Biosynthesis and Structure of Collagen*

Thayne R. Dutson
Texas A&M University

Introduction

Unlike many other biological macromolecules which have specific genetic or metabolic activities, collagen is a protein which to date has been primarily shown to be of a structural nature. Collagen is ubiquitous in the animal kingdom and contributes 20 to 30% of the total body proteins in vertebrates. Collagen is present throughout the body and is found in such diverse tissues as skin, tendon, bone, cornea, basement membrane and muscle. Its prime importance lies in its mechanical properties (tensile strength) and its ability to support and innerconnect body tissues, making possible organization of the whole body or organism. Collagen has also been shown to be important in connective tissue diseases (Miller and Matukas, 1974; Bailey, Robbins and Balian, 1974), muscle strength (Ham, 1969; Cassens, 1971) and in meat tenderness (Croll et al., 1963; 1964; Herring et al., 1965). Thus, in order to understand more completely the mechanisms by which this protein is synthesized and organized into the fibers that are so important in the structure and function of connective tissues considerable emphasis has been placed on collagen research in recent years. In most tissues, these collagen fibers occur in a variety of forms such as fine filaments, thick rope-like fibers or in laminated sheets of fibers, as depicted in figure 1.

Collagen Structure

A considerable amount of information has been known about the basic structure of collagen for many years (Traub and Piez, 1971; Gallop et al., 1972) but new information is being continually added (Bailey and Robbins, 1976). Tropocollagen, the basic molecular unit of collagen, is a triple stranded, coiled coil rod-like structure of about 300 Å in diameter. The collagen molecule (tropocollagen) contains three polypeptide chains known as α chains which are coiled together in a superhelix to form the collagen molecule (Fig. 2). Each of the α chains themselves have a unique amino acid sequence in that every third residue is glycine and they have the repeating tripeptide gly-x-y, with the x and y frequently being proline and hydroxyproline, respectively (Traub and Piez, 1971). The complete amino acid sequence of 1052 residues has been reported for the α1 (I) chain with half being from rat skin and half from calf skin collagen (Fietzek et al., 1972; Balian et al., 1972; Hulmes et al., 1973; Chien, 1975). It is now clear from this data that the alpha chains are continuous polypeptide chains in which glycine does

PRIMARY SEQUENCE

GLY-PRO-Y-GLY-X-Y-GLY-X-HYP-GLY-

TRIPLE HELIX

TWO DIMENSIONAL QUARTER STAGGER

COLLAGEN FIBER

THREE DIMENSIONAL PENTAFIBRIL QUARTER STAGGER

Fig 2
indeed occur at every third residue in the sequence gly-x-y. The sequences gly-pro-y and gly-x-hyp occur in almost equal proportions except for the short regions at the N and C terminal ends of the α chain, with hydroxyproline being almost exclusively confined to the y position (Bailey and Robins, 1976).

The high content of glycine in the α chains allows relatively unrestricted rotation which is balanced by the stereochemical properties of the pyrrolidine rings of proline and hydroxyproline that direct the chains into a polyproline II helix. Another unique property of glycine is that it has no side chain and the association of three α chains in the collagen superhelix places the α carbon atom of every third residue in the interior of the rod-like molecule, allowing the three polyproline II helices to fit nicely into a triple-stranded superhelix arrangement (Traub and Piez, 1971). The glycine residues on the inside of the inside of the molecules also form hydrogen bonds to the amino group of the peptide bond on adjacent chains. The side chains of other residues being on the outside of the tropocollagen molecule allowing them to be involved in molecular interactions controlling the organization of tropocollagen into the fibers (Bailey and Robins, 1976).

Water is also an important component of native collagen with about 0.5 g of water needed per g of tropocollagen to maintain its native conformation (Chien, 1975; Dehl, 1970). Water is implicated in inter-chain hydrogen bonds (Ramachandran and Chandrasekharan, 1968) of collagen which becomes insoluble upon complete dehydration (Yannas and Tobolsky, 1967). We have also noted the decreased solubility of freeze-dried intramuscular collagen in our laboratory, with decreases in solubility being greater if freeze-dried collagen is stored at room temperature.

Each tropocollagen molecules are formed into fibrils which probably consist of five tropocollagens arranged in a quarter stagger array known as pentafibrils (Smith, 1968) which are shown in Figure 2. The hole and overlap regions caused by the quarter stagger array of the tropocollagen molecules account for the 650 Å periodicity of collagen fibrils found with the electron microscope (Figure 2). This specific organization of collagen molecules results in fibers of very little strength which are later stahilized by the formation of inter- and intra-molecular cross links resulting in a continuous structure with very high dimensional stability. The topic of cross-linking will be covered in the following paper by Dr. McLain.

