Recent Advances in Collagen Biochemistry

Elton D. Aberle*
Edward W. Mills

Collagen is the most abundant and versatile of animal proteins. A notable feature is its astonishing plasticity and range of function. During growth and development of animals, collagen is molded into materials with a wide range of properties and shapes to fit particular biological functions. Collagen fibrils form the structural framework of tendons, bones, cartilage, basement membranes, blood vessel walls and membranous sheets and fibrous networks that support cells and organs. Collagen is involved in several diseases, some of which will probably one day afflict all of us, including arthritis, rheumatism, osteoporosis, cataracts and atherosclerosis. The biology of collagen is intimately related to understanding of wound healing and tissue repair. In the area of business, several large industries are built around the use of collagen for food and non-food purposes — glue, leather, cosmetics and gelatin. Finally, we as meat scientists concern ourselves with the properties of muscle collagen as it affects meat texture or of muscle and non-muscle collagen as an ingredient of processed meat products.

Dutson (1976) and McClain (1976) presented excellent reviews of the biosynthesis, structure and crosslinking of collagen at the 29th Reciprocal Meat Conference. The basic details described then are still valid and the essential points are as follows: The feature common to all collagen molecules is the triple helix, i.e., a coiled coil of three polypeptide units or α-chains. Each α-chain twists in a left-handed helix with three residues per turn and the three α-chains are wound into a right-handed super helix to form a molecule about 1.4 nm in diameter and 300 nm long. At least seven genetically different α-chains have been identified in higher animals and they comprise five or more different molecules (Eyre, 1980). Nomenclature for the subunit composition of collagen molecules is well established. For example, the interstitial collagens are designated as follows: $[\alpha 1(I)]_2\alpha 2$ is type I, $[\alpha 1(II)]_3$ is type II and $[\alpha 1(III)]_3$is type III.

Each α-chain contains about 1050 amino acid residues. Glycine is the only amino acid small enough to cluster down the center core of the molecule, thus it is located at every third position of α-chains in the helical portion of the molecule. For types I, II and III collagens, the general formula for α-chains in the helical part of the molecule is $(Gly\text{-}X\text{-}Y)_{340 \pm 2}$. Proline and hydroxyproline occupy the X and Y positions about one-third of the time. 4-Hydroxyproline is located only at the Y position (Ramachandran and Reddi, 1976). At both the C- and N-terminus of the molecule, there are segments (telopeptides) in which glycine does not occupy every third position and the structure is not trihelical. The telopeptides are the primary sites for crosslinking and for antigenicity of collagen molecules.

Procollagen is the biosynthetic precursor of collagen molecules eventually found packed in fibrils outside the cell (Prockop et al. 1979a, b). Procollagen has extension peptides at both the N- and C- terminus with the carboxyl extension larger than the amino extension. In type I procollagen, the N-terminus extension has a molecular weight of about 15000 and the C-terminus extension has a molecular weight of about 35000. The extension pieces (propeptides) may contain cysteine residues, some of which form disulfide bonds to crosslink the α-chains in the procollagen molecules (Bruckner et al., 1978). Propeptides are cleaved by specific procollagen peptidases outside the fibroblast, although it is believed that the N-terminus extension is cleaved first (Morris et al., 1975). The propeptides are believed to prevent fibril formation until the procollagen molecules reach the desired extracellular site. Other possible functions are to direct assembly of the correct combination of pro α-chains into molecules and to regulate triple helix formation (Eyre, 1980). Feedback inhibition of collagen biosynthesis by the propeptides has been demonstrated (Wiestner et al., 1979).

