Collagen: Molecular Diversity in the Body's Protein Scaffold

David R. Eyre

Summary. Intensive research in the last decade has revealed a wealth of detail on the mechanism of biosynthesis, molecular structure, and covalent cross-linking of collagen. Tissues of higher animals express a family of at least five genetically distinct types of collagen molecule, each apparently tailored for different construction work outside the cell. Within each genetic type of collagen, further chemical heterogeneity is also evident; the variations in hydroxylation, glycosylation, and cross-linking are dependent, for example, on tissue type, age, and hormonal status. The functional significance of collagen's molecular diversity and its control by different cells and tissues are not yet well understood but abnormalities of collagen in many human diseases keep this protein a focal molecule of medical research.

Collagen is the commonest protein in the animal world; it provides an extracellular framework for all multicellular animals. Although most abundant in fibrous connective tissues (Fig. 1), collagen appears in some shape or form in virtually every tissue. It provides (i) the ropes and straps (tendons and ligaments), (ii) woven sheets (skin and fascia), (iii) filtration membranes (glomeruli), (iv) supporting skeleton reinforced with mineral salts (bone and dentin), (v) bearing materials lubricated with proteoglycans (cartilage and intervertebral disk), and (vi) other special tissues that must be strong and yet have unusual properties, such as the light-transmitting cornea and fatigue-resistant heart valves. Only in the last decade has collagen's diversity and fine-tuning at the molecular level begun to be appreciated. It is now clear that collagen really describes a family, and probably several families, of specialized molecules, each a genetically distinct type that has evolved for a particular structural function outside the cell.

This article outlines what is known about the molecular structure of vertebrate collagens. Several reviews are available for details on the various topics (*1-4*) and Gross's earlier review gives a full account of the growth of collagen biology (*5*). Invertebrate collagens, which vary greatly in composition but are basically similar to vertebrate collagens, have been reviewed (*6*).

Common Structural Features

The definitive property of all collagen molecules is the triple helix, a unique protein conformation that is a coiled coil of three polypeptide subunits, or α chains (Fig. 2). Each α chain twists in a left-handed helix with three residues per turn, and the three chains are wound together in a right-handed superhelix to form a rodlike molecule about 1.4 nm in diameter (*7*). In the common collagens, the α chains each contain about 1050 amino acid residues and the molecule is 300 nanometers long. A nomenclature for the subunit composition of different collagen molecules has evolved where $[\alpha 1(I)]_2\alpha 2$ is type I, $[\alpha 1(II)]_3$ is type II, and $[\alpha 1(III)]_3$ is type III. The functional form of each of these three common collagens is the fibril, an ordered molecular polymer visible by electron microscopy as banded structures in the extracellular matrix of connective tissues (Fig. 1). Fibrils can range from 10 to 200 nm in diameter, depending on the collagen type and the tissue. Collagen fiber is the term usually reserved for the ordered fibrillar aggregates seen by light microscopy.

Glycine is the only amino acid small enough to cluster down the central core of the molecule, and every third residue in each α chain within the triple helix is a glycine residue. At both ends the terminal sequences (telopeptides) are not triple helical and lack glycine at every third residue (*8*). These telopeptides are primary sites of cross-linking in the molecule. A general formula for the α chains of types I, II, and III collagens minus their telopeptides is $(Gly\text{-}X\text{-}Y)_{340\pm2}$. Proline and 4-hydroxyproline together account for about a third of the X and Y positions, that is, a quarter of the total residues, with 4-hydroxyproline limited to Y. Restrictions on chain conformation imposed by the ring structures of proline and hydroxyproline strengthen the triple helix and stiffen the molecule (*7*). The special function of hydroxyproline remained a mystery until about 7 years ago, when its hydroxyl group was shown to be essential for stabilizing the triple helix, probably via intrachain or hydrogen bonds bridged through water molecules (*7, 9*). The crucial observation was that partially hydroxylated molecules, synthesized by cells in which hydroxylation was experimentally blocked, denatured (that is, the chains unwound) at a lower temperature than fully hydroxylated molecules (*9*). In short, with fewer interchain hydrogen bonds possible, the native helix is thermodynamically less stable. About 90 residues of 4-hydroxyproline per α chain are needed to preserve the triple helix of collagen molecules in solution at a body temperature of 37°C (*3*). Hydroxyproline is one of several modified amino acid residues made by posttranslational enzyme reactions that are essentially peculiar to collagen and whose main function seems to be to stabilize the molecule and the fibril.

Genetically Distinct Collagens

At least, seven different collagen α chains (comprising five different molecules or more) have been firmly identified as distinct gene products in higher animals (Table 1). As a broad concept, the different collagen types appear to be segregating into two major classes, as judged by their location and function outside the cell. Most familiar are the fibrillar, interstitial collagens that form the bulk, extracellular fabric of the major connective tissues, such as skin, bone, tendon, ligaments, and cartilage (for example, type I collagen in Fig. 1). A second class of less abundant collagens, as yet poorly defined, appears to be located in the immediate pericellular environment, perhaps as part of the cell's external skeleton, and to have a finer texture. The collagens of the specialized

The author is an assistant professor of biological chemistry in the Departments of Biological Chemistry and Orthopaedic Surgery, Harvard Medical School, at the Orthopaedic Research Laboratories, Children's Hospital Medical Center, Boston, Massachusetts 02115.

0036-8075/80/0321-1315$02.00/0
Copyright © 1980 AAAS

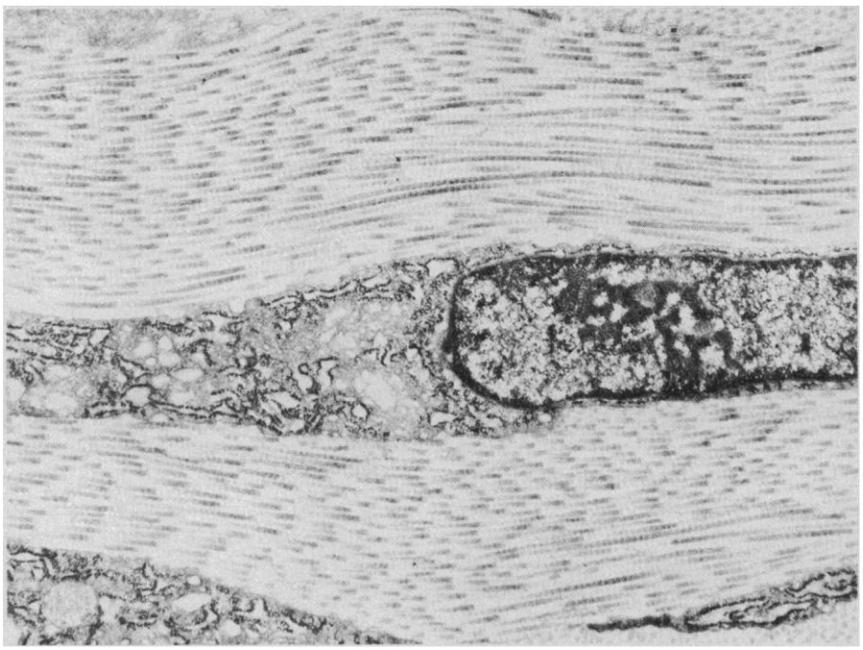

Fig. 1. Fields of coarse collagen fibrils (type I) and their parent fibroblasts. Electron-micrograph of outer periosteum. [Courtesy of M. E. Holtrop and F. D. Shapiro]

basement laminae of epithelial cell layers can be assigned to the latter, but should perhaps be considered in a specialized class of their own.

