noncollagen basement membrane macromolecules and the Type IV triple-helical domain might be facilitated by a regional unwinding of the triple helix. This concept is compatible with the lattice structure postulated for Type IV collagen matrix.

Like other collagens, Type IV collagen is synthesized and secreted as a large molecule, having both helical and nonhelical structural domains [22]. In contrast to the first class of collagens, biosynthetic studies of Type IV collagen indicate that the large nonhelical-structural domains are not substantially removed during maturation of Type IV collagen into the basement membrane matrix [23,24]. By analogy to Type I collagen fiber formation, we would expect that the presence of the bulky, nonhelical domain prevents extensive lateral associations of the

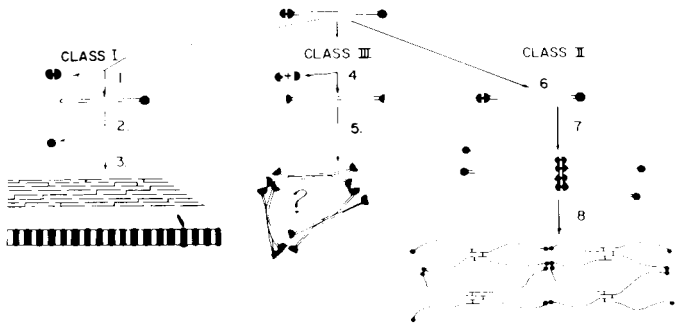

FIG 2. This model illustrates the collagen type specific extracellular modifications which determine the final fiber form. All collagens have the generalized structure illustrated in Fig 1. Specific proteolytic events modify this structure to determine how these molecules interact with one another to form specific fiber structures. The extent of these proteolytic events is different for each of the collagen classes. Class 1 collagens undergo 3 major steps. In reaction 1, the amino terminal globular domain is removed by a single proteolytic event. This liberated peptide participates in control of collagen synthesis. At step 2, the carboxy terminal structural domain is removed. This event may occur simultaneously with, and may facilitate fiber formation. Lacking the bulky, globular domains, and not having substantial amounts of neutral hexose along the length of the helical domain, the class 1 collagens are capable of extensive side-by-side interactions to form fibers. In step 3, this fiber structure is stabilized by aldehyde mediated crosslinking. The specific lateral alignment of the class 1 helical regions gives rise to the banded collagen fiber structures seen morphologically. The class 2 collagens undergo little or no proteolytic processing as illustrated in step 6. The presence of the large globular domains at either end of the molecule and the extensive addition of glucose and galactose along the triple helix prevent substantial lateral associations. As shown at step 7, the globular regions interact to form disulfide bonded aggregates. This aggregation is thought to occur between the N-terminal globular domains. The carboxy terminal globular domains are then free to interact with like regions of adjacent molecules as indicated at step 8 to form fiber structures of the basement membranes. The extent of processing of the class 3 collagens is intermediate between the other 2 classes. Step 4 in indicates the partial removal of both terminal globular domains. Lateral association of the triple-helical domain of these collagens is restricted by the high degree of neutral hexose along the triple helix and by the presence of the residual portions of the globular domains. These subunits interact to form the hypothetical structure illustrated at step 5. Presently, we have little knowledge of the structure and fiber form of the class 4 collagens.

Type IV molecules. Recent ultrastructural visualization of the Type IV molecule by rotary shadowing indicates that these nonhelical domains interact to form multimolecular aggregates of at least 4 basement membrane collagen molecules. These aggregates are stabilized by disulfide bonding. The free ends of the molecule thus associated are capable of specific interactions with the analogous portions of adjacent molecules. A model illustrating the proposed structure of the basement membrane collagen matrix is shown in Fig 2.

In summary, basement membrane collagen appears to have substantial similarities to the class 1 collagens. Both contain a large triple-helical structural domain and nontriple-helical structural domains. The proposed model of basement membrane collagen matrix differs from the fiber structures composed of class I collagens due more to differences in the extracellular processing of the initially synthesized and secreted macromolecule than due to primary structural differences. The very large residual nonhelical structural domains present in the Type IV collagen subunits of the basement membrane fiber structures, the extensive glycosylyation within the triple-helical domain, and the discontinuous nature of the triple-helical structural domain all contribute to a fibrous structure dependent more on end-to-end subunit interactions than lateral accretion.

