

## 6

**Collagens and Gelatins**

**Dr. Barbara Brodsky<sup>1</sup>, Dr. Jerome A. Werkmeister<sup>2</sup>, Dr. John A. M. Ramshaw<sup>3</sup>**

<sup>1</sup> Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School,  
Piscataway, NJ 08854, USA; Tel.: +1-732-235-4048; Fax: +1-732-235-4783;  
E-mail: "brodsky@umdnj.edu"

<sup>2</sup> CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia; Tel.: +61-3-96627239;  
Fax: +61-3-96627218; E-mail: Jerome.Werkmeister@csiro.au

<sup>3</sup> CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia; Tel.: +61-3-96627217;  
Fax: +61-3-96627218; E-mail: John.Ramshaw@csiro.au

|          |                                                                                            |     |
|----------|--------------------------------------------------------------------------------------------|-----|
| <b>1</b> | <b>Introduction</b>                                                                        | 120 |
| <b>2</b> | <b>Historical Outline</b>                                                                  | 121 |
| 2.1      | Early Background                                                                           | 121 |
| 2.2      | The Golden Age of Collagen Discoveries: Composition, Conformation, and Fibril Organization | 122 |
| 2.3      | Amino Acid Sequence, Genetic Diversity, and Precursors                                     | 123 |
| <b>3</b> | <b>Chemical Structure</b>                                                                  | 124 |
| 3.1      | Unusual Amino Acid Features and Post-translational Modifications                           | 124 |
| 3.2      | Cross-links Between Collagen Chains                                                        | 126 |
| <b>4</b> | <b>Structure of the Collagen Triple Helix</b>                                              | 127 |
| <b>5</b> | <b>Occurrence and Function of Collagen</b>                                                 | 129 |
| 5.1      | Fibril-forming Collagens                                                                   | 129 |
| 5.2      | Non-fibrillar Collagens                                                                    | 131 |
| <b>6</b> | <b>Biosynthesis and Biochemistry of Collagen</b>                                           | 131 |
| <b>7</b> | <b>Physiology</b>                                                                          | 133 |
| <b>8</b> | <b>Biodegradation</b>                                                                      | 134 |

|           |                                                           |     |
|-----------|-----------------------------------------------------------|-----|
| <b>9</b>  | <b>Polypeptide Models of Collagen</b> . . . . .           | 135 |
| <b>10</b> | <b>Molecular Genetics and Collagen Diseases</b> . . . . . | 136 |
| <b>11</b> | <b>Chemical Analysis and Detection</b> . . . . .          | 137 |
| <b>12</b> | <b>Production and Applications</b> . . . . .              | 138 |
| 12.1      | Biomedical Applications of Collagen . . . . .             | 139 |
| 12.2      | Gelatin Production and Applications . . . . .             | 140 |
| 12.3      | Recombinant Collagen and Gelatin . . . . .                | 140 |
| 12.4      | Immunological Response to Collagen Products . . . . .     | 141 |
| <b>13</b> | <b>Patents</b> . . . . .                                  | 142 |
| <b>14</b> | <b>Outlook and Perspectives</b> . . . . .                 | 146 |
| <b>15</b> | <b>References</b> . . . . .                               | 147 |

|     |                         |
|-----|-------------------------|
| Hyl | hydroxylysine           |
| Hyp | hydroxyproline          |
| P4H | prolyl-4-hydroxylase    |
| MMP | metalloproteinase       |
| OI  | osteogenesis imperfecta |
| SLS | segment long spacing    |

## 1 Introduction

Collagens are the predominant fibrous proteins in animals. They constitute a major structural component in the extracellular matrix of all multicellular animals, including sponges, invertebrates, and vertebrates. The most abundant and well-characterized collagens are found in axially banded fibrils with a 670-Å repeat (Kielty et al., 1993). These collagen fibrils make up as much as 30% of the total protein in humans and other higher animals. The role of such fibrils is to maintain tissue integrity and to provide specific mechanical properties needed for each connective tissue. For example, collagen fibrils form the basis for the strength of tendons, the compressibility of cartilage, and

the flexibility of blood vessels and skin. There are also other genetic types of collagen that do not form fibrils and are found in much smaller quantities in tissues. These play critical structural roles, including formation of the framework and permeability barrier of basement membranes, linking of fibrils to other matrix components, and anchoring of dermis to its associated basement membrane.

