Stabilization of the Triple-Helical Structure of Natural Collagen by Side-Chain Interactions†

Luigi Vitagliano,‡,§ George Némethy,§ Adriana Zagari,‡ and Harold A. Scheraga*,‖

Centro di Studio di Biocristallografia, Dipartimento di Chimica, Università di Napoli, via Mezzocannone 4, 80134 Napoli, Italy, Department of Biomathematical Sciences, Box 1023, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, New York 10029, and Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853

Received February 17, 1993; Revised Manuscript Received April 19, 1993

ABSTRACT: Conformational energy computations have been used to demonstrate that side-chain–backbone interactions contribute substantially to the stabilization of the triple-helical structure of collagen with a natural sequence. The minimum-energy conformation has been determined for a short triple-helical segment from the N-terminus of type I bovine skin collagen, containing 12 residues in each strand. In this conformation, the side chains of three Arg and four Met residues fold tightly against the triple-helical backbone, forming numerous atomic contacts with the neighboring strand. In addition, the polar groups of the three Arg and two Ser residues form hydrogen bonds with backbone carbonyl groups. The estimated total stabilization due to the side-chain interactions is about −50 kcal/mol out of a total interchain energy of −193.5 kcal/mol. The study presented here is the first application of conformational energy computations to a real sequence in the collagen molecule.

The amino acid sequence of fiber-forming collagens exhibits variations along the chain, with a local tripeptide amino acid composition of the type Gly-X-Y. The presence of various kinds of residues in positions X and Y has been shown to play an important role in the specificity of the longitudinal alignment and in the stabilization of fibrils (Hulmes et al., 1973; Hoffmann et al., 1978; Piez, 1984), and it may also have other functional significance (Piez, 1984). This investigation addresses the question of the energetic stabilization of the triple helix by interactions of the side chains with each other and with the backbone.

Our earlier conformational energy computations, reviewed elsewhere (Némethy & Scheraga, 1989), have been restricted mostly to the analysis of the structure and stability of triple helices formed by regular-sequence poly(tripeptide)s, viz., poly(Gly-Pro-Pro) (Miller & Scheraga, 1976; Némethy et al., 1992) which is the basic model for collagen, as well as poly(Gly-Pro-Hyp), poly(Gly-Pro-Ala), and poly(Gly-Ala-Pro) (Miller et al., 1980a,b; Némethy et al., 1980).[1] We have also demonstrated that the substitution of an individual Ala residue in poly(Gly-Pro-Pro), in either of the two positions, X or Y, does not cause any distortion of the triple helix (Némethy, 1981). The energetics of the local substitution of other residues in the triple helix has been investigated earlier mainly in terms of the conformational freedom of the side chains, and only in relation to the presence of Pro or Ala residues in neighboring locations (Némethy & Scheraga, 1982). The computation reported here is the first conformational energy computation of a collagen triple helix with a natural amino acid sequence.

This study deals with a fragment taken from type I bovine skin collagen, for which the amino acid sequence is known (Hoffmann et al., 1980). We have considered the structure formed by the first four tripeptides at the N-terminus of the triple helix. This segment is of particular interest because of our ongoing conformational energy computations on the N-terminal nonhelical telopeptide region of this collagen, to be reported elsewhere (L. Vitagliano, G. Némethy, A. Zagari, and H. A. Scheraga, unpublished results). The telopeptide chains are attached directly to the triple helix considered here, and their conformational freedom is influenced by this attachment. In addition, this sequence is representative of collagen in general, because it contains a variety of potentially interacting residues, including large and small, polar and nonpolar side chains.