Collagen Types

At the present time, four distinct types of collagen have been identified in various tissues, Type I, Type II, Type III and Type IV as presented in Table 1 (Miller and Matukas, 1974; Bailey and Robins, 1976; Martin et al., 1975). All four collagen types are composed of three α chains which vary only slightly in amino acid composition (Miller and Matukas, 1974). Type I, II and III collagens are aggregated
Figure 3
Table 1. Types of Collagen

<table>
<thead>
<tr>
<th>Type</th>
<th>Tissue location</th>
<th>Molecular composition</th>
<th>Distinctive features</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Skin, tendon, bone, muscle</td>
<td>$\alpha_1(I)\alpha_2$</td>
<td>$\alpha_1$ chains, 6-8 Hyl/chain low carbohydrate</td>
</tr>
<tr>
<td>II</td>
<td>Cartilage, Intervertebral discs</td>
<td>$\alpha_1(II)_3$</td>
<td>$\alpha_1$ chains only, 20-25 Hyl/chain, 10% carbohydrate</td>
</tr>
<tr>
<td>III</td>
<td>Fetal skin, cardiovascular system, Synovial membrane, cardiac and skeletal muscle</td>
<td>$\alpha_1(III)_3$</td>
<td>$\alpha_1$ chains only, 6-8 Hyl/chain, high Hyp and Gly, low carbohydrate, Cys</td>
</tr>
<tr>
<td>IV</td>
<td>Basement membrane</td>
<td>$\alpha_1(IV)_3$</td>
<td>$\alpha_1$ chains only, 60-70 Hyl/chain, high 3-Hyp, low Ala, 15% carbohydrate</td>
</tr>
</tbody>
</table>

In fibrous form while Type IV maintains an amorphous structure in the basement membrane and is apparently unable to form the fibrous structure associated with other collagen types (Bailey and Robins, 1976). These four unique collagens are composed of five different $\alpha$ chains $\alpha_1 (I)$, $\alpha_1 (II)$, $\alpha_1 (III)$, $\alpha_1 (IV)$, and $\alpha_2$, which have been determined to be different gene products, however, they probably undergo the same post-translational modifications (Martin et al., 1975).

Type I

Type I collagen is the most completely characterized collagen with the complete amino acid sequence being known (Chien, 1975). It is found in skin, bone, tendon, and muscle. Type I collagen contains two identical $\alpha_1$ chains and a different third chain, $2$ (Piez, 1967) making the molecular configuration $[\alpha_1 (I)]_2 \alpha_2$. The $\alpha_1 (I)$ chains are distinct from $\alpha_1$ chains of other collagen types in amino acid composition and sequence (Linsenmayer, 1973a, 1973b; Miller and Matukas, 1969; McClain, 1974; Bailey and Sims, 1976) and Type I collagen is the only collagen known to contain $\alpha_2$ chains (Martin et al., 1975). Type I collagen contains about 6-8 Hydroxylysine residues and has a low carbohydrate content, less than 20% of the Hydroxylysine is glycosylated (Spiro, 1969; Bailey and Robins, 1976).
Type II

Miller and Matukas (1969) were the first to identify the presence of a different collagen type in cartilage which was found to contain three identical α chains and have a molecular configuration of \([\alpha_1 (II)]_3\). The three α chains of Type II collagen differ from α1 (I) and α2 chains with respect to amino acid composition and sequence (Miller and Matukas, 1969, 1974). Type II collagen has been identified in a variety of cartilaginous structures of the chick (Miller, 1971a, 1971b, 1972, 1973; Toole et al., 1972), the bovine (Miller and Lunde, 1973; Strawich and Nimni, 1971), and the human (Miller et al., 1971; Miller and Lunde, 1973), with Type II collagen having 20-25 Hydroxylysine residues per 1000 residues and approximately 50% of the Hydroxylysine being glycosylated. Although Type II collagen was first shown to be present in cartilage, it has recently been implicated in other tissues such as intervertebral discs (Bailey and Robins, 1976).

Type III

Type III collagen, and similar collagens, have been identified from aorta, muscle tissue, synovial fluid, and granulation tissue (Wiedemann et al., 1975; Chung and Miller, 1974; Epstein, 1974; Chung et al., 1974; McClain, 1974; Bailey et al., 1975). These studies have indicated that Type III collagen is composed of three identical α chains which differ in amino acid composition from α1 (I), α2, and α1 (II) chains and has the molecular configuration \([\alpha_1 (III)]_3\). Type III collagen contains 6-8 Hydroxylysine residues per 1000 residues and has approximately 15-20% of the Hydroxylysine glycosylated. Type III collagen contains higher amounts of Hydroxyproline and contains cystine, which is apparently involved in disulfide bonds at the C terminal end and contributes to the low solubility of this collagen.