From a biomedical view, collagen has been shown to play important roles in development, wound healing, platelet activation, and angiogenesis. In addition, a growing number of hereditary diseases have been shown to be caused by mutations in collagen genes. Perturbations in collagen synthesis, degradation, or immunology also have been noted in many common diseases of complex

etiology, including arthritis, liver fibrosis, diabetes, and cancer, as well as aging.

The commercial importance of collagen relates to the wide application of gelatin as well as leather production and biomedical applications. Gelatin is derived from collagen in tissues by heat denaturation. On cooling, the chains can rewind, but not in the correct register, and the small triple-helical segments formed may further aggregate during gel formation. The uses of gelatin in films and in food are widespread.

Collagen has a uniform triple-helical conformation, which is determined by its repeating (Gly-X-Y)<sub>n</sub> sequence pattern and high imino acid content. Polypeptides with these sequence features also adopt this conformation (Heidemann and Roth, 1982). X-ray diffraction, electron microscopy, hydrodynamic methods, and a variety of spectroscopy techniques have been used extensively in characterizing collagen molecules, collagen fibrils, and gelatin. In this chapter, collagen and gelatin will be reviewed in a biopolymer context in terms of chemical structure, production, and methods of detection and analysis. The biologically unique features of collagen, including biosynthesis, functions, and disease, also will be presented. A number of excellent books and chapters on collagen describe the biochemical and biomedical aspects in further detail (Nimni; 1988–1991; Royce and Steinmann, 2002; Kadler, 1994; Bateman et al., 1995).

## 2

### Historical Outline

The field of collagen is an old one that has blossomed in multiple directions in recent years. Further details on the history of collagen and gelatin can be found in a range other of texts (e.g., Gustavson, 1956;

Ramachandran, 1967a; Fraser and MacRae, 1973; Mayne and Burgesson, 1987).

#### 2.1

##### Early Background

Collagen has been a useful material to mankind for millennia. Indeed, it is not clear when it was first recognized that hides and skins could be preserved against decay, by tanning the collagen, to give leather. While materials such as hair probably first were used as medical sutures circa 5000 years ago (see Breasted, 1930), collagen “catgut” sutures were also developed and used for repair of wounds, e.g., by Claudius Galenus, circa 175, in treating wounded gladiators (Mackenzie, 1971). The denatured products of collagen, animal glues and gelatin, were also in early use, with Pliny the Elder noting that “Boum coriis glutinum excoquitur, taurorum praecipuum” – glue is boiled from the hides of cattle, and the best from those of bulls (Pliny, circa 50). It was also recognized in historical times that collagen, as gelatin- or glue-forming material, was found widely distributed throughout all animal phyla.

The chemistry and other applications of collagen, however, developed over the last 200 years. During the 1820s, studies on the hydrolysis products of the natural substances that were later called proteins led to the isolation of a new substance from gelatin, “sugar of gelatin”, so named because of its sweet taste (Braconnot, 1820). It was distinct from sugar, and further analyses (Mulder, 1838; Gerhardt, 1846) led to a structure proposal (Cahours, 1857) for what was by then called glycine (Berzelius, 1848). Boiling acid treatment of gelatin and other proteins, initially by H<sub>2</sub>SO<sub>4</sub> but later by HCl (Bopp, 1849; Hlasiwetz and Habermann, 1873), led by 1900 to the isolation of 13 amino acids. This led to the concept that proteins are

polymers of amino acids (Fischer, 1902a; Hofmeister, 1902). Further studies on gelatin led to the discovery of hydroxyproline (Hyp) (Fischer 1902b) and of hydroxylysine (Hyl) (Van Slyke and Hiller, 1921; Van Slyke et al., 1938). However, it was only realized considerably later that these amino acids were formed by secondary modification and were not directly incorporated into protein (Stetten, 1949; Sinex and Van Slyke, 1957).

