The Biochemistry of Collagen

ARTHUR VEIS, Ph.D.

Department of Biochemistry,
Northwestern University Medical School,
Chicago, IL 60611

ABSTRACT

Collagen synthesis and subsequent extracellular stabilization and turnover are multi-step processes. The sequence of events in these processes is reviewed and correlated with collagen pathologies.

Introduction

In the past few years, significant progress has been made in uncovering the details of the biosynthesis, stabilization and turnover of collagen. Equally solid progress has been made in understanding the structure and assembly of collagen fibers. This brief review will be devoted to outlining the stages in each of these processes, with emphasis on those aspects to which connective tissue pathologies have been related.

The collagen molecule found in the extracellular space within the collagen fibers is a modified form of a precursor molecule called "procollagen." The procollagen molecule in figure 1 is comprised of three extended polypeptide chains of considerable length, each chain being in the range of 130,000 in molecular weight and having more than 1,000 amino acid residues. The three chains are assembled in a parallel fashion, that is, the amino terminal residues of each chain are in register and at the same end of the procollagen molecule. Similarly, all three carboxyl terminal residues are in register at the other end of the collagen molecule. The major central region of each molecule is a compound rod-like structure in which the three chains are wound helically about the molecular axis. The helical region extends for exactly 1,011 amino acid residues in each of the three peptide chains in the most common type of skin collagen. This helical region is characterized by having the sequence GLY-X-Y repeating in each peptide chain without deviation. Thus, glycine absolutely must appear as every third amino acid residue in each chain. Proline frequently appears in position 2 or 3 (X or Y) in the GLY-X-Y triplet. The amino acid hydroxyproline, which shall be examined in some detail, appears only in position 3 (or Y) in all mammalian collagens. The amino terminal region sequence of each chain does not contain the triplet GLY-X-Y structure and, hence, does not continue or extend the rod-like character. The conformation of this region is not known in detail as yet. In a similar fashion, the carboxyl terminal region also does not contain chain sequences with the GLY-X-Y repeating triplet and it, too, has some conformation other than that of the rod-like structure. Imagine, then, the procollagen molecule as
a stiff, rigid structure with flexible or non-rod-like short end regions.

Hydroxylysine is another amino acid found almost exclusively in collagen. This also occurs in the 3 position of the chain triplets in the helical regions. On the other hand, the non-helical amino terminal region contains several cystine residues which link the three strands of procollagen together by disulfide bonds. Cysteine does not appear in the helical regions. It is uncertain at this time if there are any sulfhydryl-mediated bonds stabilizing the carboxyl-terminal region. The procollagen molecule is also a glycoprotein and contains a few residues of galactose or glucosyl galactose bound directly to certain of the hydroxylysine residues found in the triple-helical region of the molecule.

**Peptide Chain Synthesis and Hydroxylation Reactions**

The messenger for collagen synthesis appears to be monocistronic and the synthesis of each chain occurs on polysomes of just barely the size required for the synthesis of the complete chain of somewhat more than 1,000 amino acid residues. Shortly after synthesis begins, the polynosomes appear to attach to the membrane of the rough endoplasmic reticulum and the growing or nascent peptide chains are extruded through the membrane into the cisternal region of the endoplasmic reticulum (ER) while the polynosomes are bound on the outer surface of this membrane. One of the unique features of the collagen system is that neither hydroxylysine nor hydroxyproline is incorporated directly into the polypeptide chains. Lysine appears at the position which hydroxylysine ultimately occupies and proline appears in those positions which subsequently are transformed to hydroxyproline. The hydroxylation reactions in both cases require molecular oxygen, ferrous ion, α-ketoglutarate and ascorbic acid or an equivalent electron acceptor, as shown schematically in figure 2. Separate enzymes, lysyl hydroxylase and peptidyl proline hydroxylase, are required to catalyze the hydroxylation of
lysine and proline residues, respectively. The proline hydroxylation reaction appears to occur while the nascent chains are still growing and are thus attached to the ER membrane. The situation is not so clear with regard to the location of the lysine hydroxylating system, but it is likely that lysine hydroxylation also occurs in close proximity to chain synthesis.