Procollagen undergoes substantial posttranslational modification intracellularly before it is secreted from the cell. This includes hydroxylation of proline and lysine residues and glycosylation of some hydroxylysines with galactosyl or glucosylgalactosyl side chains. 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 (Prockop et al., 1979a, b). Interstitial collagens also contain one or two 3-hydroxyprolines per α-chain. These residues are synthesized by a separate enzyme, prolyl 3-hydroxylase, and then only in the X position of Gly-X-Y when Y is already occupied by a 4-hydroxyproline (Tryggvason et al., 1977). The hydroxylases and glycosyl transferases require nonhelical chains as substrates; both hydroxylation and glycosylation stop when the triple helix is formed (Prockop et al., 1979a, b). 4-Hydroxyproline stabilizes the triple helix through hydrogen bonding (discussed later), hydroxylysine is essential for crosslinking, and the glycosyl groups may contribute to water binding by collagen fibrils.

Stability of Collagen Molecules

As a result of research in the last decade, we now

*E. D. Aberle, Department of Animal Sciences, Purdue University, West Lafayette, IN 47907

Reciprocal Meat Conference Proceedings, Volume 36, 1983.

understand more fully the factors responsible for stabilization of the three chain coiled coil structure of collagen molecules. The model that has been most favored during the last 30 years is the polyproline II helix, in which stability is achieved through interchain hydrogen bonds between carbonyl and amide groups of peptide bonds located on the inside of the triple helix and through the stereochemical properties of the pyrrolidine rings of proline and hydroxyproline residues which restrict rotation of the structure. Two models for hydrogen bond formation were first proposed. The one-bond model placed a hydrogen bond between the amide group of a Gly residue located in the 1st position of a triplet in one α-chain and the carbonyl of the residue located in the 2nd (X) position of a triplet in an adjoining chain (Rich and Crick, 1955). Thus, there is one hydrogen bond per triplet. In the two-bond model, a second hydrogen bond is located between the carbonyl of Gly and the amide of the residue in the 2nd (X) position of a triplet if the position is not occupied by an amino acid residue (Ramachandran and Kartha, 1955). On the average, there would be fewer than two hydrogen bonds per triplet with this model.

Results from studies of the thermodynamic properties of collagen have forced readjustments of concepts of collagen molecular stability. As temperature is increased, collagen's regular structure breaks (denatures) and the chains separate and fold into random structures without any residual native structure. The uniqueness of collagen in contrast to other proteins is that this change, referrred to as melting, occurs over a very narrow temperature range and proceeds with an extremely intense heat absorption of approximately 6 kJ/mol of amino acid residue or 18 to 20 kJ/amino acid triplet (Privalov, 1982). This value, which is the enthalpy of melting of molecularly dispersed collagen, is much larger than values for denaturation of any other protein. Further, the enthalpy of melting is much larger than can be explained by disruption of two hydrogen bonds between chains per amino acid triplet or by the rigidity conferred on the helix by pyrrolidine residues.

Thermal stability of molecular collagen, as indicated by melting temperature (T_m), correlates closely with the hydroxyproline content of vertebrate collagens but there is almost no correlation between proline content and melting temperature (Burjanadze, 1979). The hydroxyl group of 4-hydroxyproline located in the Y position of the triplet evidently stabilizes the helix through formation of additional hydrogen bonds; 3-hydroxyproline in either the X or Y positions or 4-hydroxyproline in the X position does not lend additional stability to the helix (Privalov, 1982).

Water and Collagen Stability

Participation of water in maintaining the collagen helix is well established (Privalov, 1982) for upon removal of water, collagen loses its regular structure. About .5 g of water is bound (and unfreezable) per gram of collagen (Dehl, 1970; Fung et al., 1974), which is equivalent to 7 to 8 water molecules per amino acid triplet. Upon thermal denaturation, the water sorption capacity of collagen is lost, indicative of the fact that the triple helical structure is required for water binding. Several models of collagen structure have been proposed to allow for water binding. Ramachandran and Ramakrishnan (1976) proposed a model in which two or three water molecules per triplet could be included as water bridges between carbonyl and amide groups of the peptide chains.