The type I collagen molecule is composed of two identical $\alpha 1$ strands and an $\alpha 2$ strand, which are highly homologous. The molecular structure is customarily denoted (Hoffmann et al., 1980; Piez, 1984) as $[(\alpha 1)_2\alpha 2]$. The position occupied by the unique $\alpha 2$ strand is not known with certainty. It has been suggested that it may occur as the first strand, i.e., in position[2] A (Bender et al., 1982; Piez & Trus, 1978) or in position B (Hoffmann et al., 1978; Traub & Fietzek, 1976). All of these proposals have been based on marginal preferences in analyses of observed sequence regularities or of distributions of neighboring residues in adjacent strands, and they are therefore not definitive. Most of the computations reported here have been carried out with the $\alpha 2$ chain in position A. Two tests have been made with the $\alpha 2$ chain placed in positions B and C, respectively, in order to check whether such a change affects the energy of the triple helix and hence whether the sequence difference between the $\alpha 1$ and $\alpha 2$ chains might play a structural role in this fragment.

† This work was supported by research grants from the National Institute on Aging (AG-00322) and the National Institute of General Medical Sciences (GM-14312), a small instrument grant (1SI5DK44714) from the National Institutes of Health, U.S. Public Health Service, and grants from the National Science Foundation (DMB90-15815 and INT-9004111) and the Italian CNR (91/3474CT03).

* To whom correspondence and reprint requests should be addressed.
‡ Università di Napoli.
§ Mount Sinai School of Medicine.
‖ Cornell University.

[1] Abbreviations: ECEPP, empirical conformational energy program for peptides; GAA, poly(Gly-Ala-Ala) triple helix; GPA, poly(Gly-Pro-Ala) triple helix; GPP, poly(Gly-Pro-Pro) triple helix; Hyp, 4-hydroxyproline; RS, regular triple helix; TS, truncated triple helix.

[2] Positions A, B, and C refer to the staggering of the three strands in the coiled-coil collagen structure with screw symmetry, as indicated in the sequence scheme in the Methods section.

0006-2960/93/0432-7354$04.00/0

The computations have been carried out by fitting the amino acid sequence under consideration onto the structure of a regular poly(Gly-Pro-Pro) or poly(Gly-X-Y) triple helix, where X is either Pro or Ala and Y is Ala, and minimizing the energy of the fitted structure. Energy minimization has also been carried out for various choices of side-chain conformations.

METHODS

Computations have been carried out using the updated version (ECEPP/3) (Némethy et al., 1992) of the ECEPP (empirical conformational energy program for peptides) algorithm[1] (Momany et al., 1975), which keeps bond lengths and bond angles fixed and varies only the dihedral angles. The general unconstrained optimization algorithm SUMSL (secant unconstrained minimization solver) was used for energy minimization (Gay, 1983). The recently computed regular triple-helical structure (Némethy et al., 1992) for CH_3CO-(Gly-Pro-Pro)$_4$-$NHCH_3$, denoted earlier as structure RS, has been used as the starting point for the computations. The three identical strands of RS are arranged according to screw symmetry (Hulmes et al., 1973; Miller & Scheraga, 1976; Miller et al., 1980a; Piez, 1984).

Regular-Sequence Poly(tripeptide)s. As the preliminary stage, preceding the computation for the natural collagen sequence, three identical regular-sequence poly(tripeptide) chains, CH_3CO-(Gly-X-Y)$_4$-$NHCH_3$, were fitted to the RS structure.[1] X-Y was selected either as Ala-Ala or as Pro-Ala. Ala-Hyp (or Ala-Pro) has not been considered, because no Gly-X-Hyp tripeptide occurs in the sequence studied here. The use of the fitting procedure is necessary when another amino acid is substituted in the place of Pro, because there are small differences in some backbone bond lengths and bond angles of various amino acids in the standard residue data set used in ECEPP (Momany et al., 1976; Némethy et al., 1992). For the sake of brevity, the triple helices with the regular sequences indicated will be referred to here as structures[1] GPP (referred to also as RS), GPA, and GAA, respectively.