The hydroxylation reactions occurring subsequent to peptide chain synthesis are not under the same kind of close control as the sequence of assembly of the residues in the peptide chains and, hence, one finds a microheterogeneity in both proline hydroxylation and lysine hydroxylation. In the same preparation of collagen, one can find that not every residue which has the capacity to become hydroxylated does in fact become hydroxylated. The two enzyme systems, the peptidyl proline and lysine hydroxylases, are not linked directly and one can find situations, as reported first by Pinnell et al., in which the hydroxylation of proline appears to occur at a normal level, whereas owing to a deficiency of lysine hydroxylase, hydroxylysine is present to a much lower extent than normal.

**Figure 2. Schematic outline of the hydroxylation process of nascent collagen peptide chains.** Note that the peptidyl prolyl hydroxylase (PPH) is probably a subunit enzyme present in inactive form and requiring activation and/or dissociation before participating in proline hydroxylation.

**Figure 3. Schematic depiction of state of collagen peptide chains at the time of registration.** The chains are in the non-helical form but have been released from the endoplasmic reticulum (ER) membrane. Triple-helix formation to the structure depicted in figure 1 follows if hydroxylation reactions have proceeded to a sufficient extent.

**Assembly of the Three-Chain Structure**

The nascent chains growing into the cisternal space of the ER are separate and in disordered conformations. One primary role of the amino terminal extensions is to interact with counterpart regions on other chains and register or correctly align the three chains necessary for each molecule. This probably occurs after the completed chains are released from the ER membrane and the correct registration is assured by formation of disulfide bonds between cysteine residues of the NH2-terminal propeptides linking the three polypeptide chains covalently into a single three-stranded procollagen molecule (figure 3). Triple-helix formation follows chain association apparently coupled with or closely controlled by the degree or extent of hydroxylation of proline. Berg and Prockop and Rosenbloom and his colleagues have convincingly demonstrated that the collagen denaturation temperature is directly proportional to the degree of hydroxylation. The current hypothesis is that hydroxylation of the released chains continues until some critical degree of hydroxylation is reached, depending upon the ambient temperature. At this point, triple-helix is formed and hydroxylation ceases.

The hydroxylated, triple-helical procollagen is ready for transport at this stage. However, like other proteins destined for transport to the extracellular space, the collagen molecule is glycosylated. Glyco-
sylation takes place on the hydroxyl group of hydroxylysine in one or two steps, mediated by glycosyl transferases, to yield peptide bound hydroxylsyl-galactose or hydroxylsyl-galactosyl-glucose. Several hydroxylysine residues in each chain are glycosylated but again some microheterogeneity is found. The relative contents of monosaccharide and disaccharide are tissue specific. In skin, the disaccharide predominates; in bone, there is a higher content of monosaccharide. The time and intracellular location of the glycosylation reactions is presently unknown.

**Secretion, Exocytosis and Conversion of Procollagen to Collagen**

In addition to its function in polypeptide chain registration, the intact NH$_2$-terminal propeptide region appears to inhibit intracellular aggregation of collagen to the fibrous form.$^8,16$ Upon exocytosis, conversion of procollagen to collagen takes place by the enzymic removal of a major section of the NH$_2$-terminal propeptide. It is our belief$^{17}$ that this conversion takes place in at least two steps and that the intermediate form has an important role in fibrillogenesis. However, others consider the presence of intermediates as artifact and that the conversion to collagen occurs in a single step. At least one enzyme involved in the procollagen to collagen conversion, procollagen peptidase, has been identified in skin and tendon. If there is more than one step in the conversion process, the first degradative cleavage probably takes place at the fibroblast membrane in close conjunction with the secretion of the procollagen molecule.