Interstitial collagens. Of the interstitial collagens, the most intensively studied is the molecule now known as type I. A decade ago it was the only know molecular type of collagen in higher animals. It accounts for about 90 percent of the collagen in the body, being almost the only collagen in bone and tendon; it is predominant in skin, fascia, and many other connective tissues (*2*). The type I molecule, $[\alpha 1(I)]_2\alpha 2$, unlike types II and III, is the product of two structural genes; it has been proposed that the one nonidentical $\alpha 2$ chain provides information, not possible in types that have three identical chains, that can regulate the way in which the molecules are oriented and cross-linked in the fibril (*10*). The type I molecule is the best developed of the collagens for forming large, well-organized fibrils and seems to be the molecule present where tensile strength is important.

The first hard evidence for genetic polymorphism of collagen in higher animals was the discovery of the $\alpha 1$(II) chain and its parent type II collagen molecule in hyaline cartilage (*11*). This molecule is the product of a single gene and has the subunit composition $[\alpha 1(II)]_3$. The nomenclature stems from the similar elution positions of $\alpha 1$(I) and $\alpha 1$(II) chains during ion exchange chromatography on carboxymethyl cellulose (*11*). Although clearly homologous to $\alpha 1$(I), the $\alpha 1$(II) chain has a different amino acid sequence (*12*) and is, therefore, a distinct gene product. Discovered in chicken cartilage, type II is now known to be the collagen of most mammalian hyaline cartilages (*2*) and is also present in the intervertebral disk (*13*) and the notochord (*14*). In the eye, it is present in the adult vitreous body (*15*) and is synthesized by cells of the developing neural retina (*16, 17*) and of the corneal epithelium (*17, 18*).

Type III collagen, of subunit composition $[\alpha 1(III)]_3$ (*19*), is another gene product, which accounts for about 10 percent of the collagen in adult skin, 50 percent in fetal skin, and 10 to 50 percent of the collagen of blood vessels, synovial membrane, uterus, spleen, granulation tissue, and reticular connective tissues in general. It is essentially absent from normal bone, tendon, and most cartilages. Type III is believed to be the collagen of the fine reticulin fibers observed by histologists, an impression supported by electron microscopy with ferritin-labeled antibodies specific for type III (*20*). These three types, I, II, and III, are the major fibrillar collagens of the body; their native fibrils exhibit the same characteristic 64- to 67-nm axial periodicity under the electron microscope (*21*).

Basement membrane and pericellular

Table 1. Genetically distinct vertebrate collagens. At least five different molecules containing seven genetically distinct α chains are present in higher animals.

Type	Molecular formula	Native polymer	Tissue distribution	Distinctive features
I	$[\alpha 1(I)]_2\alpha 2$	Fibril	Skin, tendon, bone, dentin, fascia; widespread	Low content of hydroxylysine; few sites of hydroxylysine glycosylation; broad fibrils
II	$[\alpha 1(II)]_3$	Fibril	Cartilage, nucleus pulposus, notochord, vitreous body (*2, 11–18*)	High content of hydroxylysine; heavily glycosylated; usually thinner fibrils than type I
III	$[\alpha 1(III)]_3$	Fibril	Skin, uterus, blood vessels; "reticulin" fibers generally (*19–21*)	High content of hydroxyproline; low content of hydroxylysine; few sites of hydroxylysine glycosylation; interchain disulfides between cysteines at the carboxyl end of the helix; long carboxyl telopeptide (*57*)
IV	$[\alpha\ 1(IV)]_3$ (tentative, under dispute)	Basement lamina	Kidney glomeruli, lens capsule; Descemet's membrane; basement laminae of all epithelial and endothelial cells? (*22*)	Very high content of hydroxylysine; almost fully glycosylated; relatively rich in 3-hydroxyproline; low alanine content; retains procollagen extension pieces
V	$\alpha A(\alpha B)_2$ or $(\alpha A)_3$ and $(\alpha B)_3$	Unknown	Widespread in small amounts (*24–29*); basement lamina of smooth and striated muscle cells? exoskeleton of fibroblasts and other mesenchymal cells?	High content of hydroxylysine; heavily glycosylated; low alanine content; fails to form native fibrils in vitro

collagens. Type IV collagen is the name originally given to collagen recovered from certain thick basement membranes of specialized epithelia and endothelia that include kidney glomeruli, lens capsule, and Descemet's membrane of the eye. This collagen is not fibrillar when visualized by electron microscopy, and it has a distinctive amino acid composition with a high content of glycosylated hydroxylysine (*22*). The molecular subunits include globular regions that lack the triple helix, are cross-linked by disulfide bonds and, because they are susceptible to proteolysis, are a source of heterogeneity on analysis (*23*). For some years, a single collagen type of molecular structure $[\alpha 1(IV)]_3$ has been a popular model (*22*), but this remains controversial and the subject is now an active research topic among collagen chemists (*23*). A generally accepted molecular model is still lacking, and it may be that the various types of basal laminae have each evolved their own type of collagen.

Two newly discovered collagen α chains, called αA and αB (*24*) or A and B (*25*), have been identified as minor components in pepsin digests of various tissues, including placenta (*24*), blood vessels and skin (*25*), striated muscle (*26*), cultures of smooth muscle cells (*27*), embryonic tendons (*28*), infant calvarial bone, and infant cartilage (*29*). It was originally suspected that they derived from some type of basement membrane (*25, 27*), perhaps the amorphous material surrounding smooth muscle cells (*30*) as opposed to epithelial basal laminae. Recent evidence shows that fibroblasts and other mesenchymal cells are probably surrounded by these collagens also (*29, 31*). Undefined cell-surface collagen (*32*) and microexudate collagen (*33*) have been reported in fibroblast cultures. The latter was enriched in 3-hydroxyproline as are the αA and αB chains and type IV collagen. There is still a controversy whether the αA and αB chains derive from one or two parent molecules—that is, $\alpha A(\alpha B)_2$, or $(\alpha A)_3$ and $(\alpha B)_3$. Varying ratios of αA to αB recovered from different sources suggested two independent molecules (*25, 28, 29*), but others report a consistent 2:1 ratio for αB:αA in their preparations and conclude that the parent molecule is $\alpha A(\alpha B)_2$ (*24*). There is a growing tendency to call this AB collagen type V.

Chains with similar properties to αA and αB, called X and Y, were detected in chondrocyte cultures (*34*), and two or possibly three new collagen chains have been identified in human hyaline cartilage (*35*). One of these newly described

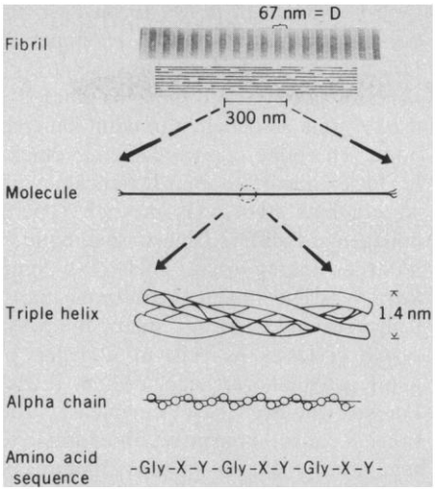

Fig. 2. Molecular features of collagen structure from primary sequence up to the fibril.

chains appears to be a highly hydroxylated and highly glycosylated form of αI(II), which may be related to the observation from studies of the amino acid sequence of αI(II) prepared from bovine cartilage (*36*) that two homologous variants of αI(II) were present in the preparation. The other two new collagen chains from cartilage resembled αA and αB in composition, but gave different electrophoretic patterns of peptides on cleavage by cyanogen bromide (CNBr) suggesting that they were genetically distinct (*35*). The bands identified as αA and αB (Fig. 3) could include the latter two newly described components.

Type I trimer. A molecule of composition $[\alpha 1(I)]_3$ has been identified as a product of various cell cultures and tumors (*34, 37*) and is present normally in skin, embryonic bones and tendons (*38*), and in rat dentin (*39*). The α chain appeared to be identical to α1(I) of type I collagen by peptide mapping, but whether it really is the same chain or a closely related genetic variant remains to be seen, as does the molecule's function. A similar molecule made by cell lines derived from the early embryo was proposed to be a distinct type, which has been referred to as type V (*40*). It would seem that the inventory of genetically distinct vertebrate collagens is by no means complete, and difficulties with nomenclature may increase as more types are discovered.