Lim (1981) proposed a rather unique model for water participation in the stabilization of collagen structure. According to this model (Figure 1) polypeptide chains do not interact

Figure 1

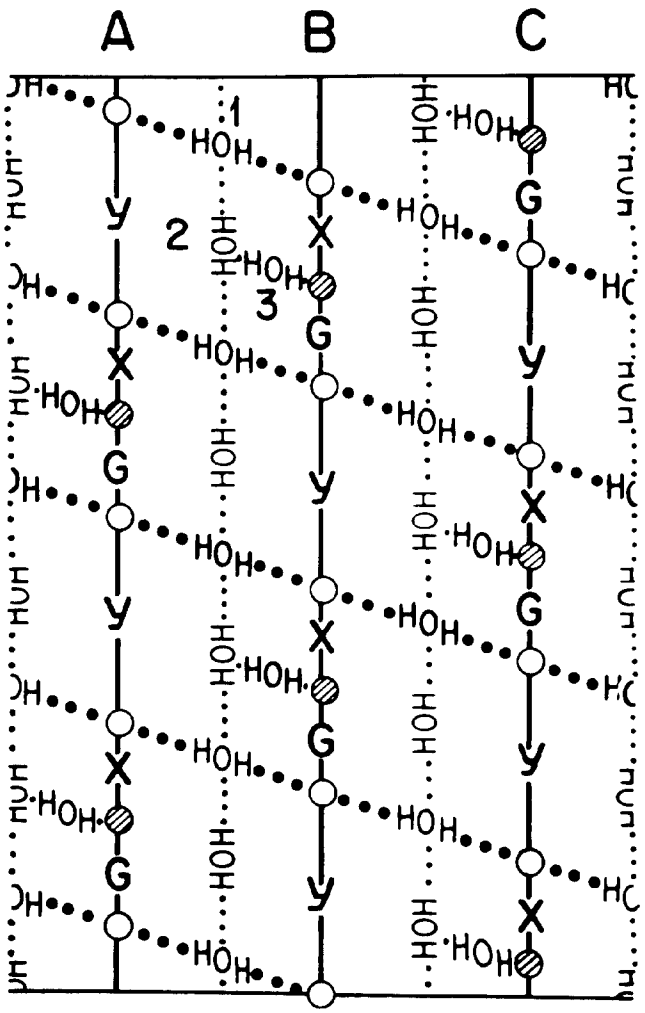

Figure 1. Diagram to illustrate formation of the water-carbonyl helix from three-helical α-chains — A, B and C. In this symmetrical triple helix, α-chains B and C are shifted upwards relative to α-chain A by, respectively, 1/3rd and 2/3rds of the height of a triplet (Gly-X-Y). Open circles denote carbonyl oxygens in the X- and Y-positions of a triplet. Shaded circles denote carbonyl oxygens of glycine residues. 1, 2 and 3 are helical, interturn and glycine water molecules, respectively. Dots represent hydrogen bonds. Large dots represent hydrogen bonds that form the water-carbonyl helix. Helical α-chains A, B and C are related to one another by a 3-fold screw axis with a right-handed twist. Reproduced with permission, from Lim, 1981. FEBS Letters 132:1.

directly with each other by hydrogen bonds. Instead, the triple helix structure is stabilized by water molecules that form interchain bridges through hydrogen bonds with carbonyl groups. The three peptide chains are joined by a left-handed water-carbonyl helix which is formed by water molecules (termed helical water molecules) and the X- and Y- carbonyl oxygens. Adjacent turns of the water-carbonyl helix are bound by water bridges; these water molecules are termed interturn water molecules. The carbonyl oxygen of each glycine residue is able to form a water bridge with every second interturn water molecule. In the water-carbonyl helix, the oxygen atom of each helical water molecule is at the center of a tetrahedron formed by two carbonyl oxygens and two oxygen atoms of interturn water molecules. Such an arrangement allows each helical water molecule to form four optimal hydrogen bonds.

