

MODEL OF THE HELICAL PORTION OF A TYPE I COLLAGEN MICROFIBRIL*

by

ELEANOR M. BROWN, GREGORY KING** AND JAMES M. CHEN***
U.S. Department of Agriculture, Agricultural Research Service
Eastern Regional Research Center
600 EAST MERMAID LANE
WYNDMOOR, PA 19038

ABSTRACT

Collagens are an ubiquitous family of extracellular proteins in the animal world. Fibril-forming collagens are the major structural proteins. In addition to their biological function, collagens provide molecular frameworks that are utilized in the medical, food and leather industries. Functionality, both biological and technological, is a consequence of the basic structure of a protein. Computerized molecular modeling has been developed into a useful tool for visualizing structure-function relationships in proteins. The unique supramolecular characteristics of collagen have made it a suitable subject for model builders over more than thirty years. The bovine type I collagen microfibril model developed at ERRC now consists of 15 polypeptide chains of 315 amino acid residues each. This model is large enough to allow a comparison of its gross structural features with the banding patterns obtained by electron microscopy. Measurable distances pertaining to the helix, triple helix, and microfibril are within experimental error of recently reported physical data on dried collagen samples. Two segments of the microfibril model, one rich in hydrophobic residues and the other rich in hydrophilic residues were examined in detail. The conformation of the hydrophobic segment more nearly matched the definition of a collagen helix (57%) than did the conformation of the hydrophilic segment (38%). The model is

potentially useful for studying mechanisms of both inter- and intra-microfibril crosslinking as well as for predicting the efficacy of specific modifications to the molecule or potential crosslinking agents.

INTRODUCTION

Collagens are a family of structural proteins of the extracellular matrix. Fibril forming collagens are the major structural proteins of skin, cartilage, bone, blood vessel walls and internal organs. In addition to their biological function, these collagens are commercially important as natural structural frameworks utilized in the medical, food and leather industries. These functions, both biological and technological, are controlled by the protein structure. An understanding of these structure-function relationships can lead to more efficient technological use of the protein and the prediction of new uses.

Several unique characteristics of the fiber-forming collagens have influenced the choice of techniques for exploring the details of the molecular structures of these proteins. Because these collagens form fibrillar structures large enough to be visualized using relatively primitive instruments, a picture of the basic organization of collagen in triple helical, microfibrillar and fibrillar structures began to emerge in the 1960's.^{1,2} Individual peptide chains are about 1050 amino acid residues long and, except for telopeptides at the amino and carboxy termini, consist entirely of a repeating Gly-X-Y pattern where at least 25%

* Presented in part at the 92nd Annual Meeting of the American Leather Chemists Association, Traverse City, Michigan, June 16-20, 1996.

** Present address, Rosenbluth International, 2401 Walnut St., Philadelphia, PA 19103, USA

*** Present address, Wyeth-Ayerst, Pearl River, NY 10965, USA

of X and Y residues are Pro or Hyp (hydroxyproline) respectively. These features of the sequence made the development of template models feasible. Early work on the preparation and use in structural studies of synthetic collagen-like peptides, reviewed by Carver and Blout,³ provided baseline data for spectroscopic evaluations of conformations in collagen. The (Gly-Pro-Pro)_n computational model of Miller and Scheraga⁴ was the start of a series of related studies that evaluated the effects of specific side chains on conformation. More extensive reviews of these and other studies leading to proposed pathways for the assembly of microfibrils were published at the start of our research⁵ and with a microfibril model of type II collagen.⁶

The structure of Type I collagen, the primary collagen of mature hides, has been widely studied. X-ray scattering studies of non-crystalline collagen⁷ and of a crystallized model peptide⁸ (Pro-Pro-Gly)₁₀ synthesized by Sakakibara and coworkers⁹ have verified the hypothesis of Ramachandran and Kartha¹⁰ that three collagen molecules assemble into triple-helical coiled coils, such that 27 to 29 amino acid residues comprise a complete rotation about the right-handed triple-helical axis. The atomic coordinates of the triple-helical peptide 3(Pro-Pro-Gly)₁₀ were determined by Okuyama and coworkers.¹¹ Electron micrographs of stained collagen fibrils^{12,13} provide some clues as to the gross packing structure of collagen. Such micrographs display a pattern of alternating light and dark bands perpendicular to the axis of the collagen fibrils that repeats every 670 Å. This 670 Å span is defined as a D interval (or D spacing). Light bands correspond to regions of more dense lateral packing, and dark bands correspond to 'gap' regions, domains of low density molecular packing first noted by Hodge and Petruska.¹⁴