Early histologists, when treating sections with acid (e.g., Gillette, 1872), would have been extracting soluble collagen; for example, tendons were observed to swell and dissolve in dilute acetic acid (Zacchariades, 1900). The initial chemical studies on soluble collagen were mainly by Nageotte and colleagues in France from the mid-1920s (Nageotte, 1927). Nageotte found that acetic acid extracts from various tissues contained soluble collagen and that this could be precipitated by salt and re-aggregated into fibrils by dialysis at neutral pH against water. These and other studies on general solubility properties under different conditions became the basis for collagen purification (Tustanovskii, 1947; Orekhovich, et al., 1948; Gross et al., 1955).

## 2.2

### **The Golden Age of Collagen Discoveries: Composition, Conformation, and Fibril Organization**

In the 1940s and 1950s, a picture of the unique features of the amino acid composition and conformation of the collagen molecule emerged, together with an understanding of the collagen fibril structure. The composition of collagen differed from any known protein; its most striking features were the high contents of glycine and of the imino acid proline (see Eastoe, 1967). The first experimental attempts to determine the primary structure, using partial acid hydro-

lysates, gave indications of a regular structure, suggesting that glycine might be present as every third residue along the chain (Schroeder et al., 1953). Comparisons between collagen and gelatin showed close similarity in composition, but gelatin had a lower ionic point as a result of amide hydrolysis and showed partial conversion of arginine to ornithine. Mature vertebrate collagen also was found to have a low total carbohydrate content, usually below 1%. Later, in 1966, it was shown that this was galactose or glucosylgalactose bound to the hydroxyl groups of Hyl (Butler and Cunningham, 1966).

In the mid-1950s, the understanding of the unique amino acid composition features of collagen, together with fiber X-ray diffraction studies, led to the first proposal of a triple-helical structure for collagen (Ramachandran and Kartha, 1954). The fiber diffraction seen for collagen in tail tendons was distinct from that seen for any other fibrous protein (Astbury, 1933), showing a strong axial reflection at 2.9 Å and a strong equatorial reflection that varied between 11 Å and 16 Å, depending on the degree of hydration (Rougvie and Bear, 1953). Improvement of the diffraction patterns by stretching (see Cowan et al., 1955) showed the transform of a rod-like structure with helical symmetry. A three-stranded, triple-helical model for collagen comprising left-handed helical chains related by a right-handed rope twist was suggested by Ramachandran and Kartha (1954) to account for the X-ray pattern features and the unusual amino acid features. This structure, in which the three chains are arranged about a common central axis, leaves no space in the inner core for C<sub>β</sub> atoms and therefore must be occupied by Gly residues. Pro residues could be accommodated in the non-glycine positions without distortion. Rich and Crick (1961) pointed out that the

initial model violated stereochemical limitations, but that modifications to the basic three-chain concept could be made to give a model that was sterically satisfactory. The modifications focused on the arrangement of the hydrogen bonds, and the model designated as RCII was preferred by virtue of stereochemical considerations.

A detailed evaluation of the chain conformation in the collagen molecule using quantitative X-ray diffraction data collected from stretched kangaroo tail tendon (Fraser et al., 1979) gave an estimated value for the unit twist of the molecular helix of  $107.1^\circ \pm 0.6^\circ$ , which is close to the value expected ( $108^\circ$ ) for a helix with 10 units in 3 turns, although there is no *a priori* reason why such a simple relationship should exist. The best solution very closely resembled the RCII structure. The model was supported by physicochemical investigations suggesting that collagen molecules consisted of three polypeptides each with length of about 1000 amino acids (Boedtker and Doty, 1956).

During this fertile time on collagen research, the molecular organization in the collagen fiber was also determined. Electron micrographs of native collagen fibers or fibrils reconstituted from soluble collagen (Hall et al., 1942; Schmitt et al., 1942) showed a simple banded appearance with cross striations at 600 Å to 700 Å, similar to the period of 670 Å observed in low angle X-ray reflections carried out by Bear (1942). This axial repeat is usually designated as the D period. Two alternative forms of collagen aggregates were also visualized: (1) the centrosymmetric structure fibrous long spacing (FLS) formed when soluble collagen was mixed with highly negatively charged compounds (Highberger et al., 1950); and (2) segment long spacing (SLS), a parallel array of in-register molecules that reflected the 300-nm molecular length of collagen (Schmitt et al., 1953). Schmidt and