CLASS 3. THE PERICELLULAR COLLAGENS

This class of collagens contains molecules composed of subunits termed $\alpha 1$(V), $\alpha 2$(V) [25,26], $\alpha 3$(V) [27] (also referred to

as B, A, and C chains), and the E and F chains isolated from vertebrate cartilage [28]. While the subunit composition of this class of collagen molecules is in dispute, a molecule containing two α1(V) chains and one α2(V) chain has been described [29,30], as well as a molecule containing three α1(V) chains [31]. The E and F chains are thought to occur in separate molecules [28]. Immunolocalization studies of the Type V collagens indicate that they are most often found in the region immediately adjacent to cells of a large variety of tissues [32]. These collagens were initially isolated from human chorioamniotic membranes, and have since been reported in such diverse tissues as bone, liver, tendon, gingiva, muscle, and skin [1]. Other studies indicate that they may also be components of basement membrane [33]. The E and F chains have thus far only been detected in cartilage, but extensive chemical or immunological surveys of other tissues have not been reported. Because the E and F chains share substantial compositional homologies with the Type V collagens, it is entirely possible that they too will be localized to pericellular tissue regions. We have no knowledge regarding the function of collagens of this class, but these localization studies suggest that these collagens form an interface between the cell surface and the surrounding matrix components. This interface may stabilize specific tissue architecture, or provide a zone of porous connective tissue to facilitate diffusion of secreted cell products.

The compositional features of the collagens of this class are not substantially different from the collagens of the other classes. Analyses of the amino acid and neutral hexose compositions of the triple-helical structural domain obtained following solubilization of these collagens by limited proteolysis with pepsin is shown in Table II. Each of these alpha chains contains one-third glycine and 20% imino acids, consistent with the requirements of a continuous triple-helical structural domain. The degree of lysylhydroxylation of each of these α chains is intermediate between the relatively low values determined for the collagens of the first class and the high degrees of lysylhydroxylation seen in the basement membrane collagens. Similarly, the content of glucose and galactose of the collagens in this class is intermediate between the values determined for the class 1 and class 2 collagens. The E and F chains of this class contain nearly all of their attached hexose as the disaccharide glucosyl-galactosyl hydroxylysine. Each of these chains also has a characteristically elevated content of histidine and an alanine content intermediate between the class 1 and class 2 collagens.

Like all other well-characterized collagens, these collagens are synthesized and secreted with large nontriple-helical structural domains at either end of the triple-helical domain [30]. Biosynthetic studies indicate that extracellular proteolysis only partially removes these nonhelical domains [34,35]. The exact nature of these proteolytic events is not completely understood, but the data are consistent with a model which predicts bulky, nontriple-helical extensions at both the amino and the carboxy terminal ends of the final product of these extracellular processing events. According to the model, molecules of this collagen class would be incapable of extensive lateral aggregation. The high content of neutral hexose along the length of the molecule supports this prediction. The fiber structure resulting from aggregation of such subunits might be expected to demonstrate extensive end-to-end interactions as well as limited lateral aggregation. Such a model is schematically present in Fig 2, though no direct evidence presently supports such a structure. The model predicts that the pericellular matrix has properties intermediate between that determined for the major structural collagens and that proposed for the matrix of basement membranes.

CLASS 4. COLLAGENS WITH DISCONTINUOUS TRIPLE HELIX

The collagens described in this section have been only recently detected, and their structures are not sufficiently well understood that they can be assigned to any of the previously described classes. Similar to Type IV collagen, all of these new collagens have triple-helical domains which are susceptible to nonspecific proteases. These new collagens have compositional characteristics substantially different from Type IV collagen. Some of them, like HMW, can be isolated from tissues which do not contain basement membrane, and therefore, do not belong to the class 2 collagens.

While there are numerous newly discovered collagenous materials which are candidates for membership in this class, three are particularly representative. These include EC collagen synthesized by endothelial cells [36], HMW and LMW collagenous fragments isolated from hyaline cartilage [37], and high molecular weight collagenous aggregates isolated from a variety of human tissues and recently described by Furuto and Miller [38,39]. EC collagen will be more fully described by Helene Sage in subsequent discussions at this meeting. Compositional studies of EC collagen indicates that this molecule contains amounts of hydroxyproline consistent with an extensive triple-helical structural domain, yet its susceptibility to a variety of proteases suggests that this helical domain is discontinuous. Following limited proteolysis, only fragments substantially smaller than collagen alpha chains are recovered. This enzyme susceptibility is characteristic of all the components of this collagen class.

Chick hyaline cartilage also contains several components which have compositional characteristics very similar to class 3 collagens, but demonstrate substantially greater susceptibility to proteolytic digestion. Two of these, termed HMW and LMW, have been partially characterized [37,40]. HMW is a pepsin-resistant fragment about 134 nanometers in length. It contains 3 subunit peptides, largely in triple-helical conformation. These 3 subunits are covalently joined by disulfide bonds. The amino acid composition of HMW is shown in Table II. Except for the presence of cysteine, the composition is strikingly similar to those determined for the class 3 collagens. HMW is believed to be a proteolytic fragment of a much larger parent molecule with pepsin-sensitive, nontriple-helical disruptions of the helical structural domain. The presence of disulfide bonding and the discontinuous nature of the triple-helical structural domain distinguish this molecule from other class 3 collagens. The organization of HMW and LMW within the cartilage matrix has not yet been determined, but their distribution is thought to be restricted to hyaline cartilage.