In the model of Lim (1981), all X- and Y- amino acid side chains are located on the surface of the triple helix. Hydrophilic side chains can form hydrogen bonds and water bridges with glycine and interturn water molecules. For instance, when 4-hydroxyproline is located in the Y position, the hydroxyl group can form a hydrogen bond with the interturn water molecule and a water bridge with the glycine water molecule, which would account for greater thermal stability as hydroxyproline content increases. According to this model, the collagen structure is supported by at least five water molecules per triplet incorporated into the structure and this amount increases significantly when hydroxyproline is in the triplet.

Variations in water content of collagen cause marked changes in T_m and enthalpy of melting (Luescher et al., 1974). At a water content of less than about 3 molecules/triplet, the enthalpy of melting decreases rapidly while T_m increases (Figure 2). This amount of water represents that

Figure 2

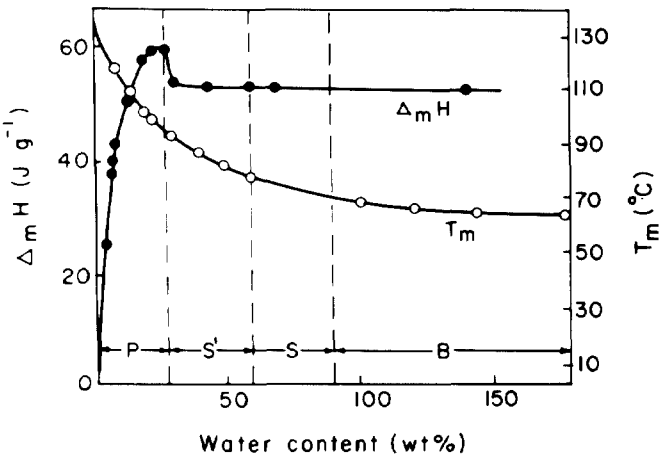

Figure 2. Enthalpies (•) and temperature of melting (0) of calf skin collagen as a function of water content. P, primary hydration water; S^1 and S, secondary hydration water; B, bulk water. Reproduced, with permission, from Luescher et al. 1974. Biopolymers 13:2489.

strongly bound to collagen and is termed primary hydration water. Additional water, termed secondary hydration water, must be bound to collagen before T_m values normal for mammalian collagen are observed.

Melting temperatures of molecularly dispersed collagens are quite close to the upper limit of physiological temperature of the donor species. For mammals, this limit is about 38° to 40°C. Regardless of the species, aggregation of molecular collagen into fibrils is accompanied by about a 25°C increase in the transition temperature (Rigby and Prossner, 1975; Rigby and Robinson, 1975). Thus intermolecular bonds formed during fibril formation reinforce collagen molecular structure. Collagen molecules in solution self-assemble into native-like fibrils through hydrogen bonds, dipole or ion-pair interactions and intermolecular water bridges. This tendency results from the fact that the pattern of charge distribution (negative, positive and uncharged amino acid side chains) along the molecule repeats every 67 nm, (Hulmes et al., 1973). When molecules are displaced axially by 67 nm, strong intermolecular interactions occur with the result that T_m increases. Intermolecular lysine- or hydroxylysine-derived crosslinks probably have lesser effects on T_m. For example, T_m of bovine intramuscular collagen increases by only 3° to 4° C during the period from 4 mo to 6 yr of age (Judge and Aberle, 1982) but the number and stability of lysine-derived crosslinks increase markedly during this time (Shimokomaki et al., 1972).