**Fibril Formation and Stabilization**

Upon the conversion of procollagen to collagen, or to an intermediate form, aggregation appears to occur spontaneously. Native collagen fibers exhibit a periodicity of 690 Å (D) while each molecule is 4.4 D in length. Thin filaments are formed first in which molecules are assembled in strands with a 4 D axial shift, that is, with a 0.4 D overlap (figure 4). Four of these strands aggregate to produce a microfibril with a diameter of $\sim$40 Å and with a shift of 1 D between molecular ends in adjacent strands. The microfibrils finally aggregate to form D periodic fibers. Proteoglycans present in the tissue may influence ultimate fiber diameter. Fibers formed by the side by side aggregation of collagen molecules and microfibrils have little intrinsic stability and low tensile strength.$^8$

High tensile strength is developed by the formation of covalent intermolecular cross-linkages between molecules. The se-

---

*Figure 4. Schematic depiction of stages in the formation of extracellular collagen fibers. Upon exocytosis, procollagen is converted to collagen or an intermediate form. These molecules aggregate to form limiting microfibrils which establish the axial periodicity of the fibrous system. The microfibrils then aggregate to produce the macroscopic collagen fibers.*
sequence of events in cross-link formation is illustrated in figure 5. As shown, the first step in cross-link formation is the production of aldehydes by the oxidative deamination of a lysine or hydroxylysine side chain to produce the "allysine" (α-amino adipic acid semialdehyde) and "hydroxy(allysine)" moieties.
Mammalian collagenases cut the triple-helix at a position about one-fourth the length from the -COOH terminal end (figure 6). The end-effects in the region of the peptide bond cleavage destabilize the triple-helix and denaturation appears to take place at body temperature. The denatured collagen chains are susceptible to other tissue (lysosomal?) enzymes and are degraded to peptides which are subsequently excreted.

**Genetically Distinct Chain Types**

In addition to the microheterogeneity already noted in the relative contents of hydroxylysine, hydroxyproline and glycosylated hydroxylysine residues, all collagens from the same animal are not the same. So far, four distinct types of collagen have been identified. In table I are shown the distribution of these four types of collagen, their location and some of their major properties. The comparisons in terms of high or low content of a particular amino acid are all with reference to the most

<table>
<thead>
<tr>
<th>Type</th>
<th>Designation</th>
<th>Molecular Formula</th>
<th>Tissue Location</th>
<th>Characteristics*</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>[α1(I)]₂α2</td>
<td>Skin, tendon, bone, dentin</td>
<td>High hydroxylysine, high degree of glycosylation, insoluble, highly cross-linked</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>[α1(II)]₃</td>
<td>Cartilage</td>
<td>High hydroxyproline, contains cysteine, highly cross-linked and insoluble, high content of glycine</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>[α1(III)]₂</td>
<td>Skin, arterial wall</td>
<td>High hydroxylysine, high degree of glycosylation, contains cysteine, linked to globular protein</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>[α1(IV)]₃</td>
<td>Basal lamina</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*All comparative statements relative to Type I collagen.

**Turnover and Remodeling: The Collagenases**

In the normal processes of growth and development, there is a continuous remodeling of the skeleton and supportive soft connective tissues. In general, dimensional change cannot be considered merely as the further accretion of collagen molecules depositing on preformed collagen fibers. Instead, existing fibers are degraded and new fibers laid down in the proper position.


TABLE II

Stages in Collagen Synthesis, Fibril Formation and Degradation as Related to Collagen Pathologies

<table>
<thead>
<tr>
<th>Stage</th>
<th>Process</th>
<th>Defect and Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intracellular</td>
<td>Assembly of polypeptide chains</td>
<td>Amino acid substitutions: decreased stability, failure to form triple-helix</td>
</tr>
<tr>
<td></td>
<td>Hydroxylation of proline</td>
<td>Deficient prolyl hydroxylase or absence of cofactor: accumulation of underhydroxylated collagen, decreased structural stability, failure to form triple-helix</td>
</tr>
<tr>
<td></td>
<td>Hydroxylation of lysine</td>
<td>Deficient lysyl hydroxylase or absence of cofactor: low hydroxylysine content, fewer sites for glycosylation and/or subsequent formation of intermolecular cross-linkages. Ehler-Danlos-VII</td>
</tr>
<tr>
<td></td>
<td>Glycosylation of hydroxylysine</td>
<td>Deficient glycosyl transferase activity: possible effect on transport</td>
</tr>
<tr>
<td></td>
<td>Chain association and triple-helix formation</td>
<td>Failure leads to accumulation (and degradation?) within cell</td>
</tr>
<tr>
<td></td>
<td>Secretion of completed collagen to extracellular space</td>
<td>Disruption of microtubular intracellular transport system</td>
</tr>
<tr>
<td>Extracellular</td>
<td>Conversion of procollagen to collagen</td>
<td>Deficiency of procollagen peptidase: Ehlers-Danlos VII - dermatosparaxis, failure to form normal fibers</td>
</tr>
<tr>
<td></td>
<td>Oxidation of lysine to allysine and subsequent cross-linking</td>
<td>Deficiency of lysyl oxidase, deficiency of Cu^{2+}, presence of dietary factors such as $-amino propionitrile: lathyrisn, loss of fiber tensile strength, Ehlers Danlos VI</td>
</tr>
<tr>
<td></td>
<td>Turnover via collagenase degradation</td>
<td>Remodeling inhibited or tissue necrosis</td>
</tr>
</tbody>
</table>