colleagues suggested, from comparison of the cross-striation period along the collagen fibril with the length of the collagen molecule, that the collagen fibril was a parallel array of molecules staggered by one-quarter (Schmitt et al., 1955), and more precise studies indicated a stagger of  $D = 670$  Å, where the molecule length is equal to  $4.4D$  periods (see Hodge, 1967). Early studies using lathyrogens, which interfere with cross-linking formation, led to the conclusion that these D-periodic fibrils contain intermolecular covalent cross-links that lead to their high tensile strength and mechanical stability.

### 2.3

#### Amino Acid Sequence, Genetic Diversity, and Precursors

Introduction of enzyme methods (e.g., Kuhn et al., 1963) gave a simple method for large-scale extraction of collagen. Initial separations of collagen chains by ion exchange chromatography (Kessler et al., 1959) led to partial degradation, but the introduction of milder conditions (Piez et al., 1960) led to successful separation and characterization of the intact subunit chains of soluble collagen (Piez et al., 1963). These data showed that the  $\alpha$  components were single polypeptide chains, with two types of these chains in the majority of tissues studied. The  $\beta$  and  $\gamma$  components consist of two and three polypeptide chains, respectively, joined together by covalent cross-links that survive denaturation. Separation of the individual chains of collagen by ion exchange chromatography (Piez et al., 1963) enabled peptides obtained by chemical cleavage, e.g., by CNBr, hydroxylamine, or enzymes (Bornstein and Piez, 1966), to be separated and used for sequence analysis. The first complete chain sequence, for the bovine  $\alpha 1(I)$  chain, was achieved in 1972 (see Fietzek and Kuhn,

1976). These data showed an excellent correlation between sequence and the cross-striations seen in SLS aggregates and intact fibrils. SLS structures prepared from a wide range of different species indicated the remarkable stability of the amino acid sequence during evolution (Mathews, 1975). Automated protein-sequencing procedures accelerated the collection of sequence data, while in recent years, DNA-sequencing technology has greatly increased the available sequence database (see Kadler 1994). With recombinant technology, the complexity of the collagen gene with a large number of coding and non-coding sequences became apparent (Wozney et al., 1981).

Until the 1970s, collagen had been considered as a single entity. Although Bazin and Delauney (1964) first noticed that granulation tissue appeared to contain a type of collagen that was distinct from type I collagen and similar to the type often associated with embryonic collagens in skin, an alternative collagen type was first clearly identified in articular cartilage by Miller (1971). Soon after, the distinct collagen of granulation tissue, skin, and other tissues was characterized as collagen type III (Epstein, 1974; Chung and Miller, 1974). At least 20 genetically distinct collagen types have now been described (Kielty and Grant, 2002). A precursor form of collagen was first observed in 1971, in biosynthetic investigations of fibroblast cultures (Layman et al., 1971) and in the skin of calves with an inherited fragile skin disease (Lenaers et al., 1971). These observations led to significant, on-going efforts using cellular and molecular biology approaches to understand the biosynthesis of collagens. This has provided data on intracellular processes including transcription, translation, chaperones, the variety of secondary modification enzymes, chain recognition, and procollagen protease activities (see, for example,

Royce and Steinmann, 1993; Bateman et al., 1995).

### 3

#### Chemical Structure

Collagen is a protein, synthesized from the 20 common amino acids, yet it is unique in terms of its amino acid composition, repeating sequence pattern, high degree of post-translational modification, and characteristic intermolecular cross-links (Kielty et al., 1993).

#### 3.1

##### Unusual Amino Acid Features and Post-translational Modifications

The amino acid composition of collagen shows an unusually high content of Gly, near 33%, and a high content of imino acids, near 20%. Examination of its sequence shows that Gly is present as every third residue, producing a (Gly-X-Y)<sub>n</sub> repeating pattern. Collagen contains a considerable amount of the unusual amino acid Hyp, which is rarely found in other animal proteins, and significant amounts of Hyl, both formed by enzymatic post-translational modification. Hyl groups are often glycosylated, resulting in Hyl-galactosyl and Hyl-galactosyl-glucosyl moieties. The Hyp is almost exclusively trans-4-Hyp, containing the OH group on the  $\gamma$ -carbon of the prolyl ring, although small amounts of 3-Hyp have been reported in invertebrates and basement membrane collagen (Mathews, 1975). The most common tripeptide sequence found in collagen is Gly-Pro-Hyp (Figure 1).