The fitting has been carried out in two steps for each sequence. Details of the procedure are given elsewhere (Vitagliano, 1993). First, the backbone of the poly(Gly-X-Y) sequence was superposed on the RS structure of GPP (Némethy et al., 1992), allowing *all* dihedral angles to vary. The backbone atoms N, C^{α}, C′, and O were used in the least-squares fitting. The values of the dihedral angles in the various tripeptides along the chains (after optimal fitting) were within ±2° of each other in the fitted structures. For use in the next step, however, the fitted structures were rendered fully regular by averaging each dihedral angle over the entire structure, in order to avoid the introduction of any end effects during the fitting. In the second step, a grid search was carried out, starting from the regularized dihedral angles obtained in the first step. In the grid search, the variable dihedral angles were $(\phi,\psi)_{Gly}$, $(\phi,\psi)_{Ala,X}$, and $(\phi,\psi)_{Ala,Y}$ for GAA and $(\phi,\psi)_{Gly}$, $(\psi)_{Pro,X}$, and $(\phi,\psi)_{Ala,Y}$ for GPA, while ω was held fixed. The grid for each dihedral angle consisted of three points, taken at the value in the regularized structure and at angles differing from it by ±5°. Thus, the total number of grid points used was $3^6 = 729$ for GAA and $3^5 = 243$ for GPA. At every grid point, an energy minimization was carried out with respect to all variable dihedral angles, constraining the corresponding dihedral angles of each chain to be the same, according to the procedure described earlier (Miller & Scheraga, 1976). In the case of GPA, a second grid search and minimization were carried out around the lowest-energy grid point reached in the first search, in order to increase the number of grid points tested.

The minimum-energy structure for GPA obtained in this grid search deviated considerably more from RS than structure GAA, as will be shown under Results and Discussion. A possible explanation for the difference in the behavior of the two sequences is the greater flexibility of a poly(Gly-Ala-Ala) chain, as compared to poly(Gly-Pro-Ala), caused by the absence of the Pro residues. The higher flexibility may permit a closer adjustment of the strands to RS.

The Natural-Sequence Collagen Fragment. The glycine residues in adjacent strands of collagen are staggered longitudinally by one residue. The triple helix starts with a Gly residue of the strand that is in position A. Consequently, one or two non-glycyl residues of strands 2 and 3, respectively, usually listed as parts of the telopeptide sequence, are actually located in the triple-helical conformation. Throughout most of this study, chain α2 has been placed in position A. Thus, the amino acid sequence of the triple helix considered here was

Gly-Pro-Met-Gly-Leu-Met-Gly-Pro-Arg-Gly-Pro-Hyp-	α2	Position A
Pro-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-Gly-Leu-	α1	Position B
Val-Pro-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-Gly-	α1	Position C

Each polypeptide chain was terminally blocked, using the CH_3CO– and –$NHCH_3$ end groups. The Arg side chains were uncharged.[3]

Starting points for the energy minimization were obtained by fitting the polypeptide backbone of the natural sequence to each of the energy-minimized regular-sequence triple helices GPP (RS), GPA, and GAA. During every energy minimization of the natural-sequence triple helix, *all* dihedral angles of each strand, and also the external variables (translational vector components and Euler angles) that specify the relative position of the three strands, were allowed to vary independently (Fossey et al., 1991). For a given choice of side-chain conformational states, the same minimum was reached when the minimization was started from each of the four regular-sequence triple helices. This test was carried out for two choices of side-chain conformations, and it indicates that the results are not sensitive to details of the fitting procedure.

The possible occurrence of side-chain–backbone and side-chain–side-chain interactions has been tested by energy minimization starting from every staggered rotameric state of the side chain of each Arg, Met, Leu, and Ser residue. For Val and Hyp, only one conformation, viz., the lowest-energy one in the triple helix (Némethy & Scheraga, 1982), has been used. There are only few conformations in which interaction between two side chains is possible, because of the spread-out locations of the side chains along the triple helix. Therefore, the exploration of side-chain conformations was carried out by varying the conformation of the side chains mentioned one at a time, with the other side chains placed in a low-energy rotameric state prior to minimization. In order to carry out exhaustive testing for possible side-chain–side-chain hydrogen bonding, an additional set of energy minimizations was carried out in which the starting dihedral angles of pairs of neighboring polar side chains were varied simultaneously. The starting conformations for the latter minimization were those rotameric

[3] It has been shown previously that the use of the neutral form of ionizable groups in the computations, in combination with the low dielectric constant used in ECEPP, adequately approximates the effects of solvent shielding and counterion effects and that computed conformations reproduce experimental data closely (Howard et al., 1973).