Molecules related to collagen. In addition to the above readily recognizable collagen molecules, two quite unrelated proteins contain collagen-like sequences. These are the Clq subcomponent of the complement system in the blood (*41*), and the enzyme acetylcholinesterase (*42*). Presumably the collagen-like regions are needed for their biological function, for example, by providing the molecules with a rigid segment, enabling them to self-aggregate, or to interact with collagen. Acetylcholinesterase is believed to be anchored to basement membranes by its collagenous tail (*42*).

A major development in collagen biology is the application of immunological techniques for identifying the various genetic species in situ (*43*). Specific antibodies have been prepared against each of the common molecular types and have been used to locate the collagens in tissue sections by standard immunohistochemical methods (*20, 44*).

Posttranslational Heterogeneity

Each of the different types of collagen is by definition the product of one or more distinct structural genes. There is also scope for variations in composition within a molecular type because of the modified amino acids derived on the newly translated polypeptides during biosynthesis or later in the extracellular environment. Inside the cells, the post-

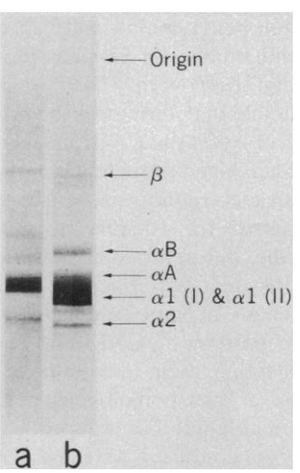

Fig. 3. Fractionation of different collagen α chains by electrophoresis in sodium dodecyl sulfate-polyacrylamide gels. This technique can resolve all known genetically distinct types of α chain except the common forms of α(I) and α1(II). If the α1(III) chain were present it would run between α1(I) and αA. (Track a) Salt-soluble collagen from lathyritic chick cartilage, extracted in the presence of protease inhibitors. (Track b) The same preparation of native collagen molecules after brief treatment with pepsin. Pepsin treatment of native molecules causes all the chains to migrate slightly farther, presumably because they have been shortened by removal of their telopeptides. The bands labeled αA and αB are identified solely on the basis of their mobility. Recent evidence from cartilage suggests that these may represent other rare genetic species distinct from αA and αB (*35*). Notice an added complexity in that two bands appear to be resolved in the αA region of track a.

translational modifications include hydroxylation of proline and lysine residues and glycosylation of hydroxylysines with galactosyl and glucosylgalactosyl side chains; outside the cell, they include conversion of lysine and hydroxylysine side chains to aldehydes, and the reactions of the aldehydes to form covalent cross-links. The enzymes prolyl 4-hydroxylase and lysyl hydroxylase attach hydroxyl groups to proline and lysine residues only in the Y position of the -Gly-X-Y- repeat (*3*). Two specific glycosyl transferases sequentially attach galactose and glucose to the hydroxyl groups of some hydroxylysines (*3*).

Although 4-hydroxyproline is the main derivative from proline, some 3-hydroxyproline is also formed. Types I, II, and III collagens contain one or two residues of 3-hydroxyproline per α chain, and these are located near the carboxyl end in α1(I), α2, and α1(II) (*8*, *45*). Types IV and V are usually richer in 3-hydroxyproline, with as many as 10 percent of the hydroxyprolines being the 3-derivative in type IV (*22*, *24*, *25*). This residue is made by a separate enzyme, prolyl 3-hydroxylase, which requires the proline to be in the X position of the -Gly-X-Y- triplet and the Y position to be already occupied with 4-hydroxyproline (*46*). The three hydroxylases and two glycosyl transferases require nonhelical chains as substrates and will not act on the triple helix (*3*, *47*). When the triple helix forms in the procollagen molecule, both hydroxylation and glycosylation stop. If helix formation is prevented or retarded by the incorporation of proline analogs into cultured cells, then extra hydroxylysine and glycosylated hydroxylysine residues appear in the synthesized collagen (*48*).

The functions of the various hydroxylated and glycosylated residues are not fully understood. Although it is established that 4-hydroxyproline stabilizes the triple helix, it is not known whether 3-hydroxyproline has the same or another function. Hydroxylysine is essential as the site for sugar attachment and for forming stable cross-links. The main function of the glycosylated hydroxylysines is still uncertain, although they do participate in cross-linking (*49*), and recent evidence suggests that native fibrils of type II collagen are more swollen with water than type I, perhaps because of the abundant carbohydrate (*50*).

Molecules of type II collagen usually contain more hydroxylysine and invariably much more glycosylated hydroxylysine than types I and III collagens. However, the extent of hydroxylation of lysine can vary substantially within each collagen type. For example, type I collagen of newly formed woven bone, both during normal development (*51*) and during repair in the adult (*52*), contains two to three times as much hydroxylysine as collagen of adult lamellar bone. The bone of hypocalcemic chicks in, for example, vitamin D deficiency also contains hydroxylysine-rich type I collagen and differs from normal bone in its cross-linking amino acids (*53*). Similarly, type II collagen of embryonic cartilage contains twice as many hydroxylysine residues as type II collagen of adult articular cartilage (*54*). It is not known whether these chemical variations occur for a purpose, for example to regulate the quality of cross-linking, or whether they are nonspecific effects of metabolic changes that alter the activities of the hydroxylases and transferases during development and aging (*55*). There is evidence, however, that glucocorticoid hormones may selectively alter the activities of the collagen hydroxylases and glycosyl transferases (*56*). This evidence supports a proposal that these enzymes, particularly prolyl hydroxylase, may be important in regulating rates of collagen biosynthesis under certain conditions (*3*).

Knowledge of Primary Structure

The task of sequencing the amino acids in the collagen α chains progressed quickly. The first data were derived from type I collagen of rat and chick (*8*), and the entire α1(I) and α2 chains of the chick are now probably sequenced. However, the most complete information applies to calf collagens (*57*, *58*). The entire sequences of the α1(I) and α2 chains of type I collagen and the α1(III) chain of type III collagen from calf are known (*58*). Extensive sequences of bovine α1(II) (*12*) and human α1(III) (*59*) chains have also been determined.

The few methionine residues in collagen (five to ten per α chain), and hence the manageable number of peptides produced by digestion with CNBr, has proved invaluable in determining the primary structure. Also, the CNBr-cleaved peptides from collagen provide, on chromatography or electrophoresis, sensitive peptide patterns for identifying and quantifying the different collagen-gene products in tissues (*60*). Each type of collagen α chain or molecule gives a characteristic pattern of CNBr-cleaved peptides, reflecting their individual sequences. One or two homologous peptides of identical length but different sequence are recovered from α1(I), α1(II), and α1(III) chains as a result of conserved methionine sites. The α1(I) chain is in fact more homologous to α1(II), with more than 80 percent of their residues identical, than to α2 where the homology is less than 70 percent (*12*).

The α chains at both ends of the collagen molecule (telopeptides) terminate in short sequences that are not triple helical (*61*); these short sequences contain 10 to 20 residues and lack glycine at every third residue. At the carboxyl terminus of the native type III molecule the nonhelical telopeptides are longer, with about 60 residues, and begin with a -Cys-Cys- (Cys, cysteine) sequence at the very end of the triple helix (*57*). These cysteine residues are not removed from native type III collagen molecules solubilized by pepsin, which clips the telopeptides farther in than the aldehydic cross-linking site but cannot attack the triple helix. Therefore, in molecules of type III prepared by pepsin digestion all three α chains are usually disulfide-bonded by these cysteines. This unusual property can be used to distinguish type III collagen from types I and II (*2*).