In contrast to the restrictive nature of every third position being Gly, the other two positions in each repeating unit, X and Y, can accommodate any of the 20 amino acids, including imino acids, without distortion.

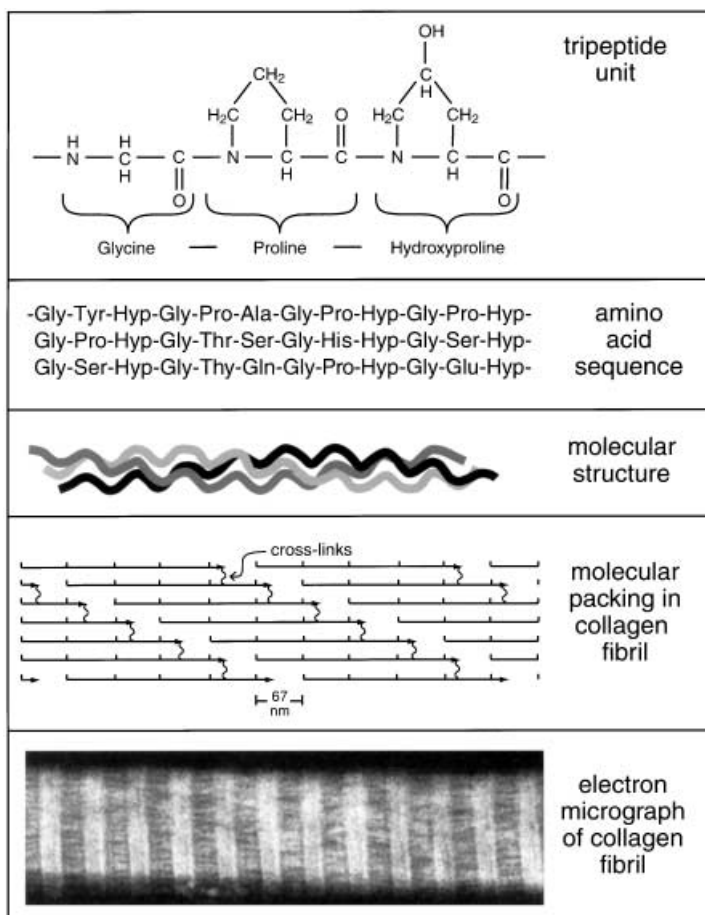


Fig. 1 The chemical structure of collagen and its higher levels of organization.

This was shown first by model building (Rich and Crick, 1961) and has been confirmed recently by studies on host–guest peptides (Persikov et al., 2000). These peptides of sequence (Gly-Pro-Hyp)<sub>3</sub>-Gly-X-Y-(Gly-Pro-Hyp)<sub>4</sub> indicate that all residues can be incorporated in either the X or Y positions, but the stability of the triple helix depends on the guest residue.

The restricted backbone of the cyclic imide group results in  $\phi, \psi$  angles very close to those found in native collagen; therefore, Pro and Hyp groups are entropically favored

to adopt the triple-helix conformation (Josse and Harrington, 1964). The high imino acid content of collagen affects its folding as well as its geometry. There is a significant probability of *cis* peptide bond configurations in the unfolded state for both Gly-Pro bonds (~14%) and Pro-Hyp bonds (~6%) (Sarkar et al., 1984). Since only *trans* peptide bonds are found in native triple-helical collagen, *cis* bonds in the unfolded chains need to undergo isomerization, an inherently slow process that limits the rate of propagation *in vitro*.