Type IV

Type IV collagen, which has not been completely characterized (Bailey and Robins, 1976), was identified by Kefalides (1971, 1973). Type IV collagen is also made up of three identical α chains with the molecular configuration \([\alpha_1 (IV)]_3\). Amino acid composition of Type IV collagens (Daniels and Chu, 1975; Kefalides, 1973) indicate that the α1 (IV) chain has a high content of Hydroxylysine (60-70 residues per 1000 residues), higher amounts of both 3 and 4 Hydroxyproline and some cystine. A high amount of carbohydrate (80% of the Hydroxylysine is glycosylated) is also associated with Type IV collagen (Bailey and Robins, 1976).

Collagen Biosynthesis

The sequence of events in collagen biosynthesis includes the production of the nascent polypeptide chains (Transcription and translation), alterations on these chains during and after translation, such
as glycosylation and hydroxylation, formation of the chains in a precursor molecule, secretion of the precursor and its conversion to collagen, and the aggregation of the collagen molecules into fibers (figure 3).

Although none of the messenger RNA's (m-RNA) that are specific for an individual collagen type have been isolated and purified, the determination that different collagen types have different amino acid contents and sequences in the chains indicate that there are separate m-RNA's for each of the five \( \alpha \) chains, \( \alpha_1 \), \( \alpha_2 \), \( \alpha_1 \) (II), \( \alpha_1 \) (III) and \( \alpha_1 \) (IV) (Miller et al., 1976; Epstein, 1974; Chien, 1975; Martin et al., 1975; Kefalides, 1973). The m-RNA's for synthesis of these chains are probably monolystronic based on size of the polysomes (Lazarides and Lukens, 1971; Kerwar et al., 1972; Diegelman et al., Benveniste, 1973). Data presented by Vuust and Piez (1970) have indicated that the m-RNA's for \( \alpha_1 \) (I) and \( \alpha_2 \) are transmitted simultaneously and that the chains are assembled by addition of individual amino acids and not as shorter polypeptide chains. The rate of translation of \( \alpha_1 \) (I) and \( \alpha_2 \) chains has been determined to be about 200 residues per minute (Vuust and Piez, 1972) and the synthesis of nacent collagen chains has been shown to occur on membrane bound ribosomes (Goldberg and Green, 1967; Diegelman et al., 1973).

Hydroxylation and Glycosylation

An unusual series of reactions take place in collagen biosynthesis which alters the structure of two of the amino acids after their incorporation into the nacent chain. Lysine and proline are hydroxylated by the enzymes proline hydroxylase and lysine hydroxylase, respectively, and hydroxylation has been shown to take place on the nacent chains prior to their release from the membrane bound ribosomes (Miller and Udenfriend, 1970; Lazarides et al., 1971). These two enzymes both require molecular oxygen, ferrous iron, \( \alpha \)-ketoglutarate, and ascorbic acid (Grant and Prockop, 1972; Miller, 1971; Miller and Matukas, 1974). Molecular oxygen has been shown to be a source of oxygen for the hydroxyl group (Prockop et al., 1963). Alpha-\( \alpha \)-ketoglutarate is required and undergoes decarboxylation upon hydroxylation of the \( \alpha \) chains (Rhoads and Underfriend, 1968). Hydroxylase activity is activated by ferous iron which may serve as a binding site for molecular oxygen (Miller and Matukas, 1974). Ascorbic acid probably acts as a reducing agent for the hydroxylation reaction and may also be involved in converting an inactive form of the enzyme to an active state (Hutton et al., 1967; Stassen et al., 1973).

Hydroxyproline has been shown to add stability to the collagen helix and that at physiological temperatures, helix formation of the procollagen molecule is dependent upon hydroxylation of proline (Sakakibara et al., 1973; Murphy and Rosenbloom, 1973). Under hydroxylated procollogen has been shown to accumulate in the cell suggesting that hydroxylation or that the complete helical procollagen is necessary for secretion (Miller and Matukas, 1974; Martin et al., 1975).
The carbohydrate groups of collagen (glucosylgalactose and galactose) are covalently bonded by O-glucosidic linkages to the carbon of hydroxylysine (Spiro, 1969). Galactose is apparently bound to hydroxylysine by a galactosyl transferase (Spiro and Spiro, 1971a) and glucose is then bound to the galactose by a glucosyl transferase (Spiro and Spiro, 1971b).