Biosynthesis—Procollagen

The biggest impact on collagen biochemistry in the last decade has been the discovery of genetic polymorphism and the existence of procollagen. The procollagen molecule is a larger, biosynthetic precursor that is trimmed by specific proteases to form the molecule eventually found packed in fibrils outside the cell (*3*, *4*). For 2 or 3 years procollagen was thought to be extended at its amino terminus only (*62*). It was later discovered to be extended at both ends, with the largest extension at the carboxyl terminus (*63*). Thus each procollagen α1(I) chain of type I procollagen has an amino terminal extension with a molecular weight of about 15,000 and a carboxyl terminal extension with a molecular weight of about 35,000. The pro α2 chain appears to have a shorter amino terminal extension than pro α1(I) (*4*). The extension piece (propeptide) at the carboxyl end contains several cysteine residues, some of which cross-link the chains within the procollagen molecule. The amino terminal propeptides of α1(I) and α2 each include a sequence of about 50 residues that is rich in hydroxyproline and together form a particularly stable segment of triple helix (*64*). Its function is unknown. This helical segment is separated from the main helix of the molecule by a short globular region, the eventual site of cleavage. The main globular

end of the amino terminal propeptides also contains cysteine residues, but in type I and type II procollagens they form only intrachain disulfide bonds (*3, 4*). Interchain disulfide bonds are present, however, at both ends of the type III procollagen molecule (*65*). The carboxyl terminal propeptide is glycosylated with sugars, notably *N*-acetylglucosamine and mannose (*66*), not found in the body of the collagen molecule and more characteristic of conventional glycoproteins.

Several important functions have been proposed for these propeptides (*62, 67*). One is to prevent formation of fibrils before the procollagen molecule reaches the desired extracellular site, that is, a transport function. Other likely functions are (i) to direct assembly of the correct combination of pro α chains into molecules, (ii) to regulate triple helix formation, and (iii) to direct fibrillogenesis outside the cell. Feedback inhibition of collagen synthesis by the propeptides has also been proposed and recently confirmed (*68*).

For types I, II, and III collagen molecules the propeptides are removed by specific procollagen peptidases, of which there are at least two enzymes, one for each end (*69*). Each type of collagen may also have its own characteristic peptidases (*4*). Both pieces are thought to be cleaved outside the cell, although the amino terminus seems to go first (*70*). A function of the longer telopeptides remaining at the carboxyl terminus of native molecules of type III collagen may be to prevent thick fibrils from forming. This possibility is supported by evidence that type IV collagen retains most of its globular propeptides for its structural role in basement laminae, which lack any obvious fibrillar structure (*22, 23*). The type IV propeptides probably contribute at least partly to the firmly bound glycoprotein fraction of basement laminae.

Messenger RNA (mRNA) for procollagen type I has been prepared and characterized quite extensively. From analyses of the translated product of such preparations, it is now known that procollagen chains, in keeping with insulin and other proteins designed for export from the cell, are synthesized on the ribosome starting with a short leader or signal sequence of 20 or more residues (*4*). This piece, at the extreme amino terminus, is rich in hydrophobic amino acids and is thought to channel the nascent polypeptide through the membrane into the cisternae of the rough endoplasmic reticulum, where it is immediately cleaved off by a protease. The entire polypeptide translated from collagen mRNA by cell-free systems in vitro is therefore being called the prepro α chain (*4*).

The first reported attempts to locate the collagen genome on human chromosomes with the use of hybrids of human and mouse cells have produced conflicting results. Genes for type I collagen were traced to chromosome 7 by one group (*71*) and to chromosome 17 by another (*72*). Although different types collagen may be coded for on different chromosomes (*73*), it seems less likely that one type will be coded for on more than one chromosome.

A recent finding is that monovalent ionophores, such as monensin, selectively block the export of procollagen from cultured fibroblasts; yet the completed molecules continue to be synthesized and appear to accumulate in membrane-bound Golgi vacuoles (*74*). Export of the protein fibronectin is also blocked, suggesting that the two proteins share the same secretory route from the cell.

Molecular Packing in Collagen Fibrils

The primary structure of the pro α chains of collagen seems to contain all the equipment needed to fold them into native molecules, to pack the molecules into fibrils (and perhaps for type IV collagen into basement laminae), and to determine the type of intermolecular cross-links eventually formed. Primary structure may also dictate how the fibrils interact with interfibrillar glycoproteins and proteoglycans that are associated with collagen in most tissues. As more is learned of the biochemistry of collagen and proteoglycans, the concept strengthens that extracellular matrices must largely self-assemble in situ, regulated primarily by the physical and chemical properties of the macromolecular monomers.

The precise axial register of collagen molecules in fibrils was established 20 years ago by Hodge and Petruska as a near quarter-stagger with overlap (*75*). The side-to-side lattice, and hence the manner of packing of molecules in three dimensions, is still a mystery, however (*76*). Axially, each molecule is displaced from all its neighbors by 0, 1, 2, 3, or 4 D periods where D equals 1/4.4 of the molecular length. This D period, or axial stagger, has now been precisely defined at 234 ± 1 residues for type I collagen (*77*). Fibrils of types I, II, and III collagens show a similar periodicity by electron microscopy which implies that they share this axial stagger. No lateral substructure is obvious in fibrils by electron microscopy, although a microfibril subunit has long been suspected. Several models for microfibrils have been proposed, and, one, the Smith five-stranded microfibril (*76, 78*), is still a favorite. It is supported experimentally by x-ray diffraction analyses of the rat tail tendon which show an equatorial reflection indicating a 3.8-nm lateral repeat, which is the theoretical width of such a microfibril (*76*). Other types of collagen fail to give this diffraction line, and it is prominent in the rat tail tendon only on stretching (*79*). Its biological significance is still in question, therefore.

An alternative proposal is that there is no ordered side-to-side lattice of molecules within a fibril, the only strict spatial register being the precise stagger axially. This predicted paracrystalline structure with an amorphous lateral organization (*80*) has been likened to a type A smectic liquid crystal (*81*). Poor lateral order would account for the diffuse intensities on the equator, compared with the sharp lines on the meridian, which most native collagens give on low-angle, x-ray diffraction.

The diameter of collagen fibrils and their weave vary greatly, depending on the tissue, even for a single genetic type of collagen. For example, type I collagen of tendon is expressed as thick, uniform fibrils packed in parallel bundles, whereas type I collagen of bone is a contorted weave of finer fibrils that are heavily mineralized in the native tissue. The basic physical properties of collagens in tendon and skeletal tissues also differ (*82*). Collagen in cartilage or decalcified bone swells little, and virtually none dissolves in dilute acid, whereas tendon collagen swells about 100-fold, and in rat tail tendon all the collagen eventually dissolves. What causes identical proteins that share the same amino acid sequence and most details of primary structure to differ so markedly in physical properties? The differences seem to be mainly due to tissue-specific variations in chemical structure and molecular location of intermolecular cross-links.

Intermolecular Cross-Linking

Cross-linking amino acids are vital for the strength and normal function of collagen fibrils. Their importance is hard to overemphasize. Their chemistry has been reviewed by Tanzer (*83, 84*). Although various compounds have now been identified, little is known still about their relative functions.

Collagen molecules cross-link spontaneously when they pack into native fi-

brils (*83*). The cross-links derive from aldehydes made extracellularly from lysyl and hydroxylysyl residues by lysyl oxidase (*85*). This oxidative deamination is the only enzymic step needed. Most of the aldehydes are located at the ends of the α chains, with one in each telopeptide sequence (*58, 86*). All the genetic types of vertebrate collagen and most invertebrate collagens exhibit this cross-linking mechanism (*6, 84, 87*). Collagen cross-linking can be blocked in experimental animals by feeding them β-aminopropionitrile and related compounds, which inhibit lysyl oxidase and produce the condition called lathyrism.

Most cross-linking amino acids were identified after chemical reduction by tritiated sodium borohydride (*88*), which stabilizes them during acid hydrolysis and labels them with tritium for isolation (*83, 84, 89*). The simplest intermolecular cross-links are aldimine addition products between the aldehydes and unmodified hydroxylysyl or lysyl side chains. However, an intramolecular cross-link, allysine aldol, was the first compound to be identified (*90*), being abundant in soluble collagen from tendon and skin. Its function has proved the hardest to understand, although it is also prominent in elastin as an intermediate in the formation of desmosine cross-links (*91*).