prevalent kind of collagen, the type I collagen of bone, skin and tendon. A particular tissue may contain more than one type of collagen chain, — for example, skin collagen is a mixture of types I and III, but in addition the ratio of contents of types I and III are age and perhaps tissue pathology dependent. Type III collagen is present in larger amounts in very young skin, whereas type I is present in overwhelming amount in mature skin. A particularly interesting feature of the various collagen types is that types II, III and IV all contain collagen molecules with only a single type of polypeptide chain, that is, their formula is [\(\alpha\)]\(_3\), whereas the more common and prevalent collagen of mature animals, skin and bone, the type I collagen, has the formula [\(\alpha_1(1)\)]\(_2\alpha_2\), that is, each molecule has two \(\alpha_1\)-chains and one \(\alpha_2\)-chain. Sequence studies which are now in progress show that there are great homologies between \(\alpha_1\) and \(\alpha_2\)-chains of type I collagen, so even in spite of the differences the chain sequences are not unrelated.

Collagen Pathologies

In table II are recapitulated the stages in collagen synthesis, stabilization and removal. An attempt shall be made to relate these stages and various collagen pathologies. In general, one can consider that the connective tissue pathologies arise from changes such as gene alterations or from outside sources such as toxic or infectious agents. Regardless of the origin of the defect, the effects of changes can be considered in two ways. Those which are directly related to the structural organization of the fibers themselves, for example a change in \(T_M\), owing to a variation in chain composition, and those disorders which are related to the control of one facet or another of the biosynthetic process itself, such as a change in the rate of some step or a blockage of a particular pathway.

At the deoxyribonucleic acid (DNA) transcriptional level, one might also consider diseases in which the ratio of one type of collagen to another in a particular
tissue is changed. This appears to be the case in osteogenesis imperfecta where in the skin of affected individuals the type III collagen predominates over the normal type I collagen level at a given age. How this is involved with the bone malformations in osteogenesis imperfecta is not clear since the bone collagen itself remains as type I. In any event, a great deal of work is being done with many disorders now in the analysis of the type of collagen chains which are present. The clinical entities known under the common heading of the Ehlers-Danlos Syndrome also appear to be directly related to defects in the collagen system, but in these cases most of the studies have indicated that the deficiencies are related to control of the hydroxylation reactions of both proline (helix stability, $T_M$) and lysine or to enzyme defects relating to lysine hydroxylase or lysyl oxidase, — both leading to defects in cross-linking.

At still another level, there has been some work relating to situations in which there is a deficiency of procollagen peptidase leading to a defect at the stage of fibril formation following normal production and exocytosis of the procollagen. Some attempt has been made to implicate changes in the basement membrane collagen as one of the causal effects of diabetes melitis. However, there is no evidence that there is any defect directly in the basement membrane collagen. It appears simply that more collagen is produced and the basement membrane is thickened.

Summary

As indicated, the collagen system is quite dynamic and subject to modulation at many stages. There appear to be direct connections between variations in collagen synthesis and postribosomal modification which are related to disease states. For the analyst, it seems that attention will have to be given in the future to assays of clinical biopsies for enzyme levels relating to proline and lysine hydroxylases, to the levels of lysyl oxidase in the extracellular tissue, and to the levels of mammalian collagenases. Some attention will also have to be given to developing standardized procedures for the identification of the ratios of chain types in various collagen tissues, and statistical studies of the distribution of such chain types in normal and pathological states.

References