Table I: Minimum-Energy Conformations of Triple Helices

sequence[a]	interchain energy[b] E_{inter} (kcal/mol)	helical parameters D^c (Å)	Θ^d (deg)
GPP (structure RS)[e,f]	−126.33	8.78	41
GAA[f]	−117.34	8.66	43
GPA[f]	−131.79	8.17	62
natural-sequence fragment[g,h]	−193.54	8.46[i]	46[i]
GPP (structure TS)[e,h]	−144.61	8.73[i]	50[i]
obsd collagen structure[j]		8.61	36

[a] See the text for the notation. [b] Total value for the three strands. [c] Translational repeat per tripeptide (Miller & Scheraga, 1976). [d] Rotational repeat per tripeptide (Miller & Scheraga, 1976). [e] Némethy et al., 1992. [f] Minimized with corresponding dihedral angles in each chain constrained to be the same. [g] The side chains are located in the energetically most favorable conformation. The triple helices are truncated; i.e., they have blunt ends. [h] Minimized with respect to all dihedral angles of each chain. [i] Averaged over the entire triple helix. See footnote 4. [j] According to the model derived by Traub et al. (1969).

Table II: Root-Mean-Square Deviations (Å) for the Backbone of the Minimum-Energy Conformations of the Triple Helices[a]

	GAA	GPA	natural-sequence fragment
GPP (structure RS)	0.27	1.43	0.74
GAA		1.36	0.64
GPA			0.85

[a] Computed for the backbone heavy atoms (N, C^α, C′, O). See Table I for the names of the structures.

states in which the distance between a hydrogen-bond donor (O or N) of the side chain and an oxygen atom in another side chain was <5.0 Å, as determined from the atomic coordinates, prior to minimization.

In summary, the procedure developed here for the energy minimization of real-sequence triple helices is as follows. (1) The backbone atoms of the specific-sequence chain are fitted to the backbone of an energy-minimized regular-sequence triple helix. (2) The energy of the triple helix is minimized with respect to all dihedral angles and external variables. (3) Step 2 is repeated for various starting sets of the side-chain conformations, selected as follows: (a) the dihedral angles of side chains that are not near other flexible side chains are varied one at a time; (b) in the case of a pair of side chains that are near each other and hence may interact with each other, their dihedral angles are varied simultaneously.

RESULTS AND DISCUSSION

Regular-Sequence Poly(tripeptide)s. The helical parameters and the interchain energy of the energy-minimized conformations are listed in Table I. The intrachain energy and hence the total energy are not listed, because the numerical values of the intrachain energy cannot be compared for different sequences (Momany et al., 1976; Némethy et al., 1992). The structures are compared to each other and to the reference structure RS in terms of the rms deviation of the backbone heavy atoms (N, C^α, C′, and O) in Table II. The GPP and GAA structures are very close to each other, as indicated by the similarity of their helical parameters and by the low value of the rms difference of the backbone coordinates, with somewhat higher deviations for the GPA structures. All interchain energies, E_{inter}, are comparable in magnitude, although E_{inter} for GAA is somewhat higher than for the other sequences, because there is little nonbonded interaction between the Ala side chain in adjacent residues.