Helix Formation and Secretion

Although the in vitro aggregation of the polyproline helix of individual chains is slow, the synthesis of the triple helical pro-collagen molecule is very rapid (Beier and Engel, 1968; Vuust and Piez, 1972). Assembly during synthesis is probably achieved by a registering of the ends of the nonpolyproline helical region of the α chains, thus locating the correct alignment for the formation of the triplehelix (Bailey and Robins, 1976). The extension peptides at the amino terminus (molecular weight 15000 Daltons) and at the carboxy terminus (molecular weight 35000 Daltons) both contain cystine residues, but disulfide bonds between α chains are found only at the carboxy terminal ends (Fessler et al., 1975; Bailey and Robins, 1976). This indicates that disulfide linkages may be involved in a rapid aggregation of the C terminal extension peptides on completion of synthesis. Upon completion of helix formation, the procollagen molecule passes through the golgi to the cell membrane by being packaged in vesicles for transport through the cells. This transport probably involves the microtubular system (Diegelmann and Peterkovsky, 1972; Warlich and Bornstein, 1972; Bailey and Robins, 1976).

Due to the fact that procollagen with its intact extension peptides will not aggregate to form collagen fibers (Martin et al., 1975; Bailey and Robins, 1976) these extension peptides must be removed before fiber formation can take place outside the cell. A number of proteases have the ability to remove the non helical peptides (Martin et al., 1975), however, part of these proteases are not present or do not function under physiological conditions. An enzyme which is capable of removing extension peptides has been isolated from calf tendon and other tissues (Iapear et al., 1971; Bornstein et al., 1972). This enzyme, called procollagen peptidase, has been shown to be dependent upon calcium for activity and to remove extension peptides indicating that this enzyme is an endopeptidase (Kohn et al., 1974). After cleavage of extension peptides, the tropocollagen molecules align themselves into fibers, according to the scheme described previously under structure.

Collagen Turnover

Although the rate of synthesis of any protein has been shown to be affected by various factors, the net accumulation of any protein is dependent upon both the rate of synthesis and the rate of degradation (Dayton et al., 1975; Bergen, 1975). Therefore, a discussion of collagen synthesis should include a discussion of collagen turnover.
Early experiments on collagen turnover have indicated that collagen turnover is very slow (Gerber et al., 1960; Popenoe and Van Slyke, 1962). Although some studies have indicated high collagen turnover rates in some tissues and certain pathological conditions, compared to other proteins, collagen is still considered metabolically inert (Bailey and Robins, 1976). In order to determine if collagen in muscles of mature animals does actually turnover, an experiment has been conducted in our laboratory which involved the feeding of β-aminoproprionitrile (BAPN) to mature ewes. BAPN, a compound which has been shown to inhibit the enzyme lysyloxidase and thus prevent the crosslinking of newly synthesized collagen, should be able to increase the solubility of collagen from mature animals if collagen is being turned over, due to the fact that the newly synthesized collagen will not be crosslinked and thus will have a higher solubility. Some preliminary results of this experiment are presented in table 2. The percent soluble collagen was increased over the controls with the percent soluble collagen being greater with greater amounts of BAPN in the diet and longer time on the BAPN treatment. This data indicates that intramuscular collagen is being turned over even in mature animals and that the state of the collagen can be altered in mature animals.

<table>
<thead>
<tr>
<th>Days on BAPN</th>
<th>20</th>
<th>20</th>
<th>40</th>
<th>40</th>
<th>60</th>
<th>60</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAPN dosage (g/day)</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Soluble collagen (%)</td>
<td>4.3</td>
<td>5.4</td>
<td>5.8</td>
<td>7.2</td>
<td>5.6</td>
<td>7.7</td>
<td>4.1</td>
</tr>
</tbody>
</table>

An enzyme, which has been shown to attack native collagen under physiological conditions, was originally isolated from tadpole tissue (Gross and Lapiere, 1962) and has recently been found in numerous other tissues (Harris and Krane, 1974; Woolley et al., 1975; McCroskery et al., 1975; Wooley et al., 1976). This enzyme has an optimal activity around pH 7.5 to 8.5, is activated by calcium ions, cleaves native collagen peptides at one site and cleaves [αI (I)]2, α2 [αI (III)]3 collagens five times more rapidly than [αI (II)]3. Although the physiological role of this enzyme has not been determined, it could be a factor in the natural turnover of collagen in various tissues.
Literature Cited


Dennis Campion: Thank you, Dr. Dutson. Consistent with the way the session is outlined in our program, I think we'll refrain from any questions at this point and hold them until the end of the program.

Dr. Phil McClain, our next speaker, obtained his Bachelor's Degree from the University of Missouri, went to Louisiana State University for his Master's Degree, and then completed a Doctorate of Philosophy Degree at Michigan State University. Dr. McClain is currently a research chemist at the Protein Nutrition Laboratory, Nutrition Institute, Agriculture Research Service, U.S. Department of Agriculture. At this time, I'd like to present Dr. McClain who will be speaking to us on the topic of chemistry of collagen crosslinking.