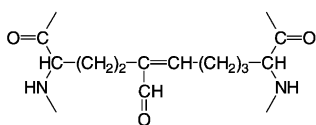
## 3.2

**Cross-links Between Collagen Chains**

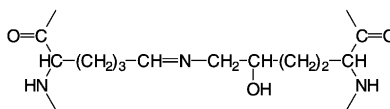
The formation of specific covalent cross-links between collagen molecules stabilizes collagen fibrils in tissues (Yamauchi and Mechanic, 1988; Bailey, 2001; Brady and Robins, 2001). Lysyl oxidase initiates collagen cross-linking by catalyzing the formation of Lys- and Hyl-derived aldehydes, i.e., allysine and hydroxyallysine, at specific residues in the telopeptides. These aldehydes then undergo a series of condensation reactions with adjacent Lys residues from the telopeptide or with specific lysine residues from the triple-helical domain to provide the initial cross-links (Figure 2). These cross-links, the reducible cross-links, can be stabilized for analysis by reduction with sodium borohydride. Further spon-

taneous reactions result in a variety of other multifunctional cross-links that characterize mature tissues (Figure 2), although only a limited number have been characterized. The number and proportions of the various cross-links show tissue specificity, and it is likely that this is regulated by the steric relationship between molecules, the type of collagens copolymerized, and glycosylation and hydroxylation of the participating amino acid residues. In mature tissues, non-enzymatic glycosylation can further cross-link the collagen through non-specific cross-links at various locations, leading to decreased solubility and a characteristic fluorescence (Monnier and Cerami, 1982). This additional cross-linking can make tissues too stiff to function properly and can affect the filtration properties of the glomerular basement membrane. Because cross-linking may

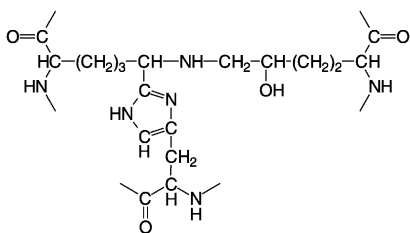
a) Aldol Condensation Product (ACP)



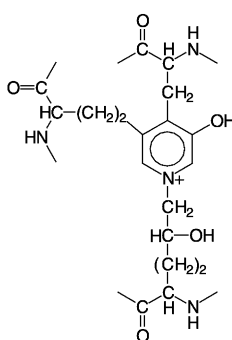
b) dehydro-Hydroxylysinonorleucine (deH-HLNL)



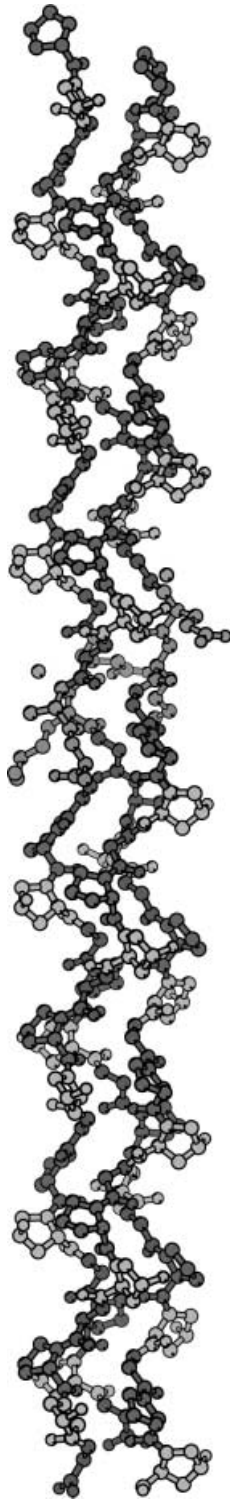
c) Histidinohydroxylysinonorleucine (HHL)



d) Pyridinolone (Py)



**Fig. 2** The chemical structures of examples of collagen cross-links.



◀ **Fig. 3** The triple-helix conformation from a peptide crystal structure (Kramer et al., 2000).