The Natural-Sequence Collagen Fragment. The energy and the helical parameters of the lowest-energy structure are given in Table I. The parameters (Chou et al., 1984) are referred to a helical axis determined from the coordinates of the C^α atoms by minimizing the sum of the squares of all of these atoms from the axis (Åqvist, 1986).[4] The translational repeat D is near the value in collagen models derived from observations (Ramachandran & Kartha, 1955; Rich & Crick, 1955; Traub et al., 1969). The difference is comparable to the variation in D for various observed and computed models (Miller & Scheraga, 1976; Némethy et al., 1992; Ramachandran & Kartha, 1955; Rich & Crick, 1955; Traub et al., 1969). The angular repeat Θ is within the range of values (36–51°) computed for the regular GPP triple helix (Miller & Scheraga, 1976; Némethy et al., 1992) and deduced from various observed models (Ramachandran & Kartha, 1955; Rich & Crick, 1955; Traub et al., 1969). It is notable that Θ is generally less precisely established in these models than D, viz., that Θ varies over a wider range than does D even for otherwise closely similar triple helices, because its numerical value depends sensitively on local details of the backbone conformation, such as small changes in individual dihedral angles (Miller & Scheraga, 1976; Némethy et al., 1992).

The rms deviation of the natural-sequence N-terminal fragment with respect to any of the regular structures considered is <0.85 Å (Table II). These deviations are smaller than those between the GPA structure and the other regular structures. Thus, the structure of the triple helix in the N-terminal fragment falls between the regular-sequence model structures, seen also from a comparison of the helical parameters (Table I). The small rms deviation (0.74 Å) between the fragment and the GPP structure suggests, in particular, that poly(Gly-Pro-Pro) is a good model system for the representation of a real-sequence triple helix, in confirmation of the widespread use of this poly(tripeptide) as a model in numerous experimental [reviewed by Bhatnagar and Rapaka (1976) and Némethy (1988)] and theoretical [reviewed by Némethy (1988) and Némethy and Scheraga (1989)] studies of collagen. The backbone N—H···O=C hydrogen-bonding pattern of the fragment is identical to that for GPP (Miller & Scheraga, 1976).

The interchain energy of the natural-sequence fragment is much lower than that of the regular-sequence model structures. Some of this energy difference arises because all dihedral angles were allowed to relax in the fragment, in contrast to the model structures (see Methods). Most of the difference can be attributed, however, to favorable interactions involving the side chains. Nonbonded and hydrogen-bonded interactions of a side chain with the backbone or a side chain in an adjacent strand provide additional stabilization of the triple helix, because all or part of the long side chains of Arg and Met residues fold against the backbone in several low-energy conformations, including the lowest-energy structure. The side-chain dihedral angles in the lowest-energy conformation are listed in Table III.

For example, all three Arg side chains are in the ttttc conformational state in the lowest-energy conformation of

[4] In our previous study (Némethy et al., 1992), the helical parameters were computed from "regularized" triple helices fitted to the energy-minimized structure, and the fitting was done by averaging the dihedral angles over the entire structure [cf. Table XII of Némethy et al. (1992)]. This procedure was used for the sake of comparison with earlier computations in which the triple helices were constrained to be regular (Miller & Scheraga, 1976; Miller et al., 1980a,b; Némethy & Scheraga, 1982; Némethy et al., 1980). In the present work, helical parameters were computed directly from the C^α coordinates, with no "regularization" of the triple helix. The use of the two methods may occasionally cause small differences in the numerical values of the helical parameters.

Table III: Side-Chain Dihedral Angles (deg) in the Lowest-Energy Structure of the N-Terminal Fragment

strand	residue	χ^1	χ^2	χ^3	χ^4	χ^5	$\chi^{6,1}$	$\chi^{6,2}$
A	Met	−169	165	−162	63			
	Leu	−177	68	54[a]	58[a]			
	Met	−174	166	−160	65			
	Arg	−174	170	−172	−159	0	180	−1
	Hyp	−176						
B	Met	−174	165	−161	63			
	Ser	179	65					
	Arg	−172	167	−171	−178	3	180	0
	Leu	178	64	54[a]	59[a]			
C	Val	−70	65[b]	55[b]				
	Met	−170	169	−171	62			
	Ser	175	−75					
	Arg	−172	170	−177	167	−1	180	0