Another component of this class of newly discovered collagens is a disulfide-bonded, high molecular weight aggregate isolated from human placenta [38,39]. Upon solubilization of placenta by limited pepsin digestion, a large complex can be isolated which contains several collagen-like fragments. The structure can be disaggregated by disulfide bond reduction to yield multiple fragments. Three major fragments, one acidic and 2 with basic properties, have been partially characterized. The major acidic component contains approximately 400 amino acids, suggesting that a triple helix containing this fragment would be approximately half the length of a class 1 triple-helical structural domain. The amino acid composition of this acidic fragment is shown in Table II. Like other collagen helical domains, this collagen contains approximately 20% imino acid. However, the content of glycine is less than would be expected for a pepsin fragment containing glycine in every third position. A very high content of half-cystine reflects the extensive disulfide bonding exhibited by the high molecular weight aggregate. This value is significantly higher than has been reported for any collagen triple-helical domain. Like Type IV collagen, lysine residues are recovered almost exclusively as hydroxylysine. All of these hydroxylysine have covalently attached neutral hexose, substantially in the disaccharide form. In addition to glucose and galactose, the peptide also contains a significant amount of glucosamine. While these fragments show substantial similarity to Type IV collagen, their compositional features and unique disulfide bond stabilized organization distinguish them from the Type IV collagen gene products. Although they represent a

new gene product, their similarity to Type IV collagen and their isolation from basement membrane-rich tissues suggest that these new collagens may be basement membrane components.

The common property of each of these new collagens is the discontinuous nature of their triple-helical structural domain. While the molecular organizations of these fragments are not well understood, they appear to derive from larger molecules having continuous triple helices about one-half the length of other collagens. One might predict that these molecules are excellent candidates for specific interactions with noncollagen components of connective tissues. As such, these molecules may serve the long-sought-after link function between proteoglycan aggregates, fibronectin, noncollagen glycoproteins, laminin, cell surface components, and the collagen matrix of a large variety of connective tissues.