In native collagen fibrils of skin or tendon the main reducible compound is a complex cross-link apparently derived by reaction of the intramolecular allysine aldol with both a histidine residue and a hydroxylysine residue. Although there is controversy whether the histidine reacts naturally or as an artifact of borohydride reduction (*92*), the few histidine residues in collagen are concentrated at cross-linking sites and are probably important in cross-linking reactions, if only as proton donors or acceptors. Moreover, a stable histidinyl cross-link has been recovered from collagen without borohydride reduction (*93*).

The main labeled compound on reacting collagen of bone or cartilage with tritiated sodium borohydride is dihydroxylysinonorleucine. It derives from two hydroxylysine residues and is fairly stable even without reduction because the initial aldimine rearranges to a ketoamine, hydroxylysino-5-keto-norleucine (*84*) (Fig. 4). In these skeletal tissues most of it is present as the glucosylgalactosyl or galactosyl derivative (Fig. 4), the nonaldehydic hydroxylysine precursor having been a glycosylated site. The function of the sugar is unknown.

The reducible cross-links gradually disappear from adult connective tissues, which suggests they are long-lived intermediates that can react further. But the anticipated mature or final cross-links have proved elusive. Indirect evidence suggesting that the aldimine cross-links were naturally reduced (*94*) was not substantiated by direct analysis in that naturally reduced cross-links were not detected (*95*). Other mechanisms of maturation have been proposed (*96, 97*), but none provide a wholly convincing mechanism supported by stoichiometric yields of the predicted mature compounds.

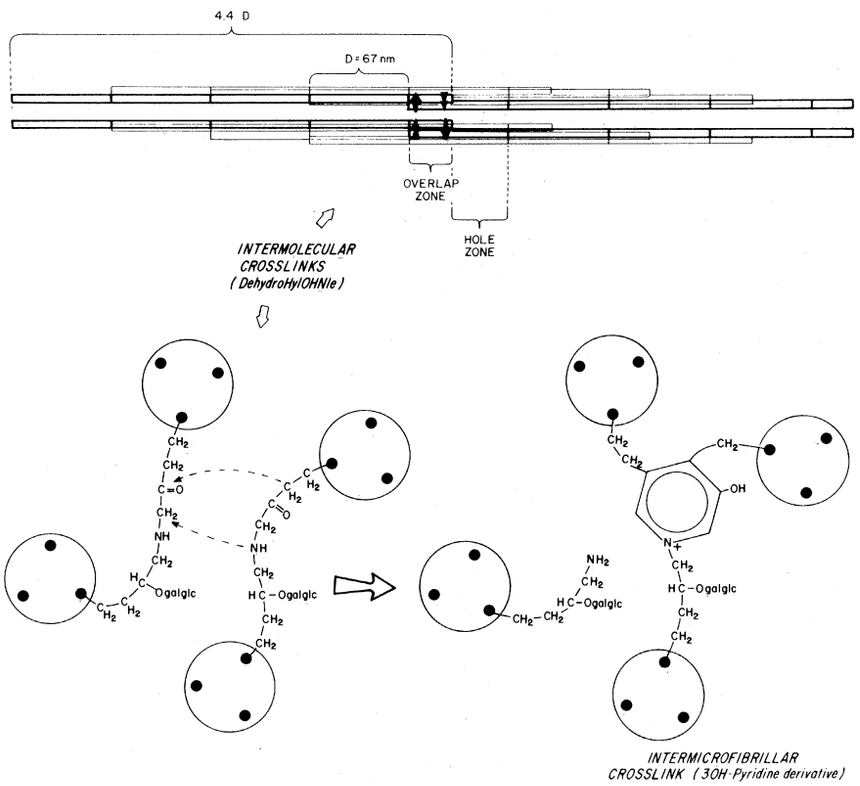

Fig. 4. Proposed mechanism (*100*) for the formation of the trivalent 3-hydroxypyridinium cross-links of collagen by interaction of two divalent cross-links. The dehydro-dihydroxylysinonorleucines are shown glycosylated and in their ketoamine configuration, their predominant form in collagens of bone and cartilage. Other reaction mechanisms are possible, although the one indicated is favored, and the product agrees with the 1,4,5 side-chained structure suggested by the data of Fujimoto (*99*). The upper molecular diagram shows how difunctional cross-links in adjacent five-stranded microfibrils can be spatially in close apposition, favoring their chemical interaction. Single-stranded filaments of head-to-tail overlapping molecules could interact equally effectively wherever they are registered as shown here.

New, Nonreducible Cross-Links

A new, fluorescent cross-linking amino acid was recently isolated from tendon collagen without reduction (*98*). It was shown to be a 3-hydroxypyridinium compound, probably derived from three hydroxylysyl residues (*99*). The same compound is especially abundant in collagens of bone and cartilage; adult articular cartilage contains about one residue per collagen molecule, with about a fifth as much in adult bone (*100*). As the dihydroxylysinonorleucine content of cartilage falls with age, the 3-hydroxypyridinium cross-link increases proportionately (*100*), suggesting a precursor-product relationship. The molar ratio of the 3-hydroxypyridinium cross-link to dihydroxylysinonorleucine reaches about 100:1 in adult cartilage (*100*). Radiochemical studies on collagen synthesized by rabbit cartilage in vivo proved that the hydroxypyridinium cross-links were derived from lysine and supported their product-precursor relationship with dihydroxylysinonorleucine (*101*).

Figure 4 projects a way that two of the ketoamine cross-links might interact to form the 3-hydroxypyridinium compound. From considerations of chemistry, molecular packing, and the observed physical properties of the skeletal tissue collagens this novel two-stage pathway of cross-linking is appealing. Reducible cross-links will be closely apposed, favoring such a reaction, wherever adjacent molecules have no axial stagger. This could occur, for instance, between abutting five-stranded microfibrils as shown in Fig. 4, or, if this packing model proves incorrect, between monomeric

strands of head-to-tail overlapping molecules that are in register (for example, the heavily outlined molecules alone in the upper part of Fig. 4). Electron microscopic evidence suggests that fibrils may assemble from small clusters of registered procollagen molecules that aggregate into an $n \times D$ (67 nm) staggered array (*102*). Such fibrils would effectively consist of staggered clusters of registered monomeric threads.

The proposed reaction can help explain why skeletal-tissue collagens resist swelling, since it would introduce extensive lateral cross-linking between all elements within fibrils. It would also allow cross-linking at contacts between fibrils, which is of particular significance for understanding the properties of cartilage with its random meshwork of fibrils and entrapped proteoglycans.

Collagen and Disease

Heritable disorders. The rapid advances in collagen biochemistry have promoted a surge of interest and new understanding on collagen pathology in disease (*103*). The long-anticipated molecular defects of collagen in various heritable diseases have begun to be defined. Thus the basic lesions in three recessive types and one sex-linked type of Ehlers-Danlos syndrome, a heritable disorder of connective tissue featuring hyperextensible skin and joints, have been traced to collagen (*104*). Three of the lesions affect enzymes responsible for posttranslational events: lysyl hydroxylase, lysyl oxidase, and procollagen peptidase; the fourth is a failure to make type III collagen and is therefore analogous to the thalassemia syndromes of hemoglobin pathology. Other inherited defects in collagen structure are anticipated and are under active study. A defect in synthesis of type I collagen is suspected in certain types of osteogenesis imperfecta (*105*).