[a] $\chi^{3,1}$, $\chi^{3,2}$. [b] $\chi^{2,1}$, $\chi^{2,2}$.

the triple helix, where they form numerous close atomic contacts (Figure 1). On the other hand, if the side chain is placed, e.g., in the g⁻ttc conformation, it extends away from the backbone and it is involved in very few close contacts. Both of these side-chain conformations have similarly low energy in a terminally-blocked single Arg residue with the collagen-like backbone conformation (Vásquez et al., 1983) [denoted as conformational state F (Zimmerman et al., 1977)], but the energy of the ttttc conformation is much lower in the triple helix, e.g., by 7.9 kcal/mol for the Arg in strand A. The decrease in energy arises in part from a hydrogen bond (described below) and in part from interchain nonbonded interactions, mostly with the neighboring strand B. The energy contribution of the side-chain interactions has also been estimated directly, by deleting the Arg side chains one by one, i.e., by removing all atoms beyond the C^β and replacing $C^\beta H_2$ with a methyl group, followed by reminimization of the energy of the triple helix. The interchain energy is always considerably higher for the structure with the deleted side chain. If the side chain of Arg in strand A is deleted, the interchain energy is increased by 8.0 kcal/mol. A similar energy difference occurs when the Arg side chain of one of the other two strands is deleted.

In a similar manner, all Met side chains are in the ttt conformational state in the lowest-energy triple-helical structure, and they interact strongly with the triple helix. Deletion beyond the C^β raises the energy by 5.6–6.4 kcal/mol. The preferred conformation could not have been predicted from a consideration of the terminally-blocked single Met residue (Vásquez et al., 1983), because the latter has no low-energy conformations with the C^α–C^β bond in the t conformational state when the backbone is in state F. This comparison serves as an additional indication that the preference for the ttt conformation arises from interactions in the triple helix.

The lowest-energy conformation for both Leu side chains is tg^+. In the case of the Leu in strand A, several other side-chain conformations have a low energy ($\Delta E < 1.0$ kcal/mol), because they can also interact with residues in the neighboring strands. No such interactions occur for the Leu in strand B, because it is at the carboxy terminus of the fragment.

The polar side chains are engaged in several hydrogen bonds in the lowest-energy conformation. The OH of the Ser side chain in strand B, in a tg^+ conformation, forms a hydrogen bond to the backbone O═C of the Ser in strand C. On the other hand, the side chain of Ser in strand C is in conformation tg^-, and its OH forms a hydrogen bond to the O═C of the succeeding Gly in the same strand. If these hydrogen bonds are broken, by rotating the side chain into the g^+g^- conformation, the energy is increased by 1.9 and 1.4 kcal/mol,