REFERENCES

1. Bornstein P, Sage H: Structurally distinct collagen types. Ann Rev Biochem 49:957–1003, 1980
2. Jimenez SA, Bashey RI, Benditt M, Yankowski R: Identification of collagen $\alpha 1$(I) trimer in embryonic chick tendons and calvaria. Biochem Biophys Res Commun 78:1354–1361, 1977
3. Miller EJ, Matukas VJ: Chick cartilage collagen: A new type of $\alpha 1$ chain not present in bone or skin of the species. Proc Natl Acad Sci 65:1264–1268, 1969
4. Butler WT, Finch JE Jr, Miller EJ: The covalent structure of cartilage collagen. J Biol Chem 252:639–643, 1977
5. Miller EJ, Epstein TH Jr, Piez KA: Identification of three genetically distinct collagens by cyanogen bromide cleavage of insoluble human skin and cartilage collagen. Biochem Biophys Res Commun 42:1024–1029, 1971
6. Fietzek F and Kühn K: The primary structure of collagen. Int Rev Connect Tissue Res 7:1–60, 1976
7. Miller EJ, Finch JE Jr, Chung E, Butler WT, Robertson PB: Sequence specific and cleavage of the native Type III collagen molecule with trypsin. Arch Biochem Biophys 173:631–637, 1976
8. Epstein EH Jr: $[\alpha 1(III)]_3$ human skin collagen. J Biol Chem 249:3225–3231, 1974
9. Lang H, Glanville RW, Fietzek PP, Kühn K: The covalent structure of calf skin Type III collagen. Hoppe-Seyler's Z Physiol Chem 360:841–850, 1979
10. Epstein EH, Munderloh NH: Isolation and characterization of CNBr peptides of human $[\alpha 1(III)]_3$ collagen and tissue distribution of $[\alpha 1(I)]_2\alpha 2$ and $[\alpha 1(III)]_3$ collagens. J Biol Chem 250:9304–9312, 1975
11. Pope FM, Martin GR, Lichtenstein JR, Penttinen RP, Gerson B, Rowe DW, McKusick VA: Patients with Ehlers-Danlos syndrome Type IV lack Type III collagen. Proc Natl Acad Sci 72:1314–1316, 1975
12. Byers PH, Holbrook KA, McGillvray B, MacLeond PM, Lowry RB: Clinical and ultrastructural heterogeneity of Type IV Ehlers-Danlos syndrome. Human Genet 47:141–150, 1979
13. Lapiere CM, Nusgens B, Pierard GE: Interaction between collagen Type I and Type III in conditioning bundles organization. Connect Tiss Res 5:21–29, 1977
14. Eyre DR, Muir H: Quantitative analyses of Types I and II collagens in human intervertebral discs at various ages. Biochem Biophys Acta 492:29–42, 1977
15. Fessler JH, Fessler LI: Biosynthesis of procollagen. Ann Rev Biochem 47:129–162, 1978
16. Simar LJ, Betz EH: Dermatosparaxis of the calf, a genetic defect of the connective tissue. Ultrastructural study of the skin. Hoppe-Seyler's Z Physiol Chem 352:13–22, 1971
17. Lapiere CM, Nusgens B: Collagen pathology at the molecular level, Biochemistry of Collagen. Edited by GN Ramachandran, AH Reddi. New York, Plenum Press, 1976, pp 377–447
18. Fessler JH, Tandberg WD: Interactions between collagen chains and fiber formation. J Super Molec Struct 3:17–23, 1975
19. Crouch E, Sage H, Bornstein P: Structural basis for apparent heterogeneity of collagens in human basement membranes: Type IV procollagen contains two distinct chains. Proc Natl Acad Sci 77:745–749, 1980
20. Robey PG, Martin GR: Type IV collagen contains two distinct chains in separate molecules. Coll Res 1:27–38, 1981
21. Ruoslahti E, Engvall E, Hayman EG: Fibronectin: Current concepts of its structure and functions. Coll Res 1:95–128, 1981
22. Sage H, Woodbury RG, Bornstein P: Structural studies on human Type IV collagen. J Biol Chem 254:9893–9900, 1979
23. Orkin RW, Gehron P, McGoodwin EB, Martin GR, Valentine T, Swarm R: A murine tumor producing a matrix of basement membrane. J Exp Med 145:204–220, 1977
24. Crouch E, Bornstein P: Characterization of a Type IV procollagen synthesized by human amniotic fluid cells in culture. J Biol Chem 254:4197–4204, 1979
25. Burgeson RE, El Adli FA, Kaitila II, Hollister, DW: Fetal membrane collagens: Identification of two new alpha chains. Proc Natl Acad Sci 73:2579–2583, 1976
26. Chung E, Rhodes RK, Miller EJ: Isolation of three collagenous components of probable basement membrane origin from several tissues. Biochem Biophys Res Commun 71:1167–1174, 1976
27. Sage H and Bornstein P: Characterization of a novel collagen chain in human placenta and its relation to AB collagen. Biochemistry 18:3815–3822, 1979
28. Burgeson RE, Hollister DW: Collagen heterogeneity in human cartilage: Identification of several new collagen chains. Biochem Biophys Res Commun 87:1124–2231, 1979
29. Bentz H, Bachinger HP, Glanville R, Kühn K: Physical evidence for the assembly of A and B chains of human placental collagen in a single triple helix. Eur J Biochem 92:563–567, 1978
30. Kumamoto C, Fessler JH: Biosynthesis of A, B procollagen. Proc Natl Acad Sci 77:6434–6438, 1980
31. Haralson MA, Mitchell WM, Rhodes RK, Kresina TF, Gay R, Miller EJ. Chinese hamster lung cells synthesize and confine to the cellular domain a collagen composed solely of B chains. Proc Natl Acad Sci 77:5206–5210, 1980
32. Gay S, Rhodes, RK, Gay RE, Miller EJ: Collagen molecules comprised of $\alpha 1$(V)-chains (B-chains): An apparent localization in the exocytoskeleton. Coll Res 1:53–58, 1981
33. Madri JA, Furthmayr H: Collagen polymorphism in the lung. Human Pathol 11:353–366, 1980
34. Kumamoto C, Fessler JH: Propeptides of procollagen V (A,B) in chick embryo crop. J Biol Chem 256:7053–7058, 1981
35. Fessler LI, Kumamoto C, Meis ME, Fessler JH: Assembly and processing of procollagen V (AB) in chick tissues. J Biol Chem in press, 1981
36. Sage H, Pritzl P, Bornstein D: A unique, pepsin-sensitive collagen synthesized by aortic endothelial cells in culture. Biochemistry 19:5747–5755, 1980
37. Reese CA, Mayne R: Minor collagens of chicken hyaline cartilage. Biochemistry, 20:918–924, 1981
38. Furuto DK, Miller EJ: Isolation of a unique collagenous fraction from limited pepsin digests of human placental tissue. J Biol Chem 255:290–295, 1980
39. Furuto DK, Miller EJ: Characterization of a unique collagenous fraction from limited pepsin digests of human placental tissue: Molecular organization of the native aggregate. Biochemistry 20:1635–1640, 1981
40. Mayne R, Reese CA, Wiedemann H: New collagenous molecules from chicken hyaline cartilage, New Trends in Basement Membrane Research. Edited by K Kühn, HH Schöne. New York, Raven Press, in press, 1981