Postmortem Changes in T_m

Several researchers have reported that melting temperature of intramuscular collagen decreases during postmortem storage (McClain, 1970; Judge and Aberle, 1982). The difference in T_m between samples obtained immediately after death and at 7 d postmortem is approximately 8°C, which is considerably greater than changes in T_m associated with physiological maturation. Recent data obtained in our laboratory (E. W. Mills, unpublished data) indicate that more than half of the change in T_m occurs during the first 12 h postmortem and that solubility of the collagen after a standard heat treatment increases as T_m decreases. Postmortem changes in collagen T_m and solubility may be attributed in part to diminished crosslinking during postmortem storage, as suggested by other researchers (Kruggel and Field, 1971; Pfeiffer et al., 1972; Judge and Aberle, 1982). But it is quite possible that the postmortem environment alters collagen stability by other means. Collagen T_m is affected by the presence of several ions. For example, .05M potassium phosphate buffer decreases T_m of intramuscular collagen 2° to 3° C from that observed in distilled water (Judge and Aberle, 1982). Low concentrations of calcium ions also decrease collagen T_m (von Hippel and Wong, 1963). Thus, the redistribution of ions that occurs during development of rigor mortis may also affect collagen stability. E. W. Mills (unpublished data) found that electrical stimulation of the carcass accelerated the postmortem drop in collagen T_m and increase in solubility, but had little effect on the total change in these variables. Changes in pH in the range of 7.1 to 5.0 have little effect on T_m (Dick and Nordwig, 1966; Judge and Aberle, 1982). However, Wu (1978) observed that muscle incubated at pH 5 yielded more neutral salt-soluble collagen

than that incubated at pH7 and theorized that the difference may have been caused by release of more collagen monomers or an increased susceptibility to protease degradation.

Molecular Packing in Collagen Fibrils

Collagen fibrils viewed by electron microscopy have a 67 nm periodicity along their length. This periodicity results from the fact that the 300 nm long molecules are staggered axially by 67 nm with respect to each other. Because 300 nm = 4.47 x 67 nm, there are gaps of .53 x 67 nm between molecular ends. As stated earlier, the pattern of charged and uncharged amino acids on the surface of collagen molecules repeats itself every 67 nm, which explains the ability of molecules in solution to self-assemble into fibrils with the correct periodicity.

Although the two-dimensional staggered arrangement is well established, much less has been known of the three-dimensional arrangement of molecules in fibrils. One widely accepted three-dimensional model is the quarter staggered pentafibril suggested by Smith (1968). Miller (1982) has considered evidence from electron microscopy, optical diffraction and x-ray diffraction of collagen fibrils, together with knowledge of the amino acid sequence and location of intermolecular lysine- and hydroxylysine-derived crosslinks, to propose a slightly different model for molecular packing in type I collagen. Each collagen molecule is regarded as composed of five segments (Figure 3) labeled 1 through 5 beginning at the C-terminus. Each segment is 67 nm in length except for segment 5, which is .47 x 67 nm long. Chemical evidence places the lysine- or hydroxy-lysine derived intermolecular crosslinks between amino acid residue 9 (in the N-terminal telopeptide of segment 5) and residue 946 (in the helical region of segment 1) of a neighboring molecule. A second crosslink is located between residue 103 (in helical region of segment 5) and residue 1047 (in the C-terminal telopeptide of segment 1) of a neighboring molecule. Thus, proven crosslinks are located between each telopeptide of one molecule and another molecule staggered by 4 x 67 nm with respect to the first. Examination of the amino acid sequence shows that residue 580 (Lys) in helical segment 3 and residue 1044 (OH-lys) in the C-terminal telopeptide of segment 1 are approximately 2 x 67 nm apart and might be able to form a crosslink (Miller, 1982). But firm evidence for this crosslink does not exist.

Views of the three-dimensional molecular packing in collagen fibrils proposed by Miller (1982) are shown in Figure 4. Crosslinks between segments 1 and 5 of the collagen molecules form a sheet of molecules staggered by 4 x 67 nm. A crosslinked polymer of the form $(1\text{-}5\text{-}1\text{-}5)_n$ can be isolated from cyanogen bromide digests of collagen (Light and Bailey, 1980a, b). The speculative crosslink between residues 580 and 1044 could link the 4 x 67 nm periodic sheets to stabilize the whole structure.