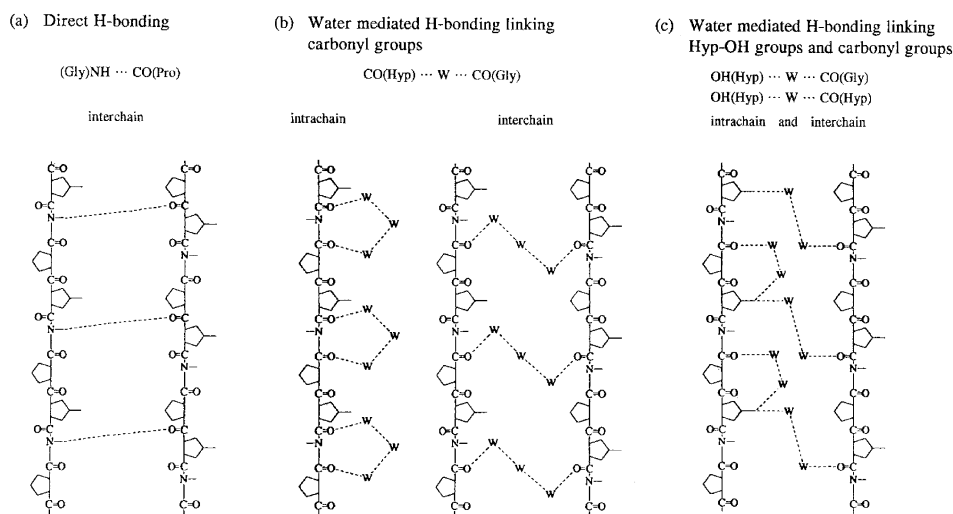
be increased in diabetic patients, it is a potential cause of the tissue-based problems that emerge in this disease.

#### 4

#### Structure of the Collagen Triple Helix

The collagen family represents a group of diverse molecular structures, yet all are linked by a common structural element – the collagen triple-helix structure (Rich and Crick, 1961; Bella et al., 1994; Brodsky and Ramshaw, 1997). The collagen triple helix is comprised of three polypeptide chains, each in a left-handed extended polyproline II-like helix (Figure 3). The three chains, which are parallel and staggered by one residue with respect to each other, are supercoiled about a common axis in a right-handed manner. Interchain peptide hydrogen bonds link the three chains. The close packing of the three chains in the triple-helix requires that every third residue in the sequence be a Gly residue, while the extended conformation of the individual chains is stabilized by a high content of imino acids.

The original triple-helix model was constructed and refined on the basis of fiber diffraction studies of tendons, but more recently, crystal structures of triple-helical peptides at resolutions higher than 2 Å have allowed elucidation of the details of the triple-helical structure at a molecular level (Bella et al., 1994; Kramer et al., 1998, 1999, 2000, 2001; Vitagliano et al., 2001; Berisio et al., 2001). These studies confirm the overall features proposed earlier. The one direct NH (Gly)...CO (X position) hydrogen



**Fig. 4** A schematic illustrating interchain hydrogen bonding and water-mediated hydrogen bonds in the collagen triple helix.

bond between chains can be visualized, and when Pro does not occupy the X position, an additional interchain hydrogen bond, NH (X)...CO (Gly), is formed through one mediating water molecule. In contrast to alpha-helices and beta-sheets, where all NH and C=O groups of the peptide backbone are involved in hydrogen bonds, the collagen triple-helix always has available backbone C=O groups and, in some cases, NH groups. An extensive hydration network is part of the response to satisfy available backbone groups, and this is visualized in the crystal structures (Figure 4). The role of Hyp in collagen stability may be related to this hydration network (Privalov, 1982) and/or to inductive effects that influence the puckering of the imide ring structure (Bretscher et al., 2001). X-ray crystallography studies indicate that imino-acid-rich sequences, such as repeating Gly-Pro-Hyp units, adopt a helical symmetry of 7/2, while imino-acid-poor regions have a symmetry closer to 10/3, the average seen in collagen fiber diffraction patterns (Kramer et al.,

1999). Such sequence-dependent variation in superhelix pitch could play a role in recognition and intermolecular interactions. The high-resolution structures also show the conformers and interactions of various side chains in the triple helix (Kramer et al., 2000, 2001). Some side chains hydrogen bond directly to available backbone carbonyls, while many interactions are mediated by water.

NMR studies on peptides with specifically labeled residues have confirmed the one-residue stagger and close packing of the three chains in solution (Li et al., 1993). It also has been possible to monitor the dynamics and folding of individual residues in the triple helix. Computational modeling has been used to clarify the features and interactions of the triple helix and, more recently, to clarify the thermodynamic features of the helix (Mooney et al., 2001).