Acquired disorders. In addition to the genetic disorders which can reveal so much about the relationship between structure and function of normal collagen, many acquired disorders indicate possible secondary abnormalities in collagen metabolism. For example, interest has focused on the synthesis of cartilage collagen in osteoarthritis (*106*), on altered collagen behavior in connective tissues and vascular basement membranes in diabetes (*22, 107*), and on the role of collagen in platelet aggregation, thrombus formation, and the blood vessel changes seen in cardiovascular disease (*108*). The influence of the collagen environment on cell differentiation and proliferation has important implications in developmental biology (*5*) and in many diseases, including cancer.

Likewise, the properties of tissue collagenases and their essential role in the degradation and turnover of collagen form a major research topic in themselves (*109*). Collagenases have special significance for understanding chronic inflammatory diseases such as rheumatoid arthritis. In this condition, articular cartilage and other fibrous tissues of the joint are eroded by proliferating cells of the synovial membrane, which degrade the collagenous tissues with neutral proteinases and a specific collagenase. The stimulus for this pathological behavior is unknown.

Aging. Because collagen turns over very slowly in most adult tissues, progressive structural changes in collagen, such as an increase in cross-linking, were formerly evoked as an underlying factor in the body's aging process. Although no longer taken seriously as a central cause of aging, age-related changes in the physical and chemical properties of collagen may nevertheless have serious effects on the performance of many tissues, including their ability to resist disease.

Unresolved Questions

Long-standing puzzles in collagen biology include what causes the collagen of bone and dentin to calcify while the same molecular type of collagen in other tissues does not. The answer probably lies not in the collagen directly but in associated matrix constituents, for example, the phosphoproteins that have long been proposed as regulators of mineral deposition (*82, 110*), or the bone protein containing γ-carboxyglutamate that is suspected of having some regulatory role in mineral metabolism (*111*). The collagen cannot be completely discounted, however. An old observation that the $\alpha 2$ chain of bone collagen contains organically bound phosphate has been confirmed, and the residue has been tentatively identified as phosphoglutamic acid (*112*).

Specifically how collagen fibrils of different types interact with other macromolecules of the extracellular matrix, notably proteoglycans and structural glycoproteins, as well as with cell membranes and blood-platelet membranes are broad questions of great medical and biological importance.

It must be assumed that the various types of collagen have evolved to suit different biological functions. However, despite extensive data on primary structure, the distinctive functions of each collagen type and the special molecular properties that underlie them are essentially unknown. Comparison of diverse species to determine the relative biological ages of the different collagen types, and which of their chemical features are rigorously conserved, may prove useful here.

References and Notes

1. G. N. Ramachandran and A. H. Reddi, Eds., *Biochemistry of Collagen* (Plenum, New York, 1976).
2. E. J. Miller, *Mol. Cell. Biochem.* **13**, 165 (1976).
3. D. J. Prockop, K. I. Kivirikko, L. Tuderman, N. A. Guzman, *N. Engl. J. Med.* **301**, 13 and 77 (1979); D. J. Prockop, R. A. Berg, K. I. Kivirikko, J. Uitto, in (*1*), pp. 163-273.
4. J. H. Fessler and L. I. Fessler, *Annu. Rev. Biochem.* **47**, 129 (1978); P. Bornstein and W. Traub, in *The Proteins*, H. Neurath and R. L. Hill, Eds. (Academic Press, New York, 1979), vol. 4, p. 411.
5. J. Gross, *Harvey Lect.* **68**, 351 (1974).
6. E. Adams, *Science* **202**, 591 (1978).
7. G. N. Ramachandran and C. Ramakrishnan, in (*1*), pp. 45-84.
8. K. A. Piez, in (*1*), pp. 1-44.
9. R. A. Berg and D. J. Prockop, *Biochem. Biophys. Res. Commun.* **52**, 115 (1973); S. Jimenez, M. Harsch, J. Rosenbloom, *ibid.*, p. 106.
10. J. P. Segrest and L. W. Cunningham, *Biopolymers* **12**, 825 (1973); D. W. L. Hukins and J. Woodhead-Galloway, *Biochem. Biophys. Res. Commun.* **70**, 413 (1976).
11. E. J. Miller and V. J. Matukas, *Proc. Natl. Acad. Sci. U.S.A.* **64**, 1264 (1969).
12. G. Francis, W. T. Butler, J. E. Finch, Jr., *Biochem. J.* **175**, 921 (1978).
13. D. R. Eyre, *Int. Rev. Connect. Tissue Res.* **8**, 227 (1979).
14. E. J. Miller and M. B. Mathews, *Biochem. Biophys. Res. Commun.* **60**, 424 (1974).
15. D. A. Swann, I. J. Constable, E. Harper, *Invest. Ophthalmol.* **11**, 735 (1972); D. A. Swann and S. S. Sotman, *Biochem. J.*, in press.
16. G. N. Smith, Jr., T. F. Linsenmayer, D. A. Newsome, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4420 (1976).
17. K. von der Mark, H. von der Mark, R. Timpl, R. L. Trelstad, *Dev. Biol. Stand.* **59**, 75 (1977).
18. T. F. Linsenmayer, G. N. Smith, Jr., E. D. Hay, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 39 (1977).
19. E. Chung and E. J. Miller, *Science* **183**, 1200 (1974); E. H. Epstein, Jr., *J. Biol. Chem.* **249**, 3225 (1974).
20. S. Gay, in *Collagen-Platelet Interaction*, H. Gastpar, K. Kühn, R. Marx, Eds. (Schattauer, New York, 1978), pp. 171-179.
21. K. Kühn and K. von der Mark, in *ibid.*, pp. 123-126; E. D. Hay, D. L. Hasty, K. L. Kiehnau, in *ibid.*, pp. 129-151.
22. N. A. Kefalides, Ed., *Biology and Chemistry of Basement Membranes* (Academic Press, New York, 1978).
23. B. G. Hudson and R. G. Spiro, *J. Biol. Chem.* **247**, 4229 (1972); J. R. Daniels and G. H. Chu, *ibid.* **250**, 353 (1975); T. Sato and R. G. Spiro, *ibid.* **251**, 4062 (1976); R. L. Trelstad and K. R. Lawley, *Biochem. Biophys. Res. Commun.* **76**, 376 (1977); S. N. Dixit, *FEBS Lett.* **85**, 153 (1978); D. Schwartz and A. Veis, *ibid.*, p. 326; K. Tryggvason and K. I. Kivirikko, *Nephron* **21**, 230 (1978). Two distinct α chains called C and D with the characteristic composition of type IV collagen have been isolated from pepsin digests of placental membranes and have been detected in lens capsule. From their molar ratios a molecular composition of C_2D was a highly tentative suggestion [T. F. Kresina and E. J. Miller, *Biochemistry* **18**, 3089 (1979)]. Other recent reports also indicate two genetically distinct α chains in type IV preparations. [R. Timpl, G. R. Martin, P. Bruckner, G. Wick, H. Weidemann, *Eur. J. Biochem.* **84**, 43 (1978); H. Sage, R. G. Woodbury, P. Bernstein, *J. Biol. Chem.* **254**, 9893 (1979); S. N. Dixit, *FEBS Lett.* **106**, 379 (1979)].
24. R. E. Burgeson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2579 (1976).

25. E. Chung, R. K. Rhodes, E. J. Miller, *Biochem. Biophys. Res. Commun.* **71**, 1167 (1976).
26. V. C. Duance, D. J. Restall, H. Beard, F. J. Bourne, A. J. Bailey, *FEBS Lett.* **79**, 248 (1977).
27. R. Mayne, M. S. Vail, E. J. Miller, *Biochemistry* **17**, 446 (1978).
28. S. A. Jimenez, R. Yankowski, R. I. Bashey, *Biochem. Biophys. Res. Commun.* **81**, 1298 (1978).
29. R. K. Rhodes and E. J. Miller, *Biochemistry* **17**, 3442 (1978).
30. R. Ross, *J. Cell Biol.* **50**, 172 (1971).
31. S. Gay and E. J. Miller, *Collagen in the Physiology and Pathology of Connective Tissue* (Fisher, New York, 1978).
32. J. R. Lichtenstein, E. A. Bauer, R. Hoyt, H. J. Wedner, *J. Exp. Med.* **144**, 145 (1976).
33. K. J. Lembach, R. E. Branson, P. B. Hewgley, L. W. Cunningham, *Eur. J. Biochem.* **72**, 379 (1977).
34. P. D. Benya, S. R. Padilla, M. E. Nimni, *Biochemistry* **16**, 865 (1977).
35. R. E. Burgeson, D. W. Hollister, B. Kwon, G. Saxon, A. Ragheb, *Biochem. Biophys. Res. Commun.* **87**, 1124 (1979).
36. W. T. Butler, J. E. Finch, Jr., E. J. Miller, *J. Biol. Chem.* **252**, 639 (1977).
37. P. K. Müller, W. N. Meigel, B. G. Pontz, K. Raisch, *Hoppe-Seylers's Z. Physiol. Chem.* **355**, 985 (1974); R. Mayne, M. S. Vail, E. J. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4511 (1975); L. Moro and B. D. Smith, *Arch. Biochem. Biophys.* **182**, 33 (1977).
38. S. Uitto, *Arch. Biochem. Biophys.* **192**, 371 (1979); S. A. Jimenez, R. I. Bashey, M. Benditt, R. Yankowski, *Biochem. Biophys. Res. Commun.* **78**, 1354 (1977).
39. E. J. Munksgaard, M. Rhodes, R. Mayne, W. T. Butler, *Eur. J. Biochem.* **92**, 609 (1978); M. Wohllebe and D. J. Carmichael, *ibid.*, p. 183.
40. C. D. Little and R. L. Church, *Arch. Biochem. Biophys.* **190**, 632 (1978).
41. R. R. Porter and K. B. M. Reid, *Nature (London)* **275**, 699 (1978).
42. T. L. Rosenberry and J. M. Richardson, *Biochemistry* **16**, 3550 (1977).
43. R. Timpl, in (*1*), pp. 319–375; _____, G. Wick, S. Gay, *J. Immunol. Methods* **18**, 165 (1977); H. K. Beard, W. P. Faulk, L. B. Conochi, L. E. Glynn, *Progr. Allergy* **22**, 45 (1977).
44. H. S. Nowack, S. Gay, G. Wick, U. Becker, R. Timpl, *J. Immunol. Methods* **12**, 117 (1976).
45. E. J. Miller and L. G. Lunde, *Biochemistry* **12**, 3153 (1973); J. M. Lane and E. J. Miller, *ibid.* **8**, 2134 (1969); A. H. Kang, S. Igarashi, J. Gross, *ibid.*, p. 3200.
46. R. M. Gryder, M. Lamon, E. Adams, *J. Biol. Chem.* **250**, 2470 (1975); K. Tryggvason, J. Risteli, K. I. Kivirikko, *Biochem. Biophys. Res. Commun.* **76**, 275 (1977).
47. A. Oikarinen, H. Anttinen, K. I. Kivirikko, *Biochem. J.* **156**, 545 (1976).
48. _____, *ibid.* **160**, 639 (1976); V.-J. Uitto, J. Uitto, W. W.-K. Kao, D. J. Prockop, *Arch. Biochem. Biophys.* **185**, 214 (1978).
49. D. R. Eyre and M. J. Glimcher, *Biochem. J.* **135**, 393 (1973).
50. M. D. Grynpas, D. R. Eyre, D. A. Kirschner, *Trans. Orthop. Res. Soc.* **5**, 13 (1980).
51. E. J. Miller, G. R. Martin, K. A. Piez, M. J. Powers, *J. Biol. Chem.* **242**, 5481 (1967).
52. R. D. Ellis, K. Hayashi, J. B. Lian, F. Shapiro, M. J. Glimcher, *Trans. Orthop. Res. Soc.* **1**, 86 (1976); M. J. Glimcher, F. D. Shapiro, R. D. Ellis, D. R. Eyre, *J. Bone Joint Surg.*, in press.
53. I. R. Dickson, D. R. Eyre, E. Kodicek, *Biochim. Biophys. Acta* **588**, 169 (1979).
54. D. R. Eyre and H. Muir, *ibid.* **492**, 29 (1977); S. Slavin, V. Shurlan, D. R. Eyre, M. J. Glimcher, *Trans. Orthop. Res. Soc.* **4**, 178 (1979).
55. H. Anttinen, A. Oikarinen, K. I. Kivirikko, *Clin. Chim. Acta* **76**, 95 (1977).
56. J. Risteli, *Biochem. Pharmacol.* **26**, 1295 (1977); A. Oikarinen, *Biochem. J.* **164**, 533 (1977); R. A. Newman and K. R. Cutroneo, *Mol. Pharmacol.* **14**, 185 (1978).
57. P. P. Fietzek, in *Collagen-Platelet Interaction*, H. Gastpar, K. Kühn, R. Marx, Eds. (Schattauer, New York, 1978), pp. 189–199.
58. P. P. Fietzek and K. Kühn, *Int. Rev. Connect. Tissue Res.* **7**, 1 (1976); H. Hofmann, P. P. Fietzek, K. Kühn, *J. Mol. Biol.* **125**, 137 (1978); Fietzek and co-workers have reported the complete sequence of the calf α1(III) chain in six papers [P. P. Fietzek, H. Allman, J. Rauterberg, W. Henkel, E. Wachter, K. Kühn, *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 809 (1979); H. Dewes, P. P. Fietzek, K. Kühn, *ibid.*, p. 821; H. Bentz, P. P. Fietzek, K. Kühn, *ibid.*, p. 833; H. Lang, R. W. Glanville, P. P. Fietzek, K. Kühn, *ibid.*, p. 841; H. Dewes, P. P. Fietzek, K. Kühn, *ibid.*, p. 851; H. Allman, P. P. Fietzek, R. W. Glanville, K. Kühn, *ibid.*, p. 861].
59. J. M. Seyer and A. H. Kang, *Biochemistry* **17**, 3404 (1978).
60. D. R. Eyre, in *Collagen-Platelet Interaction*, H. Gastpar, K. Kühn, R. Marx, Eds. (Schattauer, New York, 1978), pp. 189–199; D. R. Eyre and H. Muir, *Biochem. J.* **157**, 267 (1976).
61. A. L. Rubin, M. P. Drake, P. F. Davison, D. Pfahl, P. T. Speakman, F. O. Schmitt, *Biochemistry* **4**, 181 (1965).
62. P. Bornstein, *Annu. Rev. Biochem.* **43**, 567 (1974); G. R. Martin, P. H. Byers, K. A. Piez, *Adv. Enzymol.* **42**, 167 (1975).
63. M. L. Tanzer, R. L. Church, J. A. Yaeger, D. E. Wampler, E. D. Park, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3009 (1974).
64. D. Hörlein and P. P. Fietzek, *Arch. Int. Physiol.* **84** (Suppl. 3), 3 (1976); U. Becker, R. Timpl, O. Helle, D. J. Prockop, *Biochemistry* **15**, 2853 (1976); J. Engel, P. Bruckner, U. Becker, R. Timpl, B. Rutschmann, *ibid.* **16**, 4026 (1977).
65. H. Nowack, B. R. Olsen, R. Timpl, *Eur. J. Biochem.* **70**, 205 (1976); P. Bruckner, H. P. Bachinger, R. Timpl, J. Engel, *Eur. J. Biochem.* **90**, 595 (1978).
66. C. C. Clark and N. A. Kefalides, *J. Biol. Chem.* **253**, 47 (1978); B. R. Olsen, N. A. Guzman, J. Engel, C. Condit, S. Aase, *Biochemistry* **16**, 3030 (1977).
67. M. E. Grant and D. S. Jackson, *Essays Biochem.* **12**, 77 (1976).
68. J. R. Lichtenstein, G. R. Martin, L. D. Kohn, P. H. Byers, V. A. McKusick, *Science* **182**, 298 (1973); a recent report shows that NH_2-terminal propeptides inhibit collagen synthesis by calf fibroblasts in culture [M. Wiestner, T. Krieg, D. Hörlein, R. W. Glanville, P. P. Feitzek, P. K. Müller, *J. Biol. Chem.* **254**, 7016 (1979)].
69. J. M. Davidson, L. S. G. McEneany, P. Bornstein, *Biochemistry* **14**, 5188 (1975).
70. N. P. Morris, L. I. Fessler, A. Weinstock, J. H. Fessler, *J. Biol. Chem.* **250**, 5719 (1975).
71. B. Sykes and E. Solomon, *Nature (London)* **272**, 548 (1978).
72. C. V. Sundar Raj, R. L. Church, L. A. Klobutcher, F. H. Ruddle, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4444 (1977).
73. C. V. Sundar Raj, R. L. Church, R. P. Creagan, F. H. Ruddle, *J. Cell Biol.* **70**, A78 (1976).
74. N. Uchida, H. Smilowitz, M. L. Tanzer, *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 1868 (1979).
75. A. J. Hodge and J. A. Petruska, in *Aspects of Protein Structure*, G. N. Ramachandran, Ed. (Academic Press, New York, 1963), pp. 289–300.
76. A. Miller in (*1*), pp. 85–136.
77. D. J. S. Hulmes, A. Miller, D. A. D. Parry, K. A. Piez, J. Woodhead-Galloway, *J. Mol. Biol.* **79**, 137 (1973).
78. J. W. Smith, *Nature (London)* **219**, 157 (1968); K. A. Piez and B. L. Trus, *J. Mol. Biol.* **122**, 419 (1978).
79. D. W. L. Hukins, *Biochem. Biophys. Res. Commun.* **77**, 335 (1977).
80. T. Nemetschek and R. Hosemann, *Colloid Polym. Sci.* **251**, 1044 (1975).
81. D. W. L. Hukins and J. Woodhead-Galloway, *Mol. Cryst. Liq. Cryst.* **41**, 33 (1977); D. W. L. Hukins, *J. Theor. Biol.* **71**, 661 (1978).
82. M. J. Glimcher and S. M. Krane, in *Treatise on Collagen*, G. N. Ramachandran and B. S. Gould, Eds. (Academic Press, New York, 1968), vol. 2B, pp. 68–251; M. J. Glimcher, in *Handbook of Physiology—Endocrinology*, G. D. Aurbach, Ed. (Williams & Wilkins, Baltimore, 1976), vol. 7, pp. 25–116.
83. M. L. Tanzer, *Science* **180**, 561 (1973).
84. _____, in (*1*), pp. 137–162.
85. R. C. Siegel, *Int. Rev. Connect. Tissue Res.* **8**, 73 (1979).
86. P. M. Gallop and M. A. Paz, *Physiol. Rev.* **55**, 418 (1975).
87. M. L. Tanzer and N. A. Kefalides, *Biochem. Biophys. Res. Commun.* **51**, 775 (1973).
88. O. O. Blumenfeld and P. M. Gallop, *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1260 (1966).
89. M. L. Tanzer, *Biochim. Biophys. Acta* **133**, 584 (1967); A. J. Bailey, *ibid.* **160**, 447 (1968).
90. M. Rojkind, O. O. Blumenfeld, P. M. Gallop, *J. Biol. Chem.* **241**, 1530 (1966); P. Bornstein and K. A. Piez, *Biochemistry* **5**, 3460 (1966).
91. W. R. Gray, in *Elastin and Elastic Tissue*, L. B. Sandberg, W. R. Gray, C. Franzblau, Eds. (Plenum, New York, 1977), pp. 285–290.
92. A. J. Bailey, S. P. Robins, G. Balian, *Nature (London)* **251**, 105 (1974).
93. T. J. Housley, M. L. Tanzer, E. Henson, P. M. Gallop, *Biochem. Biophys. Res. Commun.* **67**, 824 (1975).
94. G. Mechanic, P. M. Gallop, M. L. Tanzer, *ibid.* **45**, 644 (1971).
95. D. R. Eyre and M. J. Glimcher, *Proc. Soc. Exp. Biol. Med.* **144**, 400 (1973); S. P. Robins, M. Shimokomaki, A. J. Bailey, *Biochem. J.* **131**, 771 (1973).
96. N. R. Davis, O. M. Risen, G. A. Pringle, *Biochemistry* **14**, 2031 (1975).
97. A. J. Bailey, M. H. Ranta, A. C. Nicholls, S. M. Partridge, D. F. Elsden, *Biochem. Biophys. Res. Commun.* **78**, 1403 (1977).
98. D. Fujimoto, K.-Y. Akiba, N. Nakamura, *ibid.* **76**, 1124 (1977); D. Fujimoto and T. Moriguchi, *J. Biochem. (Tokyo)* **83**, 863 (1978).
99. D. Fujimoto, T. Moriguchi, T. Ishida, H. Hayashi, *Biochem. Biophys. Res. Commun.* **84**, 52 (1978).
100. D. R. Eyre and H. Oguchi, *ibid.* **92**, 403 (1980).
101. D. R. Eyre, F. D. Shapiro, C. M. Creasman, in preparation.
102. R. R. Bruns, D. J. S. Hulmes, S. F. Therrien, J. Gross, *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 313 (1979).
103. C. M. Lapière and B. Nusgens, in (*1*), pp. 377–447; J. Uitto and J. R. Lichtenstein. *J. Invest. Dermatol.* **66**, 59 (1976).
104. V. A. McKusick, *Heritable Disorders of Connective Tissue* (Mosby, St. Louis, ed. 4, 1972); D. W. Hollister, *Pediatr. Clin. North Am.* **25**, 575 (1978).
105. B. Sykes, M. J. O. Francis, R. Smith, *N. Engl. J. Med.* **296**, 1200 (1977); R. P. Penttinen, J. R. Lichtenstein, G. R. Martin, V. A. McKusick, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 586 (1975); P. K. Müller, K. Raisch, K. Matzen, S. Gay, *Eur. J. Pediat.* **125**, 29 (1977). L. Peltonen, A. Palotie, T. Hayashi, D. J. Prockop, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 162 (1980).
106. K. Deshmukh and M. E. Nimni, *Biochem. Biophys. Res. Commun.* **53**, 424 (1973); M. E. Nimni and K. Deshmukh, *Science* **181**, 751 (1973); S. Gay, P. K. Müller, C. Lemmen, K. Remberger, K. Matzen, K. Kühn, *Klin. Wochenschr.* **54**, 969 (1976); S. Gay (*20*); D. R. Eyre, C. A. McDevitt, M. E. J. Billingham, H. Muir, *Biochem. J.*, in press.
107. R. G. Spiro, *N. Engl. J. Med.* **288**, 1337 (1973).
108. R. Ross and J. A. Glomset, *ibid.* **295**, 369 (1976).
109. J. Gross, in (*1*), pp. 275–317; E. D. Harris and S. M. Krane, *N. Engl. J. Med.* **291**, 557, 605, and 652 (1974).
110. A. Veis, A. R. Spector, H. Zamoscianyk, *Biochim. Biophys. Acta* **257**, 404 (1972); W. T. Butler, J. E. Finch, C. V. DeSteno, *ibid.*, p. 167; S. L. Lee, A. Veis, T. Glonek, *Biochemistry* **16**, 2971 (1977).
111. P. V. Hauschka, J. B. Lian, P. M. Gallop, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3925 (1975); P. A. Price, A. S. Otsuka, J. P. Poser, J. Kristaponis, N. Raman, *ibid.* **73**, 1447 (1976).
112. L. Cohen-Solal, M. Cohen-Solal, M. J. Glimcher, *ibid.* **76**, 4327 (1979).

113. Supported in part by the National Institutes of Health (grant AM 1567), the New England Peabody Home for Crippled Children, Inc., and The Medical Foundation, Inc. (Nelson E. Weeks Fund).