Figure 3

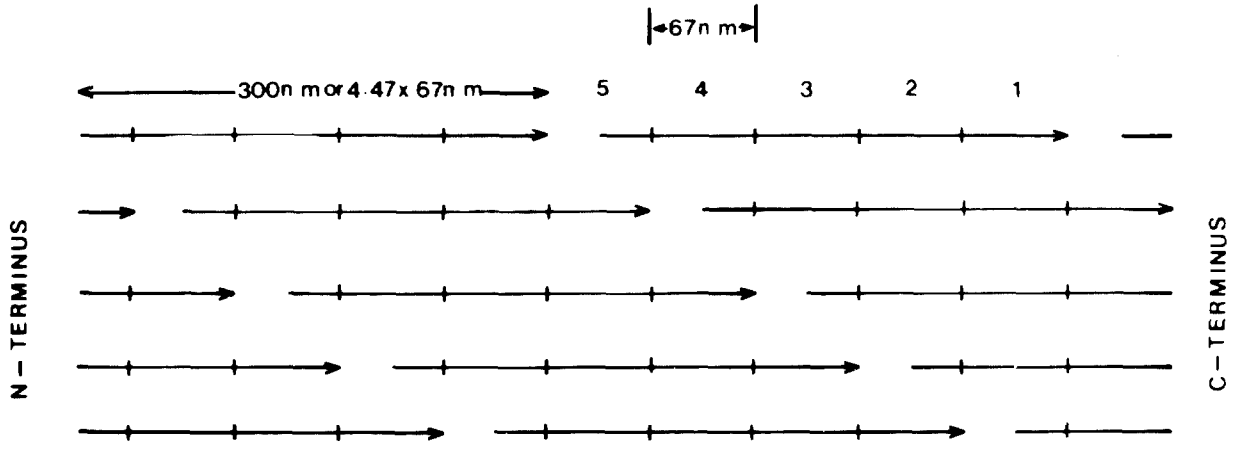

Figure 3. Two-dimensional model of the molecular packing in collagen fibrils. The 300 nm long molecules are staggered axially by 67 nm with respect to each other. Because 300 nm = 4.47 x 67 nm, there are gaps of .53 x 67 nm between molecular ends. Each molecule may be regarded as composed of five segments (labeled 1-5), each of 67 nm length except for segment 5, which is .47 x 67 nm long. Reproduced, with permission, from Miller. 1982. Trends Biochem. Sci. 7:13.

Figure 4

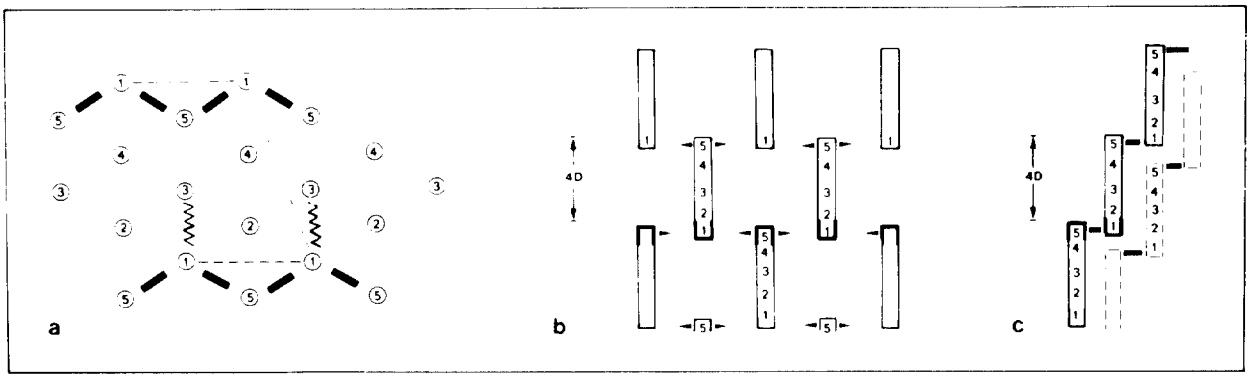

Figure 4. Views of the three-dimensional molecular packing in collagen fibrils. Collagen molecules are represented as cylinders and the five segments (labeled 1-5 as in figure 3) are marked. The 5° tilt of the molecules to the fibril axis is not shown; the molecules are represented as parallel to the fibril axis.
(a) This represents a transverse section 67 nm thick through a fibril to show the lateral disposition of the molecular segments in a quasi-hexagonal lattice. The crystallographic unit cell (a = 3.9 nm, b = 2.67 nm, γ = 104.6°) is shown by broken lines. Segments linked by the polymeric crosslink $(1\text{-}5\text{-}1\text{-}5)_n$ are shown connected by bond lines. The speculative crosslinks which link residues 580 and 1044 are shown by zig-zag lines. These speculative crosslinks would tie together sheets of the type shown in figure 3(b).
(b) A sheet of collagen molecules linked by the $(1\text{-}5\text{-}1\text{-}5)_n$ polymeric crosslinks is shown in face-on projection. The crosslinks do not lie in the plane of the page and have been represented by triangles to indicate that molecules thus linked are at different levels perpendicular to the plane of the page. The three molecules in axial register at the bottom of the diagram may be imagined as closest to the reader while those at the top are furthest from the reader. The sheet is thus stepped down as is shown in side view in figure 3(c). This sheet has a period of 4 x 67 nm when projected on to the molecular axis.
(c) In bold lines, a side view of the sheet shown in figure 3(b). This projection along a single sheet has an axial period of 4 x 67 nm, thus the $(1\text{-}5\text{-}1\text{-}5)_n$ polymeric crosslinks can not alone generate the axial period of 67 nm in collagen fibrils. Sheets similar to that in figure 3(b) must be staggered with respect to each other by 67 nm to produce the 67 nm period. Such a staggered sheet is shown in broken lines and if additional staggered sheets are added, the three-dimensional arrangement, shown in section in figure 3(a), will be formed.
Reproduced, with permission, from Miller. 1982. Trends Biochem. Sci. 7:13.

References

Bruckner, P., Bachinger, H.P., Timpl, R., Engel, J. 1978. Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple-helix ⇌ coil transition. Eur. J. Biochem. 90:595.

Burjanadze, T.V. 1979. Hydroxyproline content and location in relation to collagen thermal stability. Biopolymers 18:931.

Dehl, R.E. 1970. Collagen: mobile water content of frozen fibers. Science 170:738.

Dick, J.P., Nordwig, A. 1966. Effect of pH on the stability of the collagen fold. Arch. Biochem. Biophys. 117:466.

Dutson, T.R. 1976. Biosynthesis and structure of collagen. Proc. Recip. Meat Conf. 29:336.

Eyre, D.R. 1980. Collagen: molecular diversity in the body's protein scaffold. Science 207:1315.

Fung, B.M., Witschel, J., McAmis, L.L. 1974. State of water on hydrated collagen as studied by pulsed NMR. Biopolymers 13:1767.

Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K., Woodhead-Galloway, J. 1973. Analysis of the primary structure of collagen for the origins of molecular packing. J. Mol. Biol. 79:137.

Judge, M.D., Aberle, E.D. 1982. Effects of chronological age and postmortem aging on thermal shrinkage temperature of bovine intramuscular collagen. J. Anim. Sci. 54:68.

Kruggel, W.G., Field, R.A. 1971. Soluble intramuscular collagen characteristics from stretched and aged muscle. J. Food Sci. 36:1114.

Light, N.D., Bailey, A.J. 1980. The chemistry of collagen crosslinks. Purification and characterization of crosslinked polymeric peptide material from mature collagen containing unknown amino acids. Biochem. J. 185:371.

Light, N.D., Bailey, A.J. 1980. Polymeric C-terminal crosslinked material from type I collagen. A modified method for purification, molecular weight estimation, carbohydrate content and lipid content. Biochem. J. 189:111.

Lim, V.I. 1981. A novel structural model for collagen: water-carbonyl helix. FEBS Letters 132:1.

Luescher, M., Ruege, M., Schindler, P. 1974. Effect of hydration upon the thermal stability of tropocollagen and its dependence on the presence of neutral salts. Biopolymers 13:2489.

McClain, P.E. 1976. Chemistry of collagen crosslinking. Proc. Recip. Meat Conf. 29:350.

McClain, P.E., Creed, G.J., Wiley, E.R., Hornstein, I. 1970. Effect of postmortem aging on isolation of intramuscular connective tissue. J. Food Sci. 35:258.

Miller, A. 1982. Molecular packing in collagen fibrils. Trends Biochem. Sci. 7:13.

Morris, N.P., Fessler, L.I., Weinstock, A., Fessler, J.H. 1975. Procollagen assembly and secretion in embryonic chick bone. J. Biol. Chem. 250:5719.

Pfeiffer, N.E., Field, R.A., Varnell, T.R., Kruggel, W.G., Kaiser, I.I. 1972. Effects of postmortem aging and stretching on the macromolecular properties of collagen. J. Food Sci. 37:897.

Privalov, P.L. 1982. Stability of proteins. In C.B. Anfinsen, J.T. Edsall and F.M. Richards, eds., Advances in Protein Chemistry, Vol. 35, pp. 1-104. Academic Press, New York.

Prockop, D.J., Kivirikko, K.I., Tuderman, L., Guzman, N.A. 1979a. The biosynthesis of collagen and its disorders I. N. Eng. J. Med. 301:13.

Prockop, D.J., Kivirikko, K.I., Tuderman, L., Guzman, N.A. 1979b. The biosynthesis of collagen and its disorders II. N. Eng. J. Med. 301:77.

Ramachandran, G.N., Kartha, G. 1955. Structure of collagen. Nature. (Lond.) 176:593.

Ramachandran, G.N., Ramakrishnan, C. 1976. Molecular structure. In G. N. Ramachandran and A. H. Reddi, eds., Biochemistry of Collagen. pp. 45-84. Plenum Press, New York.

Ramachandran, G.N., Reddi, A.H. 1976. Biochemistry of Collagen. Plenum Press, New York.

Rich, A., Crick, F.H.C. 1955. The structure of collagen. Nature (Lond.) 176:915.

Rigby, B.J., Prossner, C.L. 1975. Thermal transitions of collagen from fish recovered from different depths. Comp. Biochem. Physiol. 52B:89.

Rigby, B.J., Robinson, M.S. 1975. Thermal transitions in collagen and the preferred temperature range of animals. Nature (Lond.) 253:277.

Shimokomaki, M., Elsden, D.F., Bailey, A.J. 1972. Meat tenderness: age related changes in bovine intramuscular collagen. J. Food Sci. 37:892.

Smith, J.W. 1968. Molecular pattern in native collagen. Nature (Lond.) 219:157.

Tryggvason, K., Risteli, J., Kivirikko, K.I. 1977. Separation of prolyl 3-hydroxylase and 4-hydroxylase activities and the 4-hydroxyproline requirement for synthesis of 3-hydroxyproline. Biochem. Biophys. Res. Comm. 76:275.

von Hippel, P.H., Wong, K.Y. 1963. The collagen $\rightleftharpoons$ gelatin phase transition I. Further studies of the effects of solvent environment and polypeptide chain composition. Biochemistry 2:1387.

Wiestner, M., Krieg, T., Horlein, D., Glanville, R.W., Feitzek, P.P., Muller, P.K. 1979. Inhibiting effect of procollagen peptides on collagen biosynthesis in fibroblast cultures. J. Biol. Chem.

Wu, J.J. 1978. Characteristics of bovine intramuscular collagen under various postmortem condition. Ph.D. Dissertation. Texas A&M Univ., College Station.