

An Energetic Evaluation of a "Smith" Collagen Microfibril Model

James M. Chen,^{1,2} Chun E. Kung,¹ Stephen H. Fearheller,¹ and Eleanor M. Brown¹

Received May 23, 1991

An energy minimized three-dimensional structure of a collagen microfibril template was constructed based on the five-stranded model of Smith (1968), using molecular modeling methods and Kollman force fields (Weiner and Kollman, 1981). For this model, individual molecules were constructed with three identical polypeptide chains ((Gly-Pro-Pro)_n, (Gly-Prop-Hyp)_n, or (Gly-Ala-Ala)_n, where $n=4, 12,$ and 16) coiled into a right-handed triple-helical structure. The axial distance between adjacent amino acid residues is about 0.29 nm per polypeptide chain, and the pitch of each chain is approximately 3.3 residues. The microfibril model consists of five parallel triple helices packed so that a left-handed superhelical twist exists. The structural characteristics of the computed microfibril are consistent with those obtained for collagen by X-ray diffraction and electron microscopy. The energy minimized Smith microfibril model for (Gly-Pro-Pro)₁₂ has an axial length of about 10.2 nm (for a 36 amino acid residue chain), which gives an estimated D-spacing (234 amino acids per chain) of approximately 66.2 nm. Studies of the microfibril models (Gly-Pro-Pro)₁₂, (Gly-Pro-Hyp)₁₂, and (Gly-Ala-Ala)₁₂ show that nonbonded van der Waals interactions are important for microfibril formation, while electrostatic interactions contribute to the stability of the microfibril structure and determine the specificity by which collagen molecules pack within the microfibril.

KEY WORDS: Triple helix; fibril; coiled-coil; telopeptide; cross-linking, molecular modeling.

1. INTRODUCTION

1.1. Historical Review

There are to date at least 12 known types of collagen and of these, Types I, II, and III are known as the fiber-forming collagens (Martin *et al.*, 1985; Piez, 1984; Miller, 1985; Gordon *et al.*, 1990). These three are the major structural constituents found in skin, cartilage, bone, blood vessel walls, and internal organs. Type I collagen is the most abundant and its structure has been widely studied (Piez, 1984; Chapman and Hulmes, 1984; Miller, 1976; Chapman, 1984).

Structurally, this semiflexible, rod-like molecule is approximately 300 nm in length and 1.0–1.4 nm in

diameter, depending upon the hydrated state of the protein. Type I collagen consists of two $\alpha 1$ chains and one $\alpha 2$ chain wound into a right-handed triple helix. Each of the left-handed helical polypeptide chains is composed of Gly-X-Y tripeptides, where X and Y may be any amino acid residue. Although the $\alpha 1$ and $\alpha 2$ peptide sequences consist of many different types of amino acid residues, 33% are glycine and 25% are imino acids: proline and hydroxyproline (Ramachandran and Ramakrishnan, 1976). Excluding the extrahelical terminal peptides or telopeptides, there are in total 1014 amino acid residues per polypeptide chain, such that the consensus sequence of Gly-X-Y is repeated 338 times (Piez, 1984). Types II and III, although genetically distinct, have similar physical characteristics (Martin *et al.*, 1985).

Collagen exhibits a high degree of polymorphism (Piez, 1984; Chapman and Hulmes, 1984; Brodsky and Eikenberry, 1984). Since collagen is a highly ubiquitous and multipurpose structural protein, it is

¹ U.S. Department of Agriculture, ARS, Eastern Regional Research Center, 600 E. Mermaid Lane, Philadelphia, Pennsylvania 19118.

² To whom all correspondence should be addressed.

able to form a diverse range of fibrillar structures *in vivo*. In addition, *in vitro* collagen has been observed to form structures such as segment long spacing crystals, fibrillar long spacing aggregates, obliquely banded fibrils, and nonbanded fibrils. Many of these structures have been examined using X-ray diffraction, electron microscopy, and freeze fracture analyses (Chapman and Hulmes, 1984; Brodsky *et al.*, 1982; Chew and Squire, 1986; Eikenberry and Brodsky, 1980). From these investigations, information on the three-dimensional structure of collagen fibrils has been obtained.

The axial arrangement of collagen molecules within the fibril has been studied extensively and is known to be highly ordered (Piez, 1984; Chapman and Hulmes, 1984; Chapman, 1984). In 1942, Bear, using X-ray diffraction, detected the appearance of regular crosswise striations along the length of the collagen fibril. This pattern, which can be observed using both negative and positive staining, was later labeled by Schmitt and Gross (1948). The banding also occurs with a distinct axial periodicity and is polarized in the direction of the collagen molecules in the fibril (Piez, 1984; Chapman and Hulmes, 1984). The value of the axial period (or D-period) depends upon the tissue type of the fibrils and thus ranges from 60–68 nm. However, accepted values tend to be closer to 67 nm (Woodhead-Galloway, 1984; Meek *et al.*, 1979). Given the axial distance between each adjacent amino acid residue to be about 0.29 nm, there are 234 residues per D-period.

The lateral packing of the collagen molecules within the fibrils remains unclear (Piez, 1984; Miller, 1976; Chapman, 1984; Woodhead-Galloway, 1984; Galloway, 1984). It appears from X-ray diffraction and electron microscopy analyses that the order of the packing may be dependent on the type and function of the tissue (Brodsky and Eikenberry, 1985; Brodsky *et al.*, 1982). There may also be some effects from the preparation of the sample prior to analysis (Ripamonti *et al.*, 1980). Some studies have shown the lateral packing to have crystalline properties (Miller and Wray, 1971; Miller and Parry, 1973; Parry and Craig, 1979; Squire and Freundlich; 1980). This has resulted in the development of the five-stranded helical microfibril originally proposed by Smith (1968) and the four stranded microfibril introduced by Veis and Yuan (1975) (among others: Traub, 1978; Piez and Trus, 1977, 1978; Fraser *et al.*, 1974; Hofmann *et al.*, 1978; Okuyama *et al.*, 1978). The collagen molecules in these structures are related by a 1D stagger.

Both models are able to explain the major pattern of repeated light and dark regions obtained from negative staining of transverse fibril sections. The dark regions were called "gap" regions since they were domains of low density molecular packing as had been noted by Hodge and Petruska (1963). Therefore, it was proposed that no end-to-end interactions occur between adjacent collagen molecules along the same vertical axis. The separation between the molecules was estimated to be approximately 0.6D or 140 residues (Hodge and Petruska, 1963; Smith, 1968). Conversely, the light regions resulted from denser lateral packing within the microfibril due to the overlap arrangement of adjacent collagen molecules. The length of the overlap region was about 0.4D (Hodge and Petruska, 1963; Smith, 1968). Both models emphasize the rope-like structure of the microfibril and fibril. The Smith microfibril would have an overall left-handed supercoil of pitch 20D/11 (i.e., between 115–200 nm) (Traub, 1978; Piez and Trus, 1977, 1978). The microfibril unit has been shown to exist *in vitro* and is postulated to be the intermediate step prior to fibril formation (Veis *et al.*, 1979; Na *et al.*, 1986a, b).

On the other hand, it has been inferred from other microscopy studies that the lateral packing is noncrystalline and more liquid-like (Woodhead-Galloway, 1984; Giraud-Gille, 1987). This has led to proposed models such as the octafibril and the simple two-dimensional liquid model (Woodhead-Galloway, 1984; Hosemann *et al.*, 1974). These models regard the fibril as having no intermediate substructure. ^2H and ^{13}C NMR studies have shown that there is significant mobility in the intermolecular interactions between collagen molecules, supporting the plausibility of this fluid property (Torchia and Vanderhart, 1976; Torchia *et al.*, 1985; Jelinski *et al.*, 1980).

There are also models which consider the fibril as having more crystalline than liquid properties (Miller, 1982; Piez and Trus, 1981; Hulmes and Miller, 1981). The first one is the quasihexagonal molecular crystal model, which was developed to explain the observed X-ray diffraction patterns from tendon samples (Miller, 1982; Hulmes and Miller, 1981; Hulmes *et al.*, 1981). Treating the fibril as a group of cylindrical molecules packed in a lateral array, it is logical to assume that they would be hexagonally close packed. However, this structure does not give rise to the correct X-ray reflections. By modifying the lattice spacings and tilting the collagen molecules by 4–5° to the fibril axis, an optimal quasi-hexagonal packing was obtained which gives the desired X-ray diffraction

“Smith” Collagen Microfibril Model

pattern. The other model is the compressed microfibril model (Piez and Trus, 1981). This model depicts the five stranded microfibril as being laterally compressed. This distortion results in the new microfibril occupying a unit cell similar to that for a quasi-hexagonal crystal lattice.

A possible description of fibrils would show them as having a superhelical right-handed twist. Due to the dense packing of collagen molecules, the center of the fibril would have liquid-like properties, but crystalline features would become more apparent near the periphery of the fibril (Chapman, 1984; Hulmes *et al.*, 1981). This is supported by recent studies on Type I collagen (Chew and Squire, 1986; Lees *et al.*, 1984; Hulmes *et al.*, 1985). However, the results from other X-ray diffraction studies on Type I collagen cannot be explained fully by any of the above models (Fraser *et al.*, 1983, 1987).

1.2. Objectives

At present there are no energy minimized models for the three-dimensional structure of collagen that describe inter- and intrafibrillar interactions. These inter- and intrafibrillar interactions between collagen molecules are the bases for cross-linking in fibrils, and therefore govern the strength and flexibility of the collagen fibers. In addition, such interactions are important in developing ligand-binding sites on collagen for both biological and synthetic reagents.

Given the present-day tools for molecular modeling of protein structures, it is possible to develop a full three-dimensional model for the interactions within and between collagen molecules. Although it is clear that X-ray diffraction and electron microscopy have not provided conclusive evidence of how collagen molecules pack within fibrils, it is hoped that the three-dimensional modeling of collagen will improve our understanding of the possibilities by which these biomacromolecules can interact.

The study presented here describes a prototype three-dimensional model for collagen and collagen interactions based on the “Smith microfibril,” in which five parallel collagen molecules are packed together in a circular array. This model, where the molecules are arranged in a symmetrical fashion, could provide evidence for the specific interactions which occur between the side chains of amino acid residues involved in stabilizing lateral packing (Smith, 1968; Okuyama *et al.*, 1978). Although earlier studies gave evidence of distinct regions of polar and nonpolar interactions which may occur between

adjacent molecules (Piez and Trus, 1977; Hofmann *et al.*, 1978; Trus and Piez, 1976), it is of interest to explore with this three-dimensional model of the collagen microfibril how these interactions may be possible. Our studies examine how specific amino acid interactions affect the three-dimensional structure of the “Smith” microfibril model. Three-dimensional models consisting of the collagen-like sequences (Gly-Pro-Pro), (Gly-Pro-Hyp) (where Hyp represents the modified imino acid, hydroxyproline), and (Gly-Ala-Ala) are analyzed in terms of their energetics and compared to each other. The importance of proline and hydroxyproline in stabilizing the collagen triple helix and microfibril is examined. In addition, the (Gly-Ala-Ala) models for both the triple helix and microfibril are studied to further demonstrate the importance of proline in stabilizing collagen structures. Finally, the uses of our microfibril model will be discussed.

2. METHOD

Molecular modeling of the collagen triple helix and microfibril structures was performed on an Evans and Sutherland PS390 graphics workstation interfaced with a VAX 8350 minicomputer (Digital Corp.). The molecular modeling software used was SYBYL (v5.32) developed by TRIPOS Associates, Inc. (1990). SYBYL contains a set of functions for building and optimizing biopolymers. Energy refinement of protein structures is based on a molecular mechanics method in SYBYL. The algorithms used for the minimization of large structures are based on a combination of the simplex and conjugate gradient methods. Here, the total energy (E_{tot}) function to be minimized is composed of the sum of several energy terms as given in Eq. (1).

$$E_{tot} = E_{bs} + E_{ab} + E_{op} + E_{tor} + E_{vdw} + E_e + E_{14vdw} + E_{14e} + E_{hb} \quad (1)$$

where E_{bs} is the sum of energies arising from bond stretching or compression beyond the optimum bond length; E_{ab} is the sum of energies for angles which are distorted from their optimum values; E_{op} is the sum of energies for the bending of planar atoms out of the plane; E_{tor} is the sum of the torsional energies which arise from rotations about each respective dihedral angle; E_{vdw} is the sum of energies due to nonbonded van der Waals interactions; E_e is the sum of the nonbonded electrostatic interaction energies; E_{14vdw} and E_{14e} are the sum of energies due to van der Waals and

electrostatic interactions, respectively, for atoms connected by three bonds; and E_{hb} is the sum of energies due to hydrogen bond interactions. Nonbonded van der Waals and electrostatic interactions were not considered beyond a cutoff distance of 8 Å. Using the united atoms approach, both the 1–4 van der Waals and electrostatic interactions for the collagen models were reduced by a factor of 0.5 in accordance with Weiner *et al.* (1984). Solvent molecules were not explicitly included in the models but a distance-dependent dielectric function, $\epsilon = (R_{ij} + 1)$ (where R_{ij} is the distance between atom i and atom j) was used to implicitly account for the effect of solvent, as all minimizations were carried out *in vacuo*. In addition, all minimizations were performed given a root-mean square (rms) derivative of 0.01 kcal/mol-Å as a cutoff value.

2.1. The Collagen Model

The initial step in the construction of the “Smith” microfibril model was to build a single polypeptide chain of (Gly-Pro-Pro)₄, using the parameters for each amino acid residue provided by the dictionary component of the software package. Short polypeptide segments were used to simplify the manipulation and docking procedures necessary for the construction of the collagen triple helix, where the intermolecular interactions of three polypeptide chains [e.g., three (Gly-Pro-Pro)₄] must be accommodated. Molecular modeling of the collagen models was simplified since each single chain consists of a repeat of a specific tripeptide sequence: (Gly-Pro-Pro), (Gly-Pro-Hyp), or (Gly-Ala-Ala). Each triple helix model contains a threefold (C_3) rotational symmetry about its helical axis, and each microfibril model contains a C_5 rotational symmetry about its longitudinal microfibril axis.

2.1.1. Peptide Backbone Torsional Angles

In this work, we have used the values of the peptide backbone torsional angles: ϕ , ψ , and ω reported by Miller and Scheraga (1976). These angles for (Gly-Pro-Pro) correspond to the lowest energy structure for the triple helix of (Gly-Pro-Pro)₄ as obtained from ECEPP (Empirical Conformational Energy for Proteins and Peptides) (Scheraga, 1984; Miller and Scheraga, 1976). Miller and Scheraga (1976) computed all the possible minimum energy conformations for (Gly-Pro-Pro)₄, considering both the *cis* and *trans* peptide bonds for proline residues. Many low energy

conformations were found, indicating that the single chain was quite flexible. All the minimum conformations for a polypeptide chain were combined when forming conformations for triple-stranded complexes, which were then further minimized. Unlike the single chain of collagen, the interaction of three chains formed a well-defined three-dimensional structure. The resultant global minimum conformation found was 18.8 kcal/mol more stable than the next minimum energy conformation of the triple helix. Subsequently, this conformation was shown to have a root mean square (rms) deviation of 0.3 Å (for all atoms except hydrogen atoms) when compared to a crystal structure of a triple-helical complex consisting of three synthetic (Gly-Pro-Pro)₁₀ chains (Miller and Scheraga, 1976).

2.1.2. The Triple Helix

The procedure for modeling the triple helix was as follows. First, the above torsional angles were applied to the polypeptide chain of (Gly-Pro-Pro)₄. Next, each polypeptide chain was blocked with a N-acetyl group at the amino terminus and a N-methylamide group at the carboxyl terminus to account for end effects from neighboring residues. The structure was then energy-minimized, using the Kollman force field, AMBER (Weiner and Kollman, 1981; Weiner *et al.*, 1984). Root mean square deviations between the initial and the minimized structures were then computed (rms < 0.8 Å for all atoms) and showed that they were essentially the same. Three minimized polypeptide chains of (Gly-Pro-Pro)₄ were interactively docked on the PS390 workstation so that the resulting structure was a triple helix. To accomplish this, different colors were assigned to pairs of interacting groups on the three polypeptide chains. It was then a simple process to visually dock them by bringing together all the respective colored pairs. If necessary, distance “range” constraints (SYBYL, v5.32, 1990) could be assigned to each interacting pair, such as between the oxygen of a backbone carbonyl group and the hydrogen of an amide nitrogen. For example, minimizing while constraining all the paired hydrogen-bonding interactions to remain within a given range, such as 1.70–2.30 Å.

Once the unit structure, the triple helix of a terminally blocked 3(Gly-Pro-Pro)₄, had been constructed and energy-minimized, each helical unit was docked end-to-end (i.e., carboxyl to amino terminus) in order to extend the structure. To do this, all N- and C-terminal end groups were first removed. Peptide

"Smith" Collagen Microfibril Model

bonds were then formed between neighboring polypeptide chains, thereby linking together the triple-helical segments. Prior to energy minimization of the final structure, N-acetyl and N-methylamide end groups were reassigned to the respective ends of each open polypeptide chain. For collagen, it is noted that the energy refinement of a large structure consisting of smaller previously energy-minimized structures is much more efficient than the refinement of a large unminimized structure.

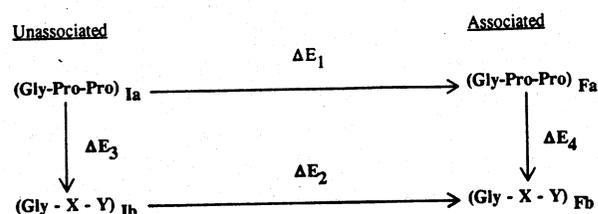
2.1.3. The Microfibril

The model of the collagen microfibril complex was constructed by the docking of five triple helices in accordance with the Smith (1968) microfibril model. The initial microfibril was constructed from the packing of two triple-helical segments of $3(\text{Gly-Pro-Pro})_{12}$, a choice which allowed us to utilize the results of Nemethy and Scheraga (1984) for the packing of two $(\text{Gly-Pro-Pro})_5$ triple helices. Initially, the interacting regions found by Nemethy and Scheraga (1984) for collagen packing were "highlighted" in color on the PS390 workstation. Two helices were then docked by pairing the colored regions, which represented the lowest energy packing of two collagen helices. The dimer was then relaxed through energy minimization. The same procedure was followed for the packing of the third, fourth, and fifth $3(\text{Gly-Pro-Pro})_{12}$ helix. Once the microfibril was assembled, the tripeptide sequences— $3(\text{Gly-Pro-Hyp})_{12}$ or $3(\text{Gly-Ala-Ala})_{12}$ —were substituted for $3(\text{Gly-Pro-Pro})_{12}$ and energy-minimized.

The positioning of the above structures was accomplished interactively and visually on the Evans and Sutherland PS390 system. All atoms of each structure were constrained to their original positions prior to docking and energy minimization. This step insures that the best possible packing or interaction occurs between molecules in their starting conformations. The SYBYL function referred to as DEFINE AGGREGATE will constrain all the atoms of a given structure from changing (SYBYL, v5.32, 1990). This AGGREGATE constraint was removed after the packing positions were optimized for each substructure polypeptide chain and the completed structure then underwent further energy refinement. This basic procedure was also used to construct the microfibril model from the initial values for the backbone dihedral angles of the collagen polypeptide chain, $(\text{Gly-Pro-Pro})_4$.

2.2. Comparison of the Computed Potential Energies for the Stabilization of the Collagen Triple Helix and Microfibril

The objective of the initial modeling work was to construct an energy-minimized triple-helical structure of the $(\text{Gly-Pro-Pro})_4$ collagen model. This initial triple-helical structure is similar to the global minimum energy complex obtained from the interaction of all low energy "allowable" conformations of the three polypeptide chains of $(\text{Gly-Pro-Pro})_4$ (Miller and Scheraga, 1976). This basic triple-helical unit was used to construct the $(\text{Gly-Pro-Pro})_{12}$ and $(\text{Gly-Pro-Pro})_{16}$ structures. The $(\text{Gly-Pro-Pro})_{12}$ structure then formed the base for the microfibril model. The triple helix and microfibril structures of $(\text{Gly-Pro-Hyp})_{12}$ and $(\text{Gly-Ala-Ala})_{12}$ were obtained through substitution of the corresponding amino acids into the minimized structures for $(\text{Gly-Pro-Pro})_{12}$ as shown in



Scheme I. Thermodynamic cycle for comparing relative stabilization energies.

Scheme I, where both Ia and Ib = initial polypeptide chain or initial triple helix structures, and both Fa and Fb = final triple helix or final microfibril structures. X and Y are the substituted amino acids. ΔE_1 and ΔE_2 are the potential energy differences between the energy of the "associated" collagen complex and the sum of energies of the "unassociated" structures which make up the complex. In the above scheme, "unassociated" simply refers to the state where each specific structure (e.g., a polypeptide chain in a triple-helical complex or a triple helix in a microfibril complex) within the complex is considered independently, not influenced by the other chains or helices.

For the triple-helical complex, ΔE_1 is defined as the difference in potential energy between that of a single collagen triple helix and the sum of the energies for the three specific polypeptide structures which form the triple helix. In similar fashion, ΔE_1 for the microfibril is defined as the difference in energy between that of a single microfibril and the sum of energies for five specific triple helices which make up the microfibril complex. ΔE_2 is the same as ΔE_1 except

that it is defined for the modified sequences, (Gly-Pro-Hyp) and (Gly-Ala-Ala).

The above schematic of a thermodynamic cycle is shown in order to compare the potential energies of the minimized structures for the triple helices and microfibrils obtained from the substitution of different tripeptide sequences. In our case, ΔE_3 and ΔE_4 would represent the potential energy differences between the energies of the original and modified structures. As a result, $\Delta\Delta E = (\Delta E_4 - \Delta E_3)$ would denote the relative stabilization of Fb as compared to Fa (see Scheme I).

3. RESULTS AND DISCUSSION

3.1. Molecular Modeling of the Collagen Triple-Helical Structure

The three-dimensional structure of a (Gly-Pro-Pro)₄ polypeptide chain was constructed using the SYBYL (v5.32, 1990) Molecular Modeling software and energy minimized using a molecular mechanics method provided by SYBYL which incorporates the Kollman force fields (Weiner and Kollman, 1981; Weiner *et al.*, 1984). The average values of the backbone dihedral angles for the minimized (Gly-Pro-Pro)₄ are given in Table I(A), as are the initial values

used to construct the polypeptide chain. It is evident that the polypeptide structures changed little after energy refinement (i.e., superposition between the starting and final structure results in an rms deviation of less than 0.8 Å for all atoms). Figure 1 shows the structure for the (Gly-Pro-Pro)₄ polypeptide chain, where the proline rings are colored green. The conformation of this peptide segment is a left-handed helical coil. The axial height between adjacent residues is approximately 0.29 nm for the energy-refined polypeptide structure (Fig. 1). The measured helical pitch for the computed polypeptide chain is about 3.3 residues.

The above (Gly-Pro-Pro)₄ structure was then used to construct the collagen triple helix which is known from X-ray diffraction data to consist of three individual polypeptide chains coiled into a right-handed helical twist. The computed (Gly-Pro-Pro)₄ structure was docked (see Methods) with two other identical chains of (Gly-Pro-Pro)₄ to reproduce the "one-banded" collagen model as proposed by Rich and Crick (1955, 1961) and as derived from conformational energy analysis (Miller and Scheraga, 1976). This model allows for a single hydrogen bond to form between a backbone carbonyl oxygen of one chain and the amide hydrogen of an adjacent polypeptide

Table I. Backbone Dihedral Angles^a

Chain	ϕ_1^b	Ψ_1	ω_1	ϕ_2	Ψ_2	ω_2	ϕ_3	Ψ_3	ω_3
A. For the polypeptide (Gly-Pro-Pro) ₄ ^c									
Initial chain	-74.0	170.0	180.0	-75.0	168.0	180.0	-75.0	153.0	-180.0
Final chain	-74.0	165.2	179.0	-72.3	162.0	179.0	-71.5	155.7	-179.0
B. For the triple helix ^d									
Gly-Pro-Pro	-77.5	170.0	178.0	-72.0	163.2	178.0	-69.2	156.2	-178.0
Gly-Pro-Hyp	-82.3	166.7	178.0	-72.4	155.7	178.0	-58.0	159.1	-178.0
Gly-Ala-Ala	-80.2	173.5	179.0	-73.0	161.7	179.0	-70.6	151.6	-179.0
C. For the microfibril ^e									
Gly-Pro-Pro	-77.8	163.2	180.0	-70.3	155.0	180.0	-60.0	148.0	180.0
Gly-Pro-Hyp	-80.2	158.9	179.0	-70.9	150.6	179.0	-59.0	154.0	179.0
Gly-Ala-Ala	-80.8	172.0	179.0	-74.0	153.3	179.0	-67.9	145.5	179.0

^a The polypeptide chain of collagen consists of the tripeptides (Gly-X-Y). The values for the given dihedral angles are averaged for each respective position of the tripeptide consensus sequence. The angles listed are in degrees and have a standard error of $\pm 0.05^\circ$.

^b The subscripts (i.e., 1, 2, and 3) refer to the conformation of each amino acid residue in the order corresponding to the tripeptide sequence, (Gly-X-Y). ϕ is the dihedral angle where rotation occurs about the peptide bond N-C α . Ψ is the dihedral angle where rotation occurs about the peptide bond C α -C'. ω is the dihedral angle where rotation occurs about the peptide bond C'-N, but this bond is constrained close to 180°.

^c The minimization conditions for the initial chain are given in Miller and Scheraga (1976) but for the final chain, see Methods.

^d The triple helix is composed of 3(Gly-X-Y)₁₆.

^e The (Gly-Pro-Pro)₁₂ microfibril consists of five triple helices packed as proposed by Smith (1968). This collagen complex was then energy-refined using AMBER (Weiner and Kollman, 1981). The corresponding microfibril structures for (Gly-Pro-Hyp)₁₂ and (Gly-Ala-Ala)₁₂ were derived by substituting these sequences into the energy-refined (Gly-Pro-Pro)₁₂ model and energy minimizing again. These are the measured averaged angles for each model.

“Smith” Collagen Microfibril Model

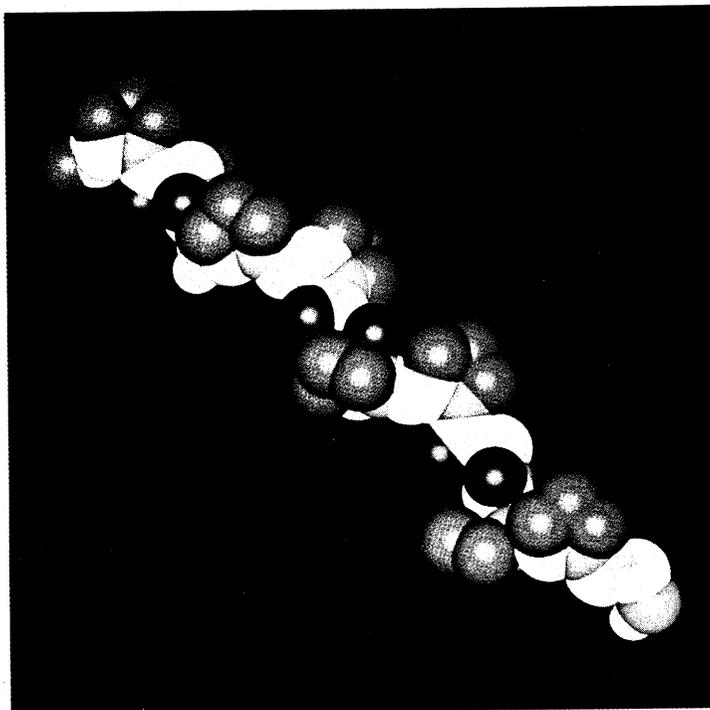


Fig. 1. A colored space-filling diagram of the $(\text{Gly-Pro-Pro})_4$ polypeptide chain after structural minimization, where the proline rings are colored green. The structure has a left-handed helical twist. The axial length between each adjacent amino acid residue is about 0.29 nm and the pitch is about 3.3 residues per turn. The amino-terminus is in the foreground.

chain. The carbonyl group involved is at the “X” position of the Gly-X-Y collagen tripeptide and the hydrogen donor is the backbone amide-nitrogen of glycine, which is always in the first position of the above tripeptide.

Figure 2 is a stereo figure showing the collagen-like triple helix after energy refinement and removal of the N- and C-terminal end groups. The dashed lines depict the single backbone hydrogen bond per tripeptide which forms between adjacent chains. The triple-helical structure for $(\text{Gly-Pro-Pro})_4$ was extended by joining together the N-termini and C-termini of four $(\text{Gly-Pro-Pro})_4$ units to form $(\text{Gly-Pro-Pro})_{16}$ (Fig. 3A). It is clear in Fig. 3A that the collagen structure has a right-handed helical twist. The pitch for the computed collagen helix is 27 amino acid residues per polypeptide chain, as indicated by the arrows in Fig. 3B, and corresponds to an axial length of approximately 7.8 nm. The resultant D spacing (234 residues per polypeptide chain) for the (Gly-Pro-Pro) collagen triple helix model is approximately 67.4 nm, close to the accepted value (Woodhead-Galloway, 1984; Meek *et al.*, 1979). The radius of this structure is approximately 0.52 nm. The average

values of the backbone dihedral angles for the minimized triple-helical structures of (Gly-Pro-Pro) , (Gly-Pro-Hyp) , and (Gly-Ala-Ala) are given in Table I(B).

3.1.1. Evaluation of ΔE_1 , ΔE_3 , ΔE_4 , and $\Delta\Delta E$ Energies for the Computed Triple Helix Models

Potential energy differences, ΔE_1 for the triple helix, $(\text{Gly-Pro-Pro})_n$, where $n=4, 12$, or 16, are listed in Table II(A). Values are shown for each of the energy terms of Eq. (1). Energy differences resulting from the substitution of hydroxyproline or alanine into single chains (ΔE_3) or the triple helices (ΔE_4) of $(\text{Gly-Pro-Pro})_n$ are summarized in parts (B) and (C) of Table II. The $\Delta\Delta E$ values in Table II(D) were computed from: $\Delta\Delta E = \Delta E_4 - \Delta E_3$ and represent the stabilization energies of the triple helix forms of $(\text{Gly-Pro-Hyp})_n$ and $(\text{Gly-Ala-Ala})_n$ relative to the reference structure $(\text{Gly-Pro-Pro})_n$. The $\Delta\Delta E$ values for $(\text{Gly-Pro-Pro})_n$ are given as zeroes since the reference $(\text{Gly-Pro-Pro})_n$ structures being related to are unmodified.

As noted earlier, about 25% of collagen consists of the amino acid residues proline and hydroxyproline (Ramachandran and Ramakrishnan, 1976). Type I

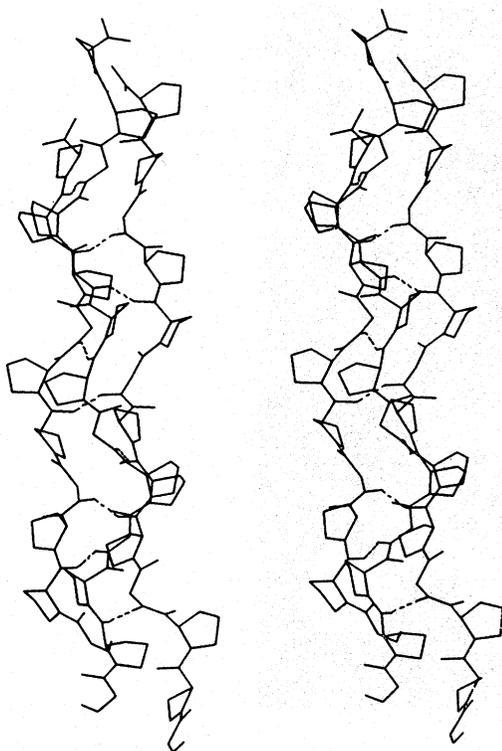


Fig. 2. A stereo figure (in the relaxed viewing mode) of the (Gly-Pro-Pro)₄ triple helix after minimization. The dashed lines indicate the single hydrogen bond which is formed per tripeptide unit in collagen. This interaction is between the amide hydrogen of Gly and the carbonyl group of the backbone of Pro (X-position) of the adjacent polypeptide chain. The measured length of the hydrogen bond is between 1.83–1.84 Å.

collagen contains several adjacent tripeptide sequences with hydroxyproline in the “Y” position (i.e., residues 865–876 and 1000–1011) (Fietzek and Kuhn, 1976). The stabilizing effect of hydroxyproline has been observed in the higher melting temperature for poly (Gly-Pro-Hyp) ($T_m = 58^\circ\text{C}$) as compared to that for poly (Gly-Pro-Pro) ($T_m = 24^\circ\text{C}$) (Berg and Prockop, 1973). In order to determine how hydroxyproline contributes to the stabilization of the collagen structure, the minimized triple-helical structure of (Gly-Pro-Pro)_n was substituted with hydroxyproline at all “Y” positions and energy minimized. The resulting triple helices were compared to the corresponding starting conformations giving rms deviations less than 0.6 Å for all backbone atoms. Comparison of the relative energies [$\Delta\Delta E$ in Table II(D)] for (Gly-Pro-Hyp)_n with the corresponding (Gly-Pro-Pro)_n show that the overall energetics are more favorable for (Gly-Pro-Hyp)_n. However, considering the fact that each energy term corresponds to a multiple of n (i.e.,

$n = 4, 12,$ and 16) tripeptide sequences, the $\Delta\Delta E_{tot}$ for (Gly-Pro-Hyp)_n is only slightly more favorable when compared to (Gly-Pro-Pro)_n.

3.1.2. Evaluation of Energies Computed for the (Gly-Pro-Hyp)₁₂ Triple Helix Model

Overall, the three-dimensional structure of the (Gly-Pro-Hyp)₄ triple helix (Fig. 3C) is similar to the (Gly-Pro-Pro)₄ triple helix (Fig. 2) but the radius is slightly larger, 0.63 nm as compared to 0.52 nm. The computed (Gly-Pro-Hyp)₄ triple-helical structure shows no evidence of hydrogen bond formation between the hydroxyl group of hydroxyproline and any backbone carbonyl oxygen atoms of the adjacent polypeptide chain (Fig. 3C). This finding is in agreement with the report of Miller *et al.* (1980) that the hydroxyl group of hydroxyproline does not appear to contribute to the stability of the collagen triple helix through interpolypeptide hydrogen bonding. Furthermore, the triple-helical structure for (Gly-Pro-Hyp) obtained using ECEPP (Scheraga, 1984) showed no interpeptide hydrogen bonds due to the hydroxyl group of hydroxyproline. However, the $\Delta\Delta H_{hb}$ (hydrogen bonding energy) values for (Gly-Pro-Hyp)_n ($n = 12$ and 16) do show stabilization [Table II(D)]. These favorable interactions may be due to the hydroxyprolines forming additional intrapolypeptide hydrogen bonds with its polypeptide backbone (Fig. 3C). Examination of the computed (Gly-Pro-Hyp)_n triple helices shows that some intrapolypeptide hydrogen bonds are formed between the hydroxyl group of hydroxyproline and the carbonyl oxygen of an adjacent C-terminal glycine residue (Fig. 3C; arrows indicate possible intrapolypeptide hydrogen bonding). This type of interaction may not be significant when the (Gly-Pro-Hyp)_n triple helices are packed in a microfibril complex, where the hydroxyl groups of specific hydroxyprolines are involved in hydrogen bonding between collagen molecules (see Microfibril section). These intrapolypeptide hydrogen bonds were not considered as being significant since solvent effects were not explicitly included in the minimizations. It has been suggested (Suzuki *et al.*, 1980) that water molecules may participate in forming intra- and interpolypeptide linkages between the hydroxyl groups of hydroxyprolines and the functional groups of the polypeptide backbone. Since the unhydrated collagen models do not show additional hydrogen bond stabilization due to the hydroxyl functional groups of hydroxyprolines and since the T_m of (Gly-Pro-Hyp) is higher than that for (Gly-Pro-Pro), it is likely that the

"Smith" Collagen Microfibril Model

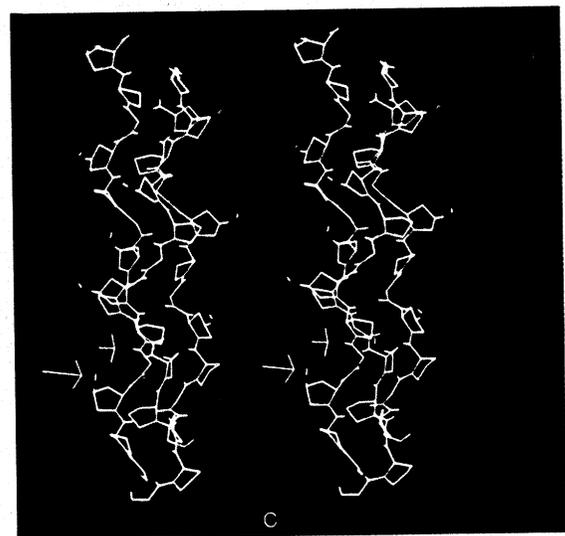
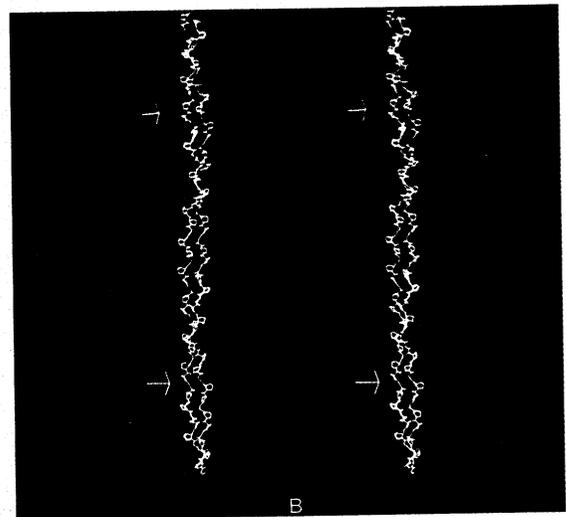
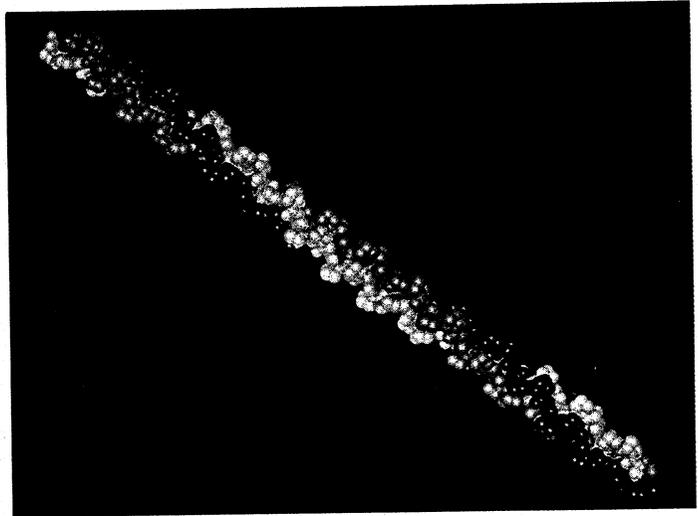


Fig. 3. (A) A space-filling model of the (Gly-Pro-Pro)_n triple helix, where $n = 16$. Each polypeptide chain is colored differently to indicate the right-handed helical twist of the collagen triple helix. The pitch of the triple helix is 27 amino acid residues, which gives an axial length of about 7.8 nm per triple-helical pitch. (B) A stereo figure of (A) (in the relaxed viewing mode) is displayed as a "stick" figure. Each polypeptide chain is shown in a different color. The arrows indicate the pitch for the collagen triple helix. (C) A stereo figure of the (Gly-Pro-Hyp)₄ triple helix. The backbone hydrogen-bonding is the same as seen in Fig. 2. In this model, the arrows indicate intrapolypeptide interactions which are observed to occur between the hydroxyl groups of hydroxyprolines and the carbonyl group of the adjacent (C-terminal) glycine residues.

Table II. Potential Energy^a Differences Calculated for the Association Steps of the Triple Helix as per Scheme

	ΔE_{bs}	ΔE_{hb}	ΔE_{ab}	ΔE_{tor}	ΔE_{op}	ΔE_{14vdW}	ΔE_{14e}	ΔE_{vdW}	ΔE_e	ΔE_{tot}
A. ΔE_1^b for association of 3(GPP) _n ^c into a triple helix										
(GPP) ₄	-0.49	-5.08	-1.24	0.70	0.12	4.07	2.61	-103.78	-85.89	-189.00
(GPP) ₁₂	-1.25	-12.25	1.29	2.84	0.11	5.86	2.20	-309.23	-234.70	-554.13
(GPP) ₁₆	-1.90	-17.15	1.30	3.34	0.22	8.15	2.88	-412.49	-325.77	-741.42
B. ΔE_3^d for substitution of (GPHP) or (GAA) into single chains of (GPP) _n										
(GPHP) ₄	-0.45	-1.07	7.74	9.25	0.45	-3.28	19.97	0.63	-29.11	4.15
(GAA) ₄	-4.27	-1.14	-220.43	-138.82	0.06	-18.53	146.74	23.28	-172.73	-385.83
(GPHP) ₁₂	-0.62	-0.11	26.08	9.33	0.23	-45.25	26.54	2.56	-15.58	3.17
(GAA) ₁₂	-10.62	-0.95	-656.66	-412.90	-0.76	-73.97	425.90	69.68	-481.92	-1142
(GPHP) ₁₆	-1.08	-0.07	38.63	13.91	0.59	-60.77	37.62	1.42	-27.98	2.26
(GAA) ₁₆	-13.48	-1.22	-876.64	-549.68	-0.94	-97.82	565.60	93.37	-642.55	-1523
C. ΔE_4^e for the substitution of (GPHP) or (GAA) into triple helices of (GPP) _n										
(GPHP) ₄	-0.53	-0.22	8.39	12.89	0.35	-2.53	22.06	-0.48	-42.83	-2.90
(GAA) ₄	-3.67	0.07	-218.79	-138.52	0.38	-20.12	146.35	46.06	-184.41	-372.63
(GPHP) ₁₂	1.14	-4.39	27.72	37.39	0.62	-12.92	51.42	-6.92	-108.81	-15.02
(GAA) ₁₂	-8.73	-1.02	-655.24	-416.92	0.61	-69.88	425.15	127.86	-508.43	-1106.6
(GPHP) ₁₆	2.37	-5.69	38.17	43.88	0.28	-20.27	60.57	-8.35	-132.38	-21.42
(GAA) ₁₆	-11.80	0.24	-874.32	-555.90	0.77	-91.38	569.32	173.97	-681.89	-1471.0
D. $\Delta\Delta E^f$ stabilization energies for triple helices of (GPHP) _n or (GAA) _n as compared to (GPP) _n										
(GPP) ₄	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(GPHP) ₄	-0.08	0.85	0.65	3.64	-0.10	0.74	2.08	-1.10	-13.72	-7.05
(GAA) ₄	0.60	1.21	1.64	0.30	3.72	-1.58	-0.39	22.78	-11.68	13.20
(GPP) ₁₂	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(GPHP) ₁₂	1.77	-4.28	1.64	28.05	0.39	32.33	24.61	-9.74	-93.23	-18.19
(GAA) ₁₂	1.89	-0.07	1.42	-4.02	1.37	4.09	-0.75	58.18	-26.51	35.59
(GPP) ₁₆	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(GPHP) ₁₆	3.45	-5.62	-0.45	29.97	-0.31	40.50	22.95	-9.78	-104.40	-23.68
(GAA) ₁₆	1.68	1.45	2.32	-6.22	1.70	6.44	3.72	80.61	-39.34	52.36

^a Energies were computed in kcal/mol using Kollman force fields with united-atoms (Blaney *et al.*, 1982). The individual ΔE terms are defined as follows: E_{bs} is the sum of energies arising from bond stretching or compression beyond the optimum bond length; E_{ab} is the sum of energies for angles which are distorted from their optimum values; E_{op} is the sum of energies for the bending of planar atoms out of the plane; E_{tor} is the sum of the torsional energies which arise from rotations about each respective dihedral angle; E_{vdW} and E_e are the sum of energies due to nonbonded van der Waals and electrostatic interactions, respectively; E_{14vdW} and E_{14e} are the sum of the van der Waals and electrostatic interaction energies, respectively, for atoms connected by three bonds; and E_{hb} is the sum of energies due to hydrogen bond interactions. Nonbonded van der Waals and electrostatic interactions were not considered beyond a cutoff distance of 8 Å.

^b $\Delta E_1 = E_{\text{triple helix}} - 3E_{\text{single chain}}$ as described in Methods.

^c GPP, GPHP, and GAA represent the peptide chains (Gly-Pro-Pro), (Gly-Pro-Hyp), and (Gly-Ala-Ala), respectively.

^d $\Delta E_3 = E_{GXY} - E_{GPP}$, for the modified polypeptide chains resulting from the substitution of the tripeptide sequences (Gly-Pro-Hyp) or (Gly-Ala-Ala) into the (Gly-Pro-Pro) polypeptide chain.

^e $\Delta E_4 = E_{GXY} - E_{GPP}$, for the modified triple helices resulting from the substitution of the tripeptide sequences (Gly-Pro-Hyp) or (Gly-Ala-Ala) into the (Gly-Pro-Pro) triple helix.

^f $\Delta\Delta E = \Delta E_4 - \Delta E_3$ as described in Methods.

specific interaction of water molecules is important in providing additional stability to the (Gly-Pro-Hyp) triple helix.

Although the structures of the nonhelical telopeptide segments which extend beyond both the N- and C-terminal helical regions of collagen are not well-established, it is thought that reverse-bends in these segments allow them to interact with the terminal helical regions (Capaldi and Chapman, 1982; Helseth *et al.*, 1979). It is also possible that these telopeptides interact with amino acid residues on

adjacent collagen molecules. Hence, hydroxyproline residues may function in the intermolecular interactions between telopeptides and adjacent collagen molecules.

3.1.3. Evaluation of Energies Computed for the (Gly-Ala-Ala)₁₂ Triple Helix Model

Alanine was substituted for proline at both the "X" and "Y" positions of triple-helical (Gly-X-Y)_n. The energy minimized structures of (Gly-Ala-Ala)_n

“Smith” Collagen Microfibril Model

and (Gly-Pro-Pro)_n were superimposable with rms deviations less than 0.3 Å for all backbone atoms. According to the study by Bhatnagar *et al.* (1988), nonbonded van der Waals interactions are the major stabilizing forces in the collagen triple helix of (Gly-Pro-Pro)₁₀. This is probably a result of the favorable stacking arrangement between adjacent proline rings in their model systems. Our studies have also shown this to be true for (Gly-Pro-Pro)_n and (Gly-Pro-Hyp)_n, where $n=4, 12,$ and 16 . It is clear from the positive energy values of $\Delta\Delta E_{vdW}$ for (Gly-Ala-Ala)_n in Table II(D) that the proline rings make important contributions to $\Delta\Delta E_{vdW}$. In addition, $\Delta\Delta E_{tot}$ values for (Gly-Ala-Ala)_n denote less stabilization of the triple helix as compared to those for (Gly-Pro-Pro)_n. It is interesting to note that the $\Delta\Delta E_c$ values for (Gly-Ala-Ala)_n are energetically favorable. This may result from a better packing arrangement of the polypeptide backbone atoms due to the fact that alanine has a smaller sidechain than proline. Furthermore, the ring-nitrogen in proline is now a nonring amide nitrogen in alanine, thereby increasing the potential for electrostatic interactions and hydrogen bond formation with carbonyl groups. These comparisons between the triple helices of (Gly-Pro-Pro)_n, (Gly-Pro-Hyp)_n, and (Gly-Ala-Ala)_n show, as indicated by their $\Delta\Delta E_{vdW}$ values, the importance of proline rings in contributing to hydrophobic interactions.

Structural parameters for the computed triple helix and “Smith” microfibril models (discussed below in Section 3.2) are summarized in Table III. Measurements were made for the triple helix and microfibril models of (Gly-X-Y)₁₂ (without the polypeptide end groups, as described in Methods) using SYBYL. The right most column shows the values extrapolated for the dipole moment of the complete

helical domain (1014 amino acid residues per chain) of each collagen model. Collagen has a permanent dipole moment of 10^3 to 10^4 Debyes (Kranck *et al.*, 1982; Umemura *et al.*, 1979). Our computed structures have values that fall within this range. It must be noted that the above literature values apply to native collagens and not to synthetic collagen models. It is clear that the presence of the hydroxyl group in hydroxyproline corresponds to an increase in the dipole moment for the computed (Gly-Pro-Hyp) collagen model as compared to the (Gly-Pro-Pro) model (Table III). The significance of the large dipole moments is that they contribute to the uniaxial alignment of the collagen molecules into microfibrils and fibrils. For instance, when collagen gels are subjected to external magnetic fields during *in vitro* self-assembly, the resulting structures formed contain collagen molecules that exhibit a high degree of uniaxial alignment (Torbet and Ronziere, 1984).

3.2. Molecular Modeling of the Collagen Microfibril

The computed three-dimensional structures for the (Gly-Pro-Pro)₁₂ and (Gly-Pro-Hyp)₁₂ “Smith” microfibrils are shown in Figs. 4A and B, respectively. Each microfibril model consists of 540 amino acid residues, 36 amino acid residues per polypeptide chain. The polypeptide backbone torsional angles for each energy-minimized microfibril model are given in Table I(C). In the (Gly-Pro-Pro)₁₂ microfibril (Fig. 4A), one of the triple helices is colored to highlight its three polypeptide chains. It is clear that the three chains form a right-handed helical twist. Figure 4C displays a cross-section of the “Smith” microfibril model as viewed from the carboxyl terminal end.

Table III. Structural Parameters for the Computed Models of Collagen and Collagen Microfibrils^a

Sequence	Radius (nm) ^b	Residue height (nm)	Pitch height (nm) ^c	Dipole moment (D) ^d
Triple helix				
(Gly-Pro-Pro)	0.52	0.288	7.78	3.2×10^3
(Gly-Pro-Hyp)	0.63	0.291	7.85	4.1×10^3
(Gly-Ala-Ala)	0.47	0.290	7.83	3.8×10^3
Microfibril				
(Gly-Pro-Pro)	1.50	0.283	53	1.1×10^4
(Gly-Pro-Hyp)	1.60	0.286	57	1.6×10^4
(Gly-Ala-Ala)	1.40	0.285	50	1.6×10^4

^a These averaged values were computed from data obtained using SYBYL (SYBYL MENDYL, v5.32, 1990).

^b The error in the radius is ± 0.05 nm.

^c The error in the pitch height is $\pm 5\%$. The pitch height of each microfibril can be computed from its tilt angle ($\sim 10^\circ$ with respect to the microfibril long axis) and the circumference of each microfibril.

^d Dipole moments are in units of Debyes.

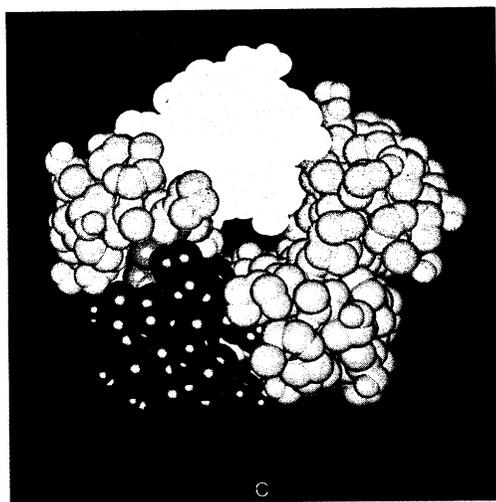
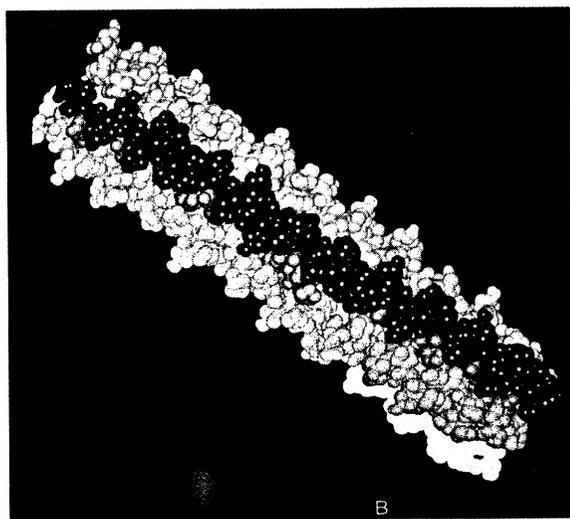
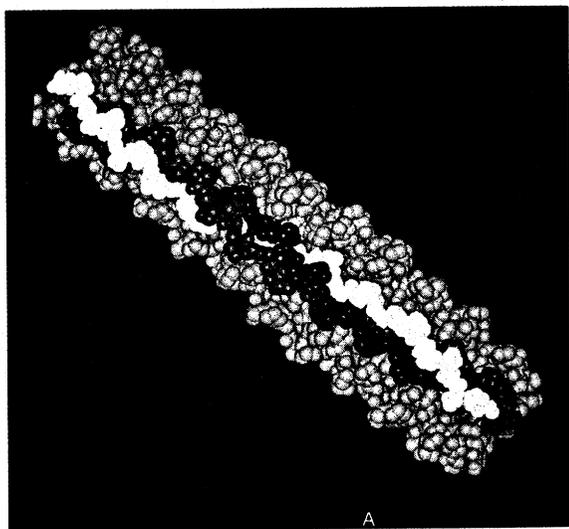


Fig. 4. (A) A space-filling model of the "Smith" microfibril which consists of $(\text{Gly-Pro-Pro})_{12}$ polypeptide chains. There are five collagen molecules within the microfibril. One collagen triple helix has each of its polypeptide chains colored differently in order to show the packing orientation and the right-hand helical twist of the individual triple helix. It is evident from our model that the microfibril has a left-handed superhelical twist. (B) A space-filling model of the microfibril for $(\text{Gly-Pro-Hyp})_{12}$. Each collagen molecule is colored differently to show the packing orientation. The residues shown in green (hydroxyprolines) are involved in specific hydrogen bond interactions between two adjacent triple helices. (There are five sets of these interactions within this microfibril model, only one set is shown in green). (C) A space-filling model of the C-terminal cross-section of the microfibril. The circular packing arrangement seen here is similar to that proposed by Smith (1968). Although our models do show free volume within the interior of the microfibril, this volume would be further reduced upon the inclusion of the amino acid side chains other than those of the imino acids.

The longitudinal distance of the microfibril model per 36 residue polypeptide chain length is approximately 10.3 nm. In the (Gly-Pro-Pro)₁₂ microfibril model, the estimated D-spacing for the lateral stagger of 234 residues is 66.2 nm, in agreement with literature values (Chapman, 1984; Meek *et al.*, 1979). Since the lateral interaction between any two adjacent molecules is not perfectly parallel but slightly tilted, this results in giving the computed microfibril a super-helical left-handed twist (Figs. 4A and B). As a result of this super-helical twisting by each triple helix in the microfibril models, the length of the D-spacing measured in the (Gly-Pro-Pro)₁₂ microfibril (66.2 nm) is shorter than that measured for the corresponding triple helix model ($D=67.4$ nm). The tilt of each microfibril can be approximated by taking the \cos^{-1} of the ratio of the axial length of the microfibril over the axial length of the triple helix. As a result, the estimated tilt of each triple-helical molecule with respect to the longitudinal axis of the microfibril is approximately 10° . The radius for the (Gly-Pro-Pro)₁₂ microfibril model is approximately 1.50 nm (Table III). In comparison, the literature values for experimental samples of collagen in tendon range from 1.40–1.75 nm, depending on the specific tissue (Piez, 1984; Chapman and Hulmes, 1984; Chapman, 1984; Bouteille and Pease, 1971). Knowing the tilt angle and the circumference of the (Gly-Pro-Pro)₁₂ microfibril model, it is estimated that the pitch for the observed super-helical left-handed twist is approximately 53 nm (Table III).

3.2.1. Evaluation of ΔE_1 , ΔE_3 , ΔE_4 and $\Delta\Delta E$ Energies for the computed "Smith" Microfibril Models

In order to determine which type of interaction contributes most to the formation of the microfibril complex and which contributes to the stabilization of the microfibril once it is formed, the potential energy differences ΔE_1 (see Scheme I) for the (Gly-Pro-Pro)₁₂ "Smith" microfibril model were calculated as they were for the triple helix. Table IV(A) displays the individual components and total potential energy differences, ΔE_1 for the (Gly-Pro-Pro)₁₂ microfibril. It is clear from the ΔE_{vdw} value that nonbonded van der Waals interactions are important for microfibril formation. Synthetic collagen molecules containing the (Gly-Pro-Pro)_n polypeptide sequence do form microcrystalline structures of laterally packed triple helices (Okuyama *et al.*, 1972, 1981; Sakakibara *et al.*, 1972; Olsen *et al.*, 1971), and our study suggests that the

packing forces present within the (Gly-Pro-Pro)₁₂ microfibril may be due mainly to van der Waals interactions.

Computed ΔE_3 and ΔE_4 energies for the "unassociated" and "associated" microfibril models of (Gly-Pro-Hyp)₁₂ and (Gly-Ala-Ala)₁₂ are listed in Table IV(B). Meanwhile, Table IV(C) shows the stabilization energies, $\Delta\Delta E$ for the (Gly-Pro-Hyp)₁₂ and (Gly-Ala-Ala)₁₂ microfibrils as compared to that for (Gly-Pro-Pro)₁₂. Also shown in Tables IV(B) and (C) are the respective energies for (Gly-Ala-Ala)_{12i} and (Gly-Ala-Ala)_{12f}, where *i* denotes energy minimization of the starting microfibril structure with all backbone atoms constrained to remain in their original positions, and *f* denotes minimization with no constraints on the backbone atoms. The rms deviations for all backbone atoms between the initial microfibril, (Gly-Pro-Pro)₁₂ and (Gly-Pro-Hyp)₁₂, was 0.75 Å. Similarly, comparing the initial microfibril (Gly-Pro-Pro)₁₂ to (Gly-Ala-Ala)_{12i} and (Gly-Ala-Ala)_{12f}, the rms deviations were observed to be 0.0 and 1.3 Å, respectively. In addition, comparison of the minimized structure of (Gly-Pro-Hyp)₁₂ to (Gly-Ala-Ala)_{12i} and (Gly-Ala-Ala)_{12f} resulted in rms values of 0.75 and 1.40 Å, respectively. The differences in the rms values may contribute in part to the observed radii for the microfibrils as shown in Table III.

3.2.2. Evaluation of Energies Computed for the "Smith" Microfibril Model of (Gly-Pro-Hyp)₁₂

Examination of the (Gly-Pro-Pro)₁₂ microfibril structure shows that no significant hydrogen bonding exists between the polypeptide backbone of adjacent triple helices due to the presence of proline rings. However, this is not the case for the (Gly-Pro-Hyp)₁₂ and (Gly-Ala-Ala)₁₂ microfibrils. The computed three-dimensional model for the (Gly-Pro-Hyp)₁₂ microfibril shows that the hydroxyl group of hydroxyproline is able to form hydrogen bonds between different collagen molecules. Figure 4B is a space-filling model of the (Gly-Pro-Hyp)₁₂ microfibril and displays the hydroxyproline-rich regions (in green) which are important for the formation of hydrogen bonds. These bonds are formed between the hydroxyl group of hydroxyproline and the backbone carbonyl of an adjacent collagen triple helix. Upon further inspection of the three-dimensional model for the (Gly-Pro-Hyp)₁₂ microfibril, a pattern of hydroxyproline interactions is evident as shown in Fig. 5A. At intervals of 6 of 9 amino acid residues,

Table IV. Potential Energy^a Differences Calculated for the Association Steps of the Microfibril as per Scheme

	ΔE_{bs}	ΔE_{hb}	ΔE_{ab}	ΔE_{tor}	ΔE_{op}	ΔE_{14vdW}	ΔE_{14e}	ΔE_{vdW}	ΔE_e	ΔE_{tot}
A. ΔE_1^b for association of 5(GPP) ₁₂ ^c into a microfibril										
(GPP) ₁₂	0.90	0.27	38.82	112.45	9.57	-56.82	-20.26	-634.86	-0.44	-550.35
B. ΔE_3^d for substitution of (GPHP) or (GAA) into triple helices of (GPP) ₁₂										
(GPHP) ₁₂	5.72	-21.95	138.58	186.93	3.10	-64.58	255.74	-34.58	-544.07	-75.11
(GAA) ₁₂	-43.64	-5.11	-3276.2	-2084.6	3.03	-349.41	2125.7	639.28	-2542.1	-5535.5
C. ΔE_4^e for the substitution of (GPHP) or (GAA) into microfibrils of (GPP) ₁₂										
(GPHP) ₁₂	9.88	-21.90	118.20	116.56	4.14	-49.10	240.93	-44.56	-940.27	-566.12
(GAA) _{12i}	-36.40	6.20	-3293.4	-2133.7	-3.35	-291.37	2157.1	1069.2	-2610.1	-5085.9
(GAA) _{12f}	-43.67	-27.42	3283.0	-2133.9	0.17	-282.96	2171.2	799.80	-2844.4	-5644.1
D. $\Delta\Delta E^f$ stabilization energies for microfibrils of (GPHP) ₁₂ or (GAA) ₁₂ as compared to (GPP) ₁₂										
(GPP) ₁₂	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(GPHP) ₁₂	4.16	0.05	-20.38	-70.37	1.04	15.47	-14.81	-9.90	-396.20	-491.01
(GAA) _{12i}	7.24	11.31	-17.20	-49.08	-0.32	57.67	31.34	429.87	-67.99	449.60
(GAA) _{12f}	-0.03	-22.31	-6.77	-49.29	-2.86	66.45	45.51	160.52	-302.23	-108.56

^a Energies were computed in kcal/mol using Kollman force fields with united-atoms (Blaney *et al.*, 1982). The individual ΔE terms are defined as follows: E_{bs} is the sum of energies arising from bond stretching or compression beyond the optimum bond length; E_{ab} is the sum of energies for angles which are distorted from their optimum values; E_{op} is the sum of energies for the bending of planar atoms out of the plane; E_{tor} is the sum of the torsional energies which arise from rotations about each respective dihedral angle; E_{vdW} and E_e are the sum of energies due to nonbonded van der Waals and electrostatic interactions, respectively; E_{14vdW} and E_{14e} are the sum of the van der Waals and electrostatic interaction energies, respectively, for atoms connected by three bonds and E_{hb} is the sum of energies due to hydrogen bond interactions. Nonbonded van der Waals and electrostatic interactions were not considered beyond a cutoff distance of 8 Å.

^b $\Delta E_1 = E_{\text{microfibril}} - 5E_{\text{triple}}$, as described in Methods.

^c GPP, GPHP, and GAA represent the tripeptide sequences (Gly-Pro-Pro), (Gly-Pro-Hyp), and (Gly-Ala-Ala), respectively.

^d $\Delta E_3 = E_{GXY} - E_{GPP}$, for the modified triple helices resulting from the substitution of the tripeptide sequences (Gly-Pro-Hyp) or (Gly-Ala-Ala) into the (Gly-Pro-Pro) triple helix.

^e $\Delta E_4 = E_{GXY} - E_{GPP}$, for the modified microfibrils resulting from the substitution of the tripeptide sequences (Gly-Pro-Hyp) or (Gly-Ala-Ala) into the (Gly-Pro-Pro) microfibril model.

^f $\Delta\Delta E = \Delta E_4 - \Delta E_3$ as described in Methods.

hydroxyprolines face toward the interior of the microfibril. The hydroxyl groups of adjacent hydroxyprolines are seen (Fig. 5A as green regions) to hydrogen bond with adjacent carbonyl backbone groups and also with each other. A close-up view of one of these hydroxyproline regions is shown in Fig. 5B. The arrows point to the interacting residues in the interior of the microfibril. This type of hydrogen bond interaction in the (Gly-Pro-Hyp)₁₂ microfibril provides additional stability to the microfibril complex. The observed pattern suggests possible sites where cross-linking interactions may be significant within the interior of collagen microfibrils.

In contrast, the $\Delta\Delta E_{hb}$ value for the (Gly-Pro-Hyp)₁₂ microfibril model in Table IV(C) seem to indicate that no additional hydrogen bonds were formed. As noted previously, intrapolypeptide hydrogen bonds do exist in the initial polypeptide chain and triple helix for (Gly-Pro-Hyp)₁₂. It is likely that hydrogen bond formation between collagen molecules within the microfibril model is a result of the breaking of some intrapolypeptide hydrogen bonds (Fig. 3C).

Inspection of the $\Delta\Delta E_e$ term in Table IV(C) for (Gly-Pro-Hyp)₁₂ shows that electrostatic interactions resulting from the additional hydroxyl functionalities are significant. Therefore, as observed in the three-dimensional models, it can be stated that hydroxyprolines in (Gly-Pro-Hyp)₁₂ provide further stabilization to the packing of collagen molecules within the microfibril as compared to the microfibril structure of (Gly-Pro-Pro)₁₂.

3.3.3. Evaluation of Energies Computed for the "Smith" Microfibril Models of (Gly-Ala-Ala)_{12i} and (Gly-Ala-Ala)_{12f}

Table IV(C) shows the $\Delta\Delta E$'s for (Gly-Ala-Ala)_{12i} and (Gly-Ala-Ala)_{12f} microfibrils. The (Gly-Ala-Ala)_{12i} microfibril was minimized while constrained to the original (Gly-Pro-Pro)₁₂ microfibril structure (i.e., rms deviations for all backbone atoms = 0.0). The (Gly-Ala-Ala)_{12i} structure was further minimized after removal of all constraints (rms = 1.3 Å) to give (Gly-Ala-Ala)_{12f}. The $\Delta\Delta E_{vdW}$

“Smith” Collagen Microfibril Model

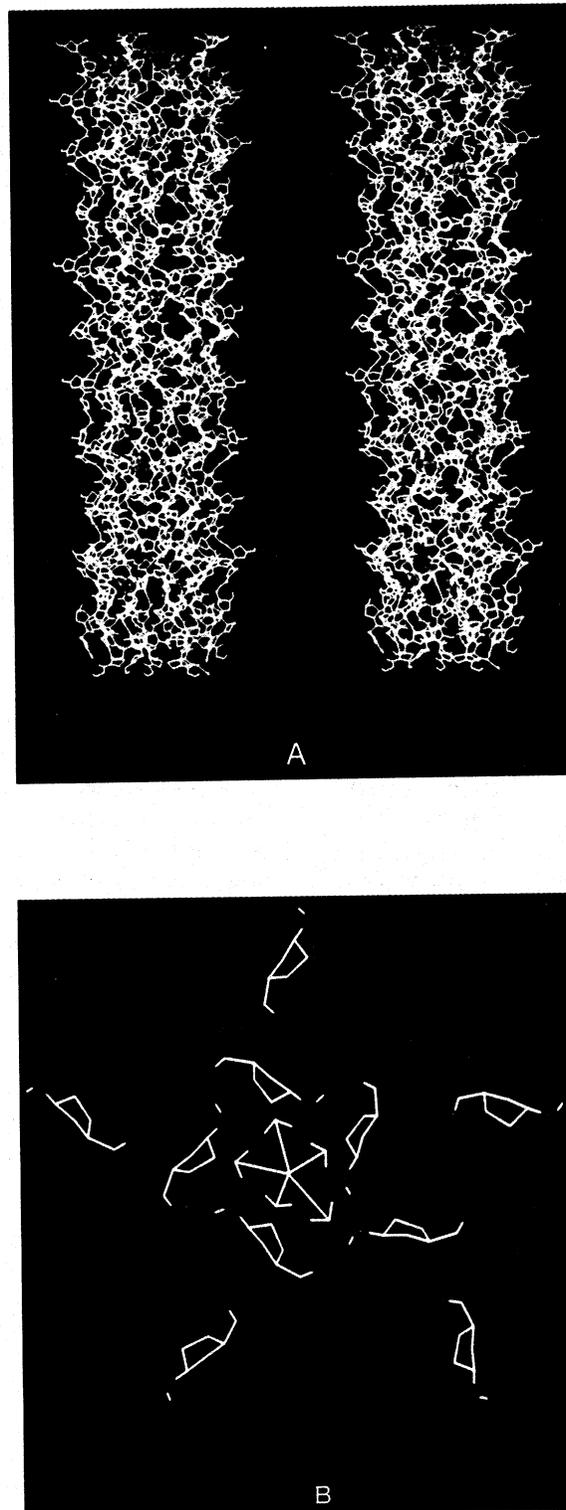


Fig. 5. (A) A stereo figure (in the relaxed viewing mode) indicating an important type of interaction occurring in the interior of the microfibril for $(\text{Gly-Pro-Hyp})_{12}$. At intervals of 6 or 9 amino acid residues along the long axis of the microfibril, five hydroxyproline rings (one per triple helix) face the center of the microfibril. In the circular array (shown in green), each hydroxyproline forms a hydrogen bond to the backbone carbonyl or hydroxyl group of an adjacent hydroxyproline. (B) A view of one of the interaction regions shown in (A). The hydroxyprolines point toward the center of the microfibril. The green arrows indicate the interacting hydroxyprolines.

for (Gly-Ala-Ala)_{12i} is unfavorable compared to (Gly-Pro-Pro)₁₂. This indicates that the van der Waals interactions resulting from the presence of proline are very important in microfibril formation.

In the case of the (Gly-Ala-Ala)_{12f} microfibril, hydrogen bonding and electrostatic energies are also significant when compared to (Gly-Pro-Pro)₁₂ [Table IV(C)]. Hydrogen bonding between adjacent collagen molecules is observed in the (Gly-Ala-Ala)_{12f} microfibril. However, unlike the (Gly-Pro-Hyp)₁₂ microfibril, where the hydroxyproline forms hydrogen bonds to adjacent collagen molecules, the hydrogen bonds existing in (Gly-Ala-Ala)_{12f} are formed between backbone amide hydrogens and the carbonyl groups of adjacent collagen molecules. The alanine side chain is smaller in comparison to the proline ring and thus allows for closer packing between the individual triple helices. The proline ring-nitrogen is now a nonring amide in alanine and thus contributes to hydrogen bonding and electrostatic interactions [$\Delta\Delta E_{hb}$ and $\Delta\Delta E_e$ for (Gly-Ala-Ala)_{12f} in Table IV(C)]. Although the $\Delta\Delta E_{tot}$ in Table IV(C) for (Gly-Ala-Ala)_{12f} is negative compared to that for (Gly-Pro-Pro)₁₂, it is unlikely that a synthetic peptide of (Gly-Ala-Ala)₁₂ will form a collagen triple helix or microfibril under normal conditions, given its flexible polypeptide chain. However, Type I collagen does contain several (Gly-Ala-Ala) sequences and numerous (Gly-Ala-Y) and (Gly-X-Ala) sequences. It is possible that these sequences serve to allow for increased free volume in cases where bulky sidechains must pack effectively. In addition, these regions may allow for "tighter" packing between collagen molecules. It would be interesting to determine the structural effects of incorporating amino acids with bulky sidechains into the (Gly-Ala-Ala) sequence of Type I collagen.

4. CONCLUSION

Our computed three-dimensional models for the triple-helical and microfibrillar structures of collagen have structural parameters which agree with X-ray diffraction data and structural findings obtained in other studies. The microfibrillar structures that were developed are based on the model proposed by Smith (1968). Structural information obtained from studies which have identified possible interactions between two triple helices of (Gly-Pro-Pro)₅ (Nemethy and Scheraga, 1984) were used in order to assist the packing of five triple helices of (Gly-Pro-Pro)₁₂ in the "Smith" microfibril. The results show that when proline is present in both the "X" and "Y" positions:

(Gly-Pro-Pro)_n, van der Waals interactions stabilize the microfibril complex. However, when polar amino acid residues such as hydroxyproline or residues with small side chains such as alanine are present, electrostatic interactions become more important. Our computed models do indicate that nonbonded van der Waals interactions are crucial for microfibril formation. Nevertheless, it is probably electrostatic interactions that further increase the stability of the microfibril and determine the specificity by which collagen molecules interact with one another (Meek *et al.*, 1979).

The three-dimensional structure of the (Gly-Pro-Hyp)₁₂ microfibril shows several interesting features. First, the axial tilt of each collagen triple helix within the "Smith" microfibril model results in the microfibril having a superhelical left-handed twist. The pitch estimated for this superhelical twist is in the range of 50–57 nm for the three-dimensional models described (Table III), but this range may vary depending on the axial tilt of each molecule within the packing of the microfibril. It is possible that the packing of collagen molecules varies with tissue type (i.e., skin, tendon, and bone) and tissue state (wet and dry). For example, a three-dimensional computer image of a series of experimental cross-sectional profiles of mineralized collagen fibrils from bone (Lee and Glimcher, 1991) shows that the diameter of this collagen fibril varies periodically along the fibril long axis. It is possible that the tilt of each collagen molecule within this fibril varies along the fibril long axis. Furthermore, substitution of specific Type I collagen sequences into the computed microfibril models will allow for the comparison of collagen packing for different regions of the triple helix (i.e., comparing regions containing many charged and polar residues with regions containing mainly hydrophobic or nonpolar residues). Second, the computed three-dimensional microfibril models can also be modified by removing one triple helix or a part of it in order to model regions known as "gaps" as described in the "Smith" microfibril model (Smith, 1968). Hence, comparison between the gap and overlap regions of the Smith microfibril can be studied. In addition, modeling of the N- and C-terminal nonhelical telopeptide structures is possible by initially docking each flexible telopeptide segment into the corresponding rigid model of the N- or C-terminal microfibril domain [i.e., energy minimization or molecular dynamics using distance or distance range constraints can be applied here, SYBYL (v5.32)]. Third, the microfibril model allows one to examine the specific interactions that exist in three dimensions.

"Smith" Collagen Microfibril Model

The pattern describing hydroxyproline interactions as shown in Fig. 5B reveals that interactions between collagen molecules are specific and nonrandom. Initially, these microfibril models can be modified by substituting the actual collagen sequences into the models. Energy refinement of the modified models will allow for the determination of which regions of the microfibril packing are high in energy and subsequently, which regions require modification of the packing model. These models may also reveal clues to how specific point mutations (Prockop, 1990) effect collagen interaction, structure, and hence may reveal how certain collagen diseases arise.

Further development of the collagen models will allow for the determination of the specific parameters required to develop optimal site-specific reagents for collagen. The three-dimensional model of collagen also allows for the precise study of the structure-function relationships evolving from amino acid substitutions. The proposed model for the collagen microfibril offers a realistic system whereby the inter- and intrapolypeptide interactions can be examined. The model permits direct concentration on any particular interaction region of the collagen microfibril. In addition, the polypeptide sequences of different types of fiber-forming collagens can be incorporated and examined with this model. The interactions studied here can then be correlated with other three-dimensional models for the collagen fibril. At present, studies are in progress which examine the characteristics of the microfibril after the incorporation of the Type I calf skin collagen amino acid sequence.

ACKNOWLEDGMENTS

The authors would like to thank Drs. Gregory King, Tom Kumosinski, and Mark Hurle (Smith Kline Beecham) for helpful suggestions and review of this manuscript. Appreciation is also given to the Computer Unit at ERRC. We would like to thank Diane Koch for her efforts in helping to prepare this manuscript.

NOTE ADDED IN PROOF

The coordinates for the collagen models above have been deposited with the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, N.Y. 11973, from which copies are available.

REFERENCES

- Bear, R. (1942). *J. Am. Chem. Soc.* **64**, 727.
- Berg, R. A., and Prockop, D. J. (1973). *Biochem. Biophys. Res. Comm.* **52**, 115-120.
- Bhatnagar, R. S., Pattabingman, N., Sorensen, K. R., Langridge, R., MacElroy, R. D., and Renugopalakrishnan, V. (1982). *J. Biomol. Struct. Dynam.* **104**, 6424-6434.
- Blaney, J. M., Weiner, P. K., Dearing, A., Kollman, P. A., Jorgensen, E. C., Oatley, S. J., Burrige, J. M., and Blake, C. C. F. (1982). *J. Am. Chem. Soc.* **104**, 6424-6434.
- Bouteille, M., and Pease, D. C. (1971). *J. Ultrastruct. Res.* **35**, 311-338.
- Brodsky, B., and Eikenberry, E. (1985). In *Annals of the New York Academy of Science: Biology, Chemistry and Pathology of Collagen*, Vol. 460 (Fleischmajer, R., Olsen, B. R., Kuhn, K., eds.), The New York Academy of Sciences, New York, pp. 73-85.
- Brodsky, B., Eikenberry, E. F., Belbruno, K., and Sterling, K. (1982). *Biopolymers* **21**, 935-951.
- Capaldi, M. J., and Chapman, J. A. (1982). *Biopolymers* **21**, 2291-2313.
- Chapman, J. A. (1984). In *Connective Tissues Matrix: Topics in Molecular and Structural Biology*, Vol. 5 (Hukins, D. W. L., ed.), Verlag Chemie, Weinheim, pp. 89-132.
- Chapman, J. A., and Hulmes, D. J. S. (1984). In *Ultrastructure of the Connective Tissue Matrix* (Ruggeri, A., Motta, P. M., eds.), Martinus Nijhoff Publishers, Boston, pp. 1-33.
- Chew, M. W. K., and Squire, J. M. (1986). *J. Biol. Macromol.* **8**, 27-36.
- Eikenberry, E. F., and Brodsky, B. (1980). *J. Mol. Biol.* **144**, 397-404.
- Fietzek, P. P., and Kuhn, K. (1976). In *International Review of Connective Tissue Research*, Vol. 7 (Hall, D., and Jackson, D. S., eds.), Academic Press, New York, pp. 1-60.
- Fraser, R. D. B., MacRae, T. P., and Miller, A. (1987). *J. Mol. Biol.* **193**, 115-125.
- Fraser, R. D. B., MacRae, T. P., Miller, A., and Suzuki, E. (1983). *J. Mol. Biol.* **167**, 497-521.
- Fraser, R. D. B., Miller, A., and Parry, D. A. D. (1974). *J. Mol. Biol.* **83**, 281-283.
- Galloway, J. (1984). *TIBS* **9**, 233-238.
- Giraud-Guille, M.-M. (1987). *Mol. Cryst. Liq. Cryst.* **153**, 15-30.
- Gordon, M. K., Gerecke, D. R., Dublet, B., Van Der Rest, M., Sugrue, S. P., and Olsen, B. R. (1990). In *Annals of the New York Academy of Science: Structure, Molecular Biology and Pathology of Collagen*, Vol. 580 (Fleischmajer, R., Olsen, B. R., Kuhn, K., eds.), The New York Academy of Sciences, New York, pp. 8-16.
- Helseth, Jr., D. L., Lechner, J. H., and Veis, A. (1979). *Biopolymers* **18**, 3005-3014.
- Hodge, A. J., and Petruska, A. J. (1963). In *Aspects of Protein Structure* (Ramachandran, G. N., ed.), Academic Press, London, pp. 289-300.
- Hofmann, H., Fietzek, P. P., and Kuhn, K. (1978). *J. Mol. Biol.* **125**, 137-165.
- Hosemann, R., Dressig, W., and Nemetschek, Th. (1974). *J. Mol. Biol.* **83**, 275-280.
- Hulmes, D. J. S., Holmes, D. F., and Cummings, C. (1985). *J. Mol. Biol.* **183**, 473-477.
- Hulmes, D. J. S., Jesior, J.-C., Miller, A., Berthet-Colominas, C., and Wolff, C. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 3567-3571.
- Hulmes, D. J. S., and Miller, A. (1979). *Nature* **282**, 878-880.
- Hulmes, D. J. S., and Miller, A. (1981). *Nature* **293**, 239-240.
- Jelinski, L. W., Sullivan, C. E., and Torchia, D. A. (1980). *Nature* **284**, 531-534.
- Kranck, H., Bernengo, J. C., and Vasilescu, D. (1982). *Appl. Phys. Commun.* **2**, 189-202.

- Lee, D. D., and Glimcher, M. J. (1991). *J. Mol. Biol.* **217**, 487-501.
- Lees, S., Pineri, M., and Escoubes, M. (1984). *Int. J. Biol. Macromol.* **6**, 133-136.
- Martin, G. R., Timpl, R., Muller, P. K., and Kuhn, K. (1985). *TIBS* **10**, 285-287.
- Meek, K. M., Chapman, J. A., and Hardcastle, R. A. (1979). *J. Biol. Chem.* **254**, 10,710-10,714.
- Miller, A. (1982). *TIBS* **7**, 13-18.
- Miller, A. (1976). In *Biochemistry of Collagen* (Ramachandran, G. N., and Reddi, A. H., eds.), Plenum Press, New York, pp. 85-136.
- Miller, A., and Parry, D. A. D. (1973). *J. Mol. Biol.* **75**, 441-447.
- Miller, A., and Wray, J. S. (1971). *Nature* **230**, 437-439.
- Miller, E. J. (1985). In *Annals of the New York Academy of Sciences: Biology, Chemistry and Collagen*, Vol. 460 (Fleischmajer, R., Olsen, B. R., and Kuhn, K., eds.), The New Academy of Sciences, New York, pp. 1-13.
- Miller, M. H., Nemethy, G., and Scheraga, H. A. (1980). *Macromol.* **13**, 470-478.
- Miller, M. H., and Scheraga, H. A. (1976). *J. Polymer Sci. Symp.* **54**, 171-200.
- Na, G. C., Butz, L. J., and Carroll, R. J. (1986a). *J. Biol. Chem.* **261**, 12,290-12,299.
- Na, G. C., Butz, L. J., Bailey, D. G., and Carroll, R. J. (1986b). *Biochemistry* **25**, 958-966.
- Nemethy, G., and Scheraga, H. A. (1989). *Bull. Inst. Chem. Res.* **66**, 398-408.
- Nemethy, G., and Scheraga, H. A. (1984). *Biopolymers* **23**, 2781-2799.
- Nemethy, G., and Scheraga, H. A. (1986). *Biochemistry* **25**, 3184-3188.
- Okuyama, K., Arnott, S., Takayanagi, M., and Kakudo, M. (1981). *J. Mol. Biol.* **152**, 427-443.
- Okuyama, K., Nobuhiro, G., and Takayanagi, M. (1978). *Chem. Lett.* 509-512.
- Okuyama, K., Tanaka, N., Ashida, T., Kakudo, M., Sakakibara, S., and Kishida, Y. (1972). *J. Mol. Biol.* **72**, 571-576.
- Olsen, B. R., Berg, R. A., Sakakibara, S., Kishida, Y., and Prockop, D. J. (1971). *J. Mol. Biol.* **57**, 589-595.
- Parry, D. A. D., and Craig, A. S. (1979). *Nature* **282**, 213-215.
- Piez, K. A. (1984). In *Extracellular Matrix Biochemistry* (Piez, K. A., and Reddi, A. H. eds.), Elsevier, New York, pp. 1-39.
- Piez, K. A., and Trus, B. L. (1981). *Biosci. Rep.* **1**, 801-810.
- Piez, K. A., and Trus, B. L. (1977). *J. Mol. Biol.* **110**, 701-704.
- Piez, K. A., and Trus, B. L. (1978). *J. Mol. Biol.* **122**, 419-432.
- Prockop, D. J. (1990). *J. Biol. Chem.* **265**, 15,349-15,352.
- Ramachandran, G. N., and Ramakrishnan, C. (1976). In *Biochemistry of Collagen* (Ramachandran, G. N., ed.) Plenum Press, New York, pp. 45-84.
- Rich, A., and Crick, F. H. C. (1961). *J. Mol. Biol.* **3**, 483-506.
- Rich, A., and Crick, F. H. C. (1955). *Nature* **176**, 915-916.
- Ripamonti, A., Roveri, N., Braga, D., Hulmes, D. J. S., Miller, A., and Timmins, P. A. (1980). *Biopolymers* **19**, 965-975.
- Sakakibara, S., Kishida, Y., Okuyama, K., Tanaka, N., Ashida, T., and Kakudo, M. (1972). *J. Mol. Biol.* **65**, 371-373.
- Scheraga, H. A. (1984). *Carlsberg Res. Commun.* **49**, 1-55.
- Schmitt, F. O., and Gross, J. (1948). *J. Am. Leather Chem. Assoc.* **43**, 658-675.
- Smith, J. W. (1968). *Nature* **219**, 157-158.
- Squire, J. M., and Freundlich, A. (1980). *Nature* **288**, 410-413.
- Suzuki, E., Fraser, R. D. B., and MacRae, T. P. (1980). *Int. J. Biol. Macromol.* **2**, 54-56.
- SYBYL MENDYL, (1990). *Macromolecular Modeling Software*, version 5.32, TRIPOS Associates, Inc.
- Torbet, J., and Ronziere, M. C. (1984). *Biochem. J.* **219**, 1057-1059.
- Torchia, D. A., Hiyama, Y., Sarkar, S. E., and Sullivan, C. E. (1985). *Biopolymers* **24**, 65-75.
- Torchia, D. A., and Vanderhart, D. L., (1976). *J. Mol. Biol.* **104**, 315-321.
- Traub, W. (1978). *FEBS Lett.* **92**, 114-120.
- Trus, B. L., and Piez, K. A. (1976). *J. Mol. Biol.* **108**, 705-732.
- Umamura, S., Sakamoto, M., Hayakawa, R., and Wada, Y. (1979). *Biopolymers* **18**, 25-34.
- Weis, A., Miller, A., Leibovich, S. J., and Traub, W. (1979). *Biochim. Biophys. Acta* **576**, 88-98.
- Weis, A., and Yuan, L. (1975). *Biopolymers* **14**, 895-900.
- Weiner, P. K., and Kollman, P. A. (1981). *J. Comp. Chem.* **2**, 287-303.
- Weiner, S. J., Kollman, P. A., Case D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. A. (1984). *J. Am. Chem. Soc.* **106**, 765-784.
- Woodhead-Galloway, J. (1984). In *Connective Tissue Matrix: topics in Molecular and Structural Biology*, Vol. 5 (Hukins, D. W. L., ed.) Verlag Chemie, Weinheim, pp. 133-160.