The structural intermediates between the triple helix and collagen fibers are poorly defined. A variety of models for these intermediates have been proposed based on observed staining patterns and the length of the triple-helical amino acid sequence.¹⁵⁻²³ It is generally agreed that the fibril can be described as an aggregate of packing units containing four to six triple helices. In the model, proposed by Smith² a microfibril is defined as a bundle of five triple-helical molecules, where adjacent triple helices are staggered longitudinally by 1 D interval. Longitudinally-neighboring triple helices are separated by a 'gap' region approximately 0.6 D interval in length. Cross-sections of the model exhibit a regular pentagonal geometry.

The choice of the Smith microfibril model as a template for our computer model was influenced by the fact that a complete type I collagen molecule 4.4 D intervals in length (5.0 D intervals when a gap region is included) may be represented with minimal redundancy in a 5-molecule microfibril segment 1 D interval in length, due to the relative longitudinal staggering of the triple-helical molecules by 0, 1, 2, 3 and 4 D intervals. Although the microfibril has not been identified experimentally as a separate species,²³ it is used here to visualize intra- and inter-helical interactions in a manageably sized fragment of a collagen fiber. The center-to-center distance of 10 Å between adjacent triple-helical molecules in the model compares well with the value of 10.5 Å indicated by X-ray diffraction data for dry collagen fibrils.²⁴

In preliminary studies we constructed a 36-residue long segment of bovine type I collagen.^{5,25,26} In the present work, we have extended the microfibril model to include 15 polypeptide chains, each 315 amino acid residues in length, of bovine Type I collagen. This model is large enough to be useful in comparing gross structural features with the images of stained collagen obtained by electron microscopy (EM), in predicting the effectiveness of proposed tanning agents for crosslinking collagen molecules to stabilize leather, as well as in evaluating the probable efficacy of other industrial processes using collagen. At the same time this model is small enough to be manipulated on a mini-computer or workstation.

METHODOLOGY

The three-dimensional computer model of the bovine type I collagen microfibril was constructed using SYBYL[†] molecular modeling software (Tripos Associates, St. Louis, MO, USA) on a SGI Indigo workstation (Silicon Graphics, Mountain View, CA, USA) and on the Cray at the North Carolina State Super Computing Center. Previous studies performed at this laboratory^{5,25,26} describe the assembly of computer models of relatively short segments of collagen-like molecules. These molecules are the building blocks that were used in the construction of the present microfibril model.²⁷ To construct the present type I collagen microfibril model which includes a full D-space (234 amino acid residues in length), some modifications to the earlier procedures were necessary and have been published in detail.²⁷ Briefly, triple-helical 3(Gly-Pro-Hyp)₁₂ models were joined with a splicing operation until the model structure 3(Gly-

† Mention of brand or firm names does not constitute an endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.

Pro-Hyp)₆₉ was achieved. Five copies of this model triple-helical molecule were bundled together to form the microfibril 5[3(Gly-Pro-Hyp)₆₉]. Finally, two 5[3(Gly-Pro-Hyp)₆₉] microfibrils were spliced together to produce a longer microfibril 5[3(Gly-Pro-Hyp)₁₀₅]. The 15 chain microfibril model was refined by energy minimization at both the 207 and 315 residue lengths.

After the completion of the structural scaffolding for the type I collagen microfibril model, the actual type I sequence could be substituted into the model. In order to minimize the disruption of secondary structure, side chains were modified in a step-wise manner, such that at each modification affected side chains were increased in size by at most one non-hydrogen atom. The bulkier side chains were thus allowed to slowly grow into their sites within the protein.²⁷ Thus, Pro or Hyp residues were converted to Ala in a single step, but the conversion of Pro or Hyp to Arg was done with four intermediate steps (Val, Leu, Met, and Lys).

RESULTS AND DISCUSSION

The completed microfibril model 5[3(Gly-X-Y)₁₀₅] with the actual type I sequence is displayed in Figure 1. The

functional characteristics of collagen side chains are illustrated in the coloring of the model. The collagen scaffold consisting of peptide backbone and the Gly, Pro, and Hyp side chains are in white. Other side chains have been colored by type as follows: hydrophobic side chains (Ala, Ile, Leu, Met, Phe, Val) in gold; neutral polar side chains (Asn, Gly, Ser, Thr) in green; positively charged side chains (Arg, Lys, His) in purple and negatively charged side chains (Asp, Glu) in red. As can be seen from this figure, the different types of residues are not distributed uniformly throughout the length of the molecule, but instead form regions of relatively high and low hydrophobicity.

One goal of this research was to test the computer model of the Smith microfibril against the banding pattern observed in electron micrographs of stained collagen (Figure 2). Each fibril is composed of several thousand microfibrils, presumably with every possible angular orientation. In order to compare the computer model to the micrograph, we created an "average" computer model by manipulating the colors of the atoms. All atoms belonging to charged side chains were assigned one color and the rest of the atoms were assigned a different color. A color filtering operation in which all atoms within the same cross-sectional slice

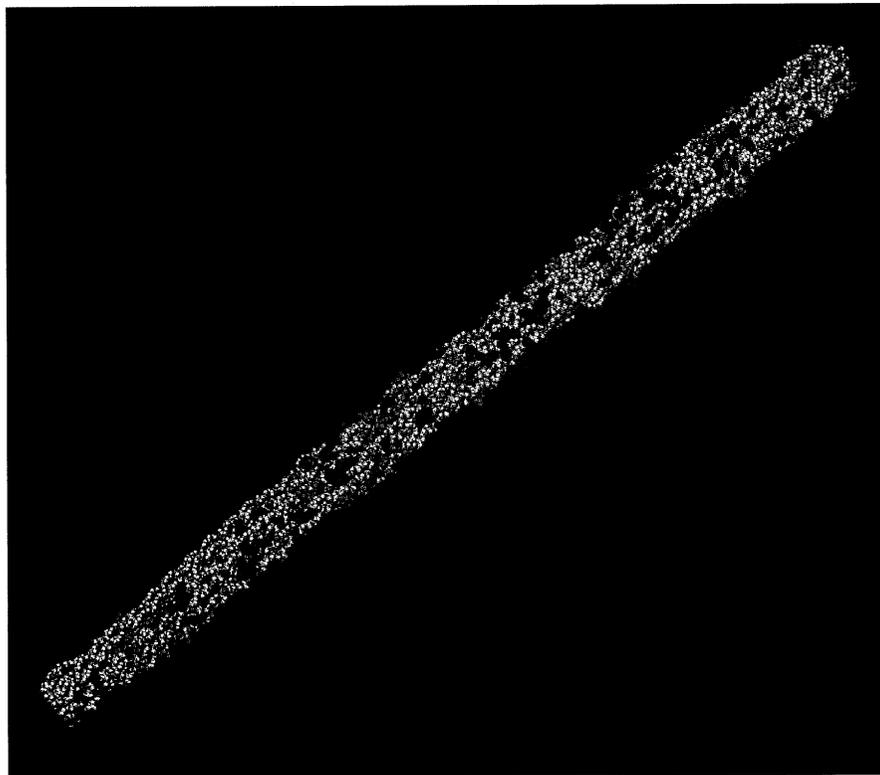


FIGURE 1. — A 400 Å segment of the microfibril model colored to illustrate the functional characteristics of side chains. Hydrophobic side chains (Ala, Ile, Leu, Met, Phe, Val) are in gold, neutral polar side chains (Asn, Gln, Ser, Thr) are in green, positively charged side chains (Arg, Lys, His) are in purple, negatively charged side chains (Asp, Glu) are in red and the collagen scaffold consisting of peptide backbone and the Gly, Pro, and Hyp side chains are in white.

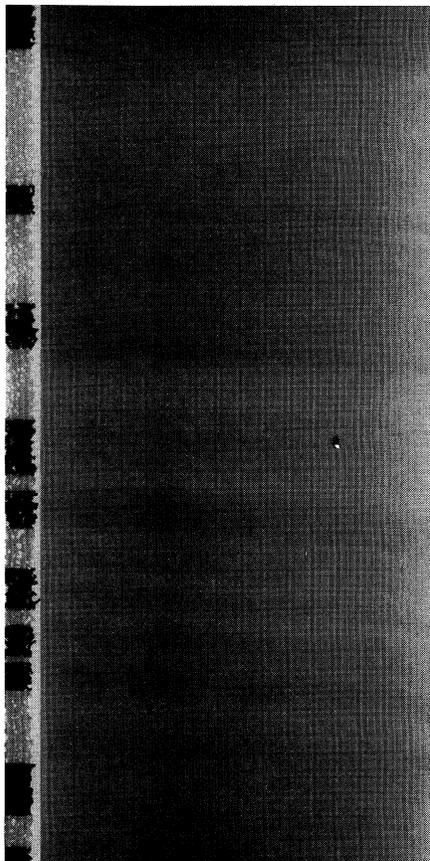


FIGURE 2. — Electron micrograph of a segment-long-spacing crystallite of calf skin collagen. The banding pattern obtained from our computer model has been aligned with the micrograph.

were given the same color was then carried out. All colors were then converted to shades of gray in order to provide a more direct comparison with the micrograph. As can be seen in Figure 2 the computer model displays a banding pattern similar to that seen in the electron micrograph of calf skin collagen. In addition to reproducing the banding pattern, the molecular dimensions of the model are within experimental error of those reported for dried collagen samples.²⁷

Another goal of this research is to use the model in the exploration of possible mechanisms for the kinds of crosslinks produced by different tanning methods. Metal tannages, chromium in particular, are likely to crosslink acidic residues, while vegetable tannins are more likely to interact with hydrophobic sites. Because crosslinks that occur naturally in collagen have been reported to involve residues in the nonhelical portions of collagen as well as in the helical parts,²⁸ it seemed worthwhile to determine how closely segments of the microfibril containing functional side chains conform to the helical definition. With this goal in mind, two short segments of the microfibril were extracted from the model. This extraction could be considered the equivalent of looking at a small region of the molecule with

a magnifying glass. One segment (Figure 3a) is rich in hydrophilic side chains while the other (Figure 3b) is richer in hydrophobic side chains.

The definition of secondary structure in proteins is derived from the dihedral angles for the amino acid residues along the peptide backbone. In the construction of the $3(\text{Gly-Pro-Hyp})_n$ template the dihedral angles ϕ and ψ of Miller and Scheraga⁴ were used. These are $\phi = -74^\circ$, $\psi = 170^\circ$ for Gly, $\phi = -75^\circ$, $\psi = 168^\circ$ for the first Pro, and $\phi = -75^\circ$, $\psi = 153^\circ$ for the second Pro or Hyp. The definition²⁹ used with crystallographic or model data to identify collagen type helices is $\phi = -65^\circ \pm 10^\circ$, $\psi = 145^\circ \pm 30^\circ$. In the hydrophilic segment, 38% of the residues had conformations that met this definition, 57% of those in the hydrophobic segment did. Considering only those residues that are not Gly, Pro or Hyp, 28% of the hydrophilic segment and 33% of the hydrophobic segment met the definition for a collagen type helical structure. These data are most readily illustrated by Ramachandran plots³⁰ (Figure 4), where it can be seen that points representing the hydrophobic segment (Figure 4a) are more tightly clustered around the X-ray defined value (-65° , 145°) than those representing the hydrophilic segment (Figure 4b).

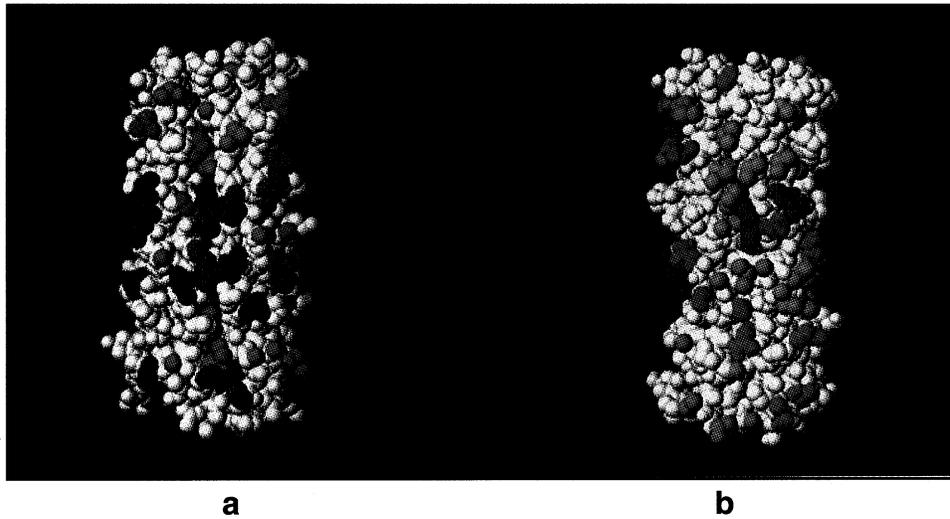


FIGURE 3. — (a) Enlarged view of a 281 residue segment of the microfibril rich in hydrophilic residues, particularly charged residues. (b) Enlarged view of a 281 residue segment of the microfibril rich in hydrophobic residues. The coloring scheme is the same as in Figure 1.

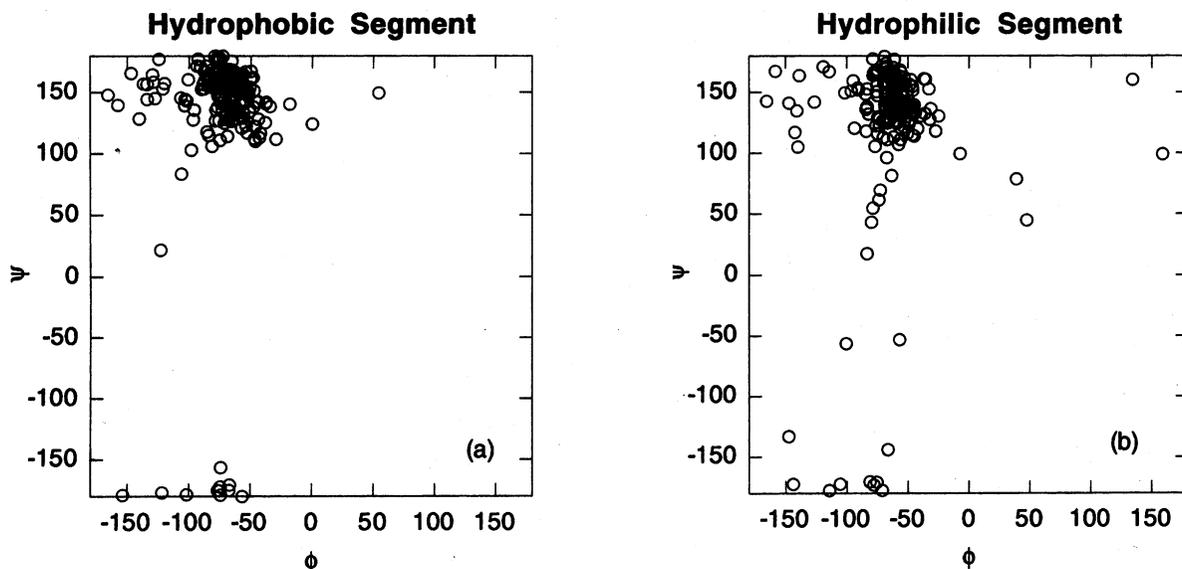


FIGURE 4. — Ramachandran plots from the (a) hydrophobic and (b) hydrophilic segments of the microfibril model.

The overall dimensions of the model segments were also measured and the data are summarized in Table I. Although the hydrophilic and hydrophobic segments excised from the microfibril contain 281 residues each, the backbone structures are not identical. The diameter of a cross section of the hydrophilic segment, including side chains, is on average 10% greater than that of the hydrophobic segment. Ignoring the contribution of side chains, the diameter of a cross section of the backbone of the hydrophilic segment is 20% greater than that of the hydrophobic segment. The greater variation in the measurements of the hydrophilic segment is consistent with the lower amount of helical

character in this segment. The hydrophobic segment was about 10% longer than the hydrophilic segment, so that the total volumes occupied by the two segments are virtually identical. These results are consistent with the idea that the hydrophobic effect makes a large contribution to the stabilization of the triple helix and microfibril structures.³¹ They are also consistent with a relatively open structure in the hydrophilic regions where mineral tannages are taken up. The recent development of tanning with duets (complexes of metals with vegetable tannins)³² is most likely a mechanism for stabilizing the relative positions of hydrophobic and hydrophilic regions of collagen.

TABLE I
Dimensions of Microfibril Segments^a

	Hydrophilic	Hydrophobic	HL/HB ^b
Diameter ^c			
average	35.5	32.0	1.1
stdev	8.0	4.7	
Backbone ^d			
average	34.3	28.7	1.2
stdev	7.2	2.2	
Length			
average	52.9	59.0	0.9
stdev	2.1	2.2	
Volume ^e	20517	20667	1.0

a Each segment consists of 281 amino acid residues.

b HL/HB is the ratio of hydrophilic to hydrophobic.

c Average diameter of the segment including side chains, dimensions are in Å

d Average diameter not including side chains.

e Volume is in Å³.

For the leather manufacturer, collagen has always been a major raw material. Increasingly collagen (in the form of gelatin recovered from chrome-tanned solid wastes) is also a major byproduct of leather manufacturing. The ability to examine the collagen structure in detail and look at regions with different characteristics may be expected to lead to new chemistries for tanning or for modifications of gelatin that may increase the value of a product. The computer model used in this work will be made available upon request to researchers wishing to evaluate it for other research projects. A mini-computer or workstation with an appropriate molecular modeling software package will be necessary in order to use this model.

ACKNOWLEDGEMENTS

The authors thank Dr. Peter Cooke and Ms. Irma Sweeney for the electron micrograph.

REFERENCES

1. Ramachandran, G. N.; Structure of collagen at the molecular level. in *Treatise on Collagen* (G. N. Ramachandran ed.) Academic Press, London, pp. 103-183, 1967.

2. Smith, J. W.; Molecular Pattern in native collagen. *Nature* **219**, 157-158, 1968.
3. Carver, J. P. and Blout, E. R.; Polypeptide models for collagen. in *Treatise on Collagen* (G. N. Ramachandran ed.) Academic Press, London, pp. 441-523, 1967.
4. Miller, M. H. and Scheraga, H. A.; Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled coil conformation. I. poly(glycyl-prolyl-prolyl). *J. Polymer Sci., Polym. Symp.* **54**, 171-200, 1976.
5. Chen, J. M., Kung, C. E., Fearheller, S. H. and Brown, E. M.; An energetic evaluation of a "Smith" collagen microfibril model. *J. Protein Chem.* **10**, 535-552, 1991.
6. Chen, J. M. and Sheldon, A.; Three-dimensional energy-minimized model of human type II "Smith" collagen microfibril. in *Molecular Modeling from Virtual Tools to Real Problems*. (T. F. Kumosinski and M. N. Liebman eds.) American Chemical Society Symp. Ser. 576 pp. 139-170, 1994.
7. Ramachandran, G. N.; Molecular architecture of collagen. *JALCA* **63**, 160-191, 1968.

8. Okuyama, K., Tanaka, N., Ashida, T., Kakudo, M., Sakakibara, S. and Kishida, Y.; An X-ray study of the synthetic polypeptide (P-P-G)₁₀. *J. Mol. Biol.* **72**, 571-576, 1972.
9. Sakakibara, S., Kishida, Y., Kikuchi, Y., Sakai, R. and Kakiuchi, K.; Synthesis of poly (L-prolyl-L-prolylglycyl) of definite molecular weights. *Bull. Chem. Soc. Jpn.* **41**, 1273-1273, 1968.
10. Ramachandran, G. N. and Kartha, G.; Structure of collagen. *Nature* **174**, 269-270, 1954.
11. Okuyama, K., Tanaka, N., Ashida, T. and Kakudo, M.; Structure analysis of a collagen model polypeptide, (pro-pro-gly)₁₀. *Bull. Chem. Soc. Jpn.* **49**, 1805-1810, 1976.
12. Mould, A. P., Hulmes, D. J. S., Holmes, D. F., Cummings, C., Sear, C. H. J. and Chapman, J. A.; D-Periodic assemblies of type I procollagen. *J. Mol. Biol.* **211**, 581-594, 1990.
13. Bruns, R. R. and Gross, J.; Band pattern of the segment-long-spacing form of collagen. Its use in the analysis of primary structure. *Biochemistry* **12**, 808-815, 1973.
14. Hodge, A. J. and Petruska, A. J.; Studies with the electron microscope on ordered aggregates of the tropocollagen molecule. in *Aspects of Protein Structure*. (G. N. Ramachandran ed.) Academic Press, London, pp. 289-300, 1963.
15. Fraser, R. D. B., Miller, A. and Parry, D. A. D.; Packing of microfibrils in collagen. *J. Mol. Biol.* **83**, 281-283, 1974.
16. Veis, A. and Yuan, L.; Structure of the collagen microfibril. A four-strand overlap model. *Biopolymers* **14**, 895-900, 1975.
17. Piez, K. A. and Trus, B. L.; Microfibrillar structure and packing of collagen: Hydrophobic interactions. *J. Mol. Biol.* **110**, 701-704, 1977.
18. Piez, K. A. and Trus, B. L.; Sequence regularities and packing of collagen molecules. *J. Mol. Biol.* **122**, 419-432, 1978.
19. Traub, W.; Molecular assembly in collagen. *FEBS Lett.* **92**, 114-120, 1978.
20. Hofmann, H., Fietzek, P. P. and Kuhn, K.; Role of polar and hydrophobic interactions for the molecular packing of type I collagen: A 3-D evaluation of the amino acid sequence. *J. Mol. Biol.* **125**, 137-165, 1978.
21. Okuyama, K., Go, H. and Takayanagi, M.; Analysis of microfibrillar structure of collagen based on $\alpha_1(I)$ chain sequence and 7/1 helix model of tropocollagen. *Chem. Lett.*, 509-512, 1978.
22. Fraser, R. D. B., MacRae, T. P. and Suzuki, E.; Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463-481, 1979.
23. Fraser, R. D. B., MacRae, T. P., Miller, A. and Suzuki, E.; Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167**, 497-521, 1983.
24. Rich, A. and Crick, F. H. C.; The molecular structure of collagen. *J. Mol. Biol.* **3**, 483-506, 1961.
25. Chen, J. M., Fairheller, S. H. and Brown, E. M.; Three-dimensional energy-minimized models for calf skin type I collagen triple helix and microfibril: I. The triple helical model. *JALCA* **86**, 475-486, 1991.
26. Chen, J. M., Fairheller, S. H. and Brown, E. M.; Three-dimensional energy-minimized models for calf skin type I collagen triple helix and microfibril: II. The microfibril model. *JALCA* **86**, 487-497, 1991.
27. King, G., Brown, E. M. and Chen, J. M.; Computer model of a bovine type I collagen microfibril. *Protein Engineering* **9**, 43-49, 1996.
28. Weiss, T. J., Miller, A. and Bradshaw, J. P.; Cross-linkage sites in type I collagen fibrils studied by neutron diffraction. *J. Mol. Biol.* **213**, 1-5, 1990.
29. Garnier, J. and Robson, B.; The GOR method for predicting secondary structures in proteins. in *Prediction of Protein Structure and the Principles of Protein Conformation* (G. D. Fasman ed) Plenum Press, New York, pp. 417-465, 1989.
30. Ramachandran, G. N., Sasisekharan, V.; Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283-437, 1968.
31. Almela, M., Maldonado, F., Otero, A. and Costalopez, J.; The hydrophobic effect and its importance to collagen. *J. Soc. Leather Technol. Chem.* **76**, 157-161, 1992.
32. Lampard, G.; Vegetable mineral duets. *Leather*, 47-48, 1996.