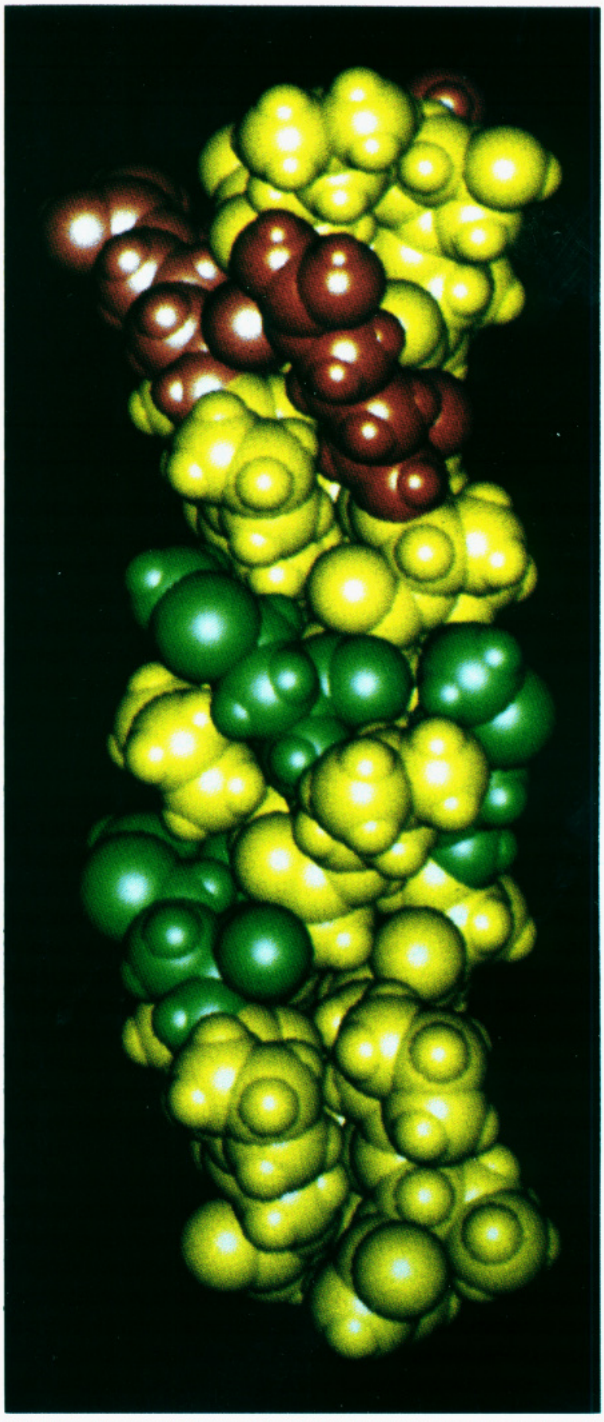

FIGURE 1: Space-filling model of the N-terminal triple helix of the natural-sequence collagen fragment in the lowest-energy computed conformation. The N-terminus of the polypeptide chains is at the bottom; the C-terminus, at the top. The Arg residues are colored red; the Met residues, green; and the rest of the structure, yellow. The Arg residue of the strand in position A (shown in a brighter shade of red) is located in the upper center of the figure. The large side chains fold against the triple-helical backbone, as discussed in the text.

respectively. All three arginines form an $N^\epsilon H\cdots O{=}C$ hydrogen bond to a backbone carbonyl group in the next strand. The locations of these hydrogen bonds are listed in Table IV, together with the distances from the acceptor O atom to the H atom and to the (O or N) hydrogen donor atom. The H···O distance is about 1.9 Å in all of them, a distance which is close to the unperturbed value of 1.80 Å for this bond length (Momany et al., 1975), and the bonds are nearly linear. The only exception is the hydrogen bond involving the Ser in strand

Table IV: Side-Chain–Backbone Hydrogen Bonds in the Lowest-Energy Structure of the N-Terminal Fragment

hydrogen bond	donor		acceptor		atomic distance d(H···O) (Å)	atomic distance d(D···O)[a] (Å)
	strand	residue	strand	residue		
$N^{\epsilon}H\cdots O{=}C$	A	Arg 9	B	Arg 10	1.90	2.90
$N^{\epsilon}H\cdots O{=}C$	B	Arg 10	C	Arg 11	1.92	2.90
$N^{\epsilon}H\cdots O{=}C$	C	Arg 11	A	Hyp 12	1.93	2.93
$O^{\gamma}H\cdots O{=}C$	B	Ser 7	C	Ser 8	1.93	2.76
$O^{\gamma}H\cdots O{=}C$	C	Ser 8	C	Gly 9	1.99	2.96

[a] D = donor atom (N^{ϵ} for Arg, O^{γ} for Ser).

B, which is stretched by 0.2 Å and is slightly bent.

Side-chain–backbone hydrogen bonds can also be formed in some other conformations of the Arg side chains, but the total energy is at least 3 kcal/mol above the energy of the most favorable conformation discussed above. Therefore, these conformations do not contribute to the stability of the triple helix in the sequence considered here. No side-chain–side-chain hydrogen bonds occurred in any low-energy conformation in this sequence. The possibility of interactions between polar side chains in collagen has also been pointed out in a recent computation (Katz & David, 1990), in which the formation of salt bridges between charged Arg and Glu or Asp side chains was considered, in a collagen-like conformation of a *single* polypeptide chain. Similar interchain salt bridges have been computed for the triple helix (Katz & David, 1992). The sequence studied here, however, does not contain any anionic side chains.

In summary, the lowest-energy triple helix (line 4 of Table I) is stabilized by interactions involving mainly the Arg, Ser, and Met side chains. The structure is shown in Figure 1. The folding of the Arg side chain of strand A against the backbone is seen clearly at the center of the figure. The energy can be compared with that of structure[1] TS (line 5 of Table I) of the poly(Gly-Pro-Pro) triple helix, which was also minimized by allowing all dihedral angles to vary independently (Némethy et al., 1992). It can be estimated from the comparison of the interchain energies in Table I that the side-chain interactions contribute about −50 kcal/mol, out of a total of −193.54 kcal/mol, to the energy of stabilization of this triple helix.

Permutation of the Placement of the α2 Chain. In all of the computations described so far, the α2 chain has been placed in position A. In order to test any possible structural role of this chain in this fragment, the energy was also minimized for triple helices in which the α2 chain is in positions B and C, respectively, with the initial side-chain conformations as listed in Table III. The residues at the N-terminus differed for these triple helices, for the reason discussed in the Methods section. When the α2 chain was in position B, the sequence of each strand was

```
-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-Gly-Leu-Hyp-   α1   Position A
-Pro-Gly-Pro-Met-Gly-Leu-Met-Gly-Pro-Arg-Gly-Pro-   α2   Position B
-Val-Pro-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-Gly-   α1   Position C
```

When the α2 chain was in position C, the sequence was

```
-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-Gly-Leu-Hyp-   α1   Position A
-Pro-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-Gly-Leu-   α1   Position B
-Gly-Pro-Gly-Pro-Met-Gly-Leu-Met-Gly-Pro-Arg-Gly-   α2   Position C
```

The interchain energies are −193.54, −187.90, and −188.30 kcal/mol when the α2 chain is placed in positions A, B, and C, respectively. The energy is somewhat lower for position A, suggesting that this position might be preferred energetically for the α2 chain, as a result of more favorable side-chain interactions. On the other hand, some of the energy differences arise because the Arg residues are closer to the C-terminus of the short fragment in the permuted sequences, and hence their interactions with the rest of the triple helix are weaker. The result obtained here does not allow us to draw conclusions concerning the possible energetic stabilization of this position for the entire collagen molecule. It demonstrates, however, that such side-chain interactions could play a role in the preferred positioning of the α2 chain, as suggested in an earlier investigation (Traub & Fietzek, 1976).

CONCLUSIONS

In this first application of conformational energy computations to a real sequence in a collagen molecule, a general procedure has been developed for the efficient survey and energy minimization of real-sequence triple helices. The method introduced here can be used in studies of other sequences in collagen.

The computations have demonstrated that the model triple helix (Miller & Scheraga, 1976; Némethy et al., 1992), generated using regularly repeating poly(tripeptide)s such as poly(Gly-Pro-Pro), is not distorted by the substitution of residues containing various side chains. Furthermore, these substitutions do not result in an increase in the interchain energy, indicating that the three strands can pack efficiently, irrespective of the actual sequence. Therefore, the results confirm the validity of the use of simple poly(tripeptide)s as model compounds in studies of collagen.

Large side chains, such as those of Arg and Met, can be accommodated on the surface of the triple helix in such a way that they interact with a neighboring strand by means of nonbonded and (in the case of Arg) hydrogen-bonded interactions. It is notable that Arg is one of the most frequently occurring amino acids in type I collagen (Hulmes et al., 1973; Hoffmann et al., 1978; Piez, 1984). Our results suggest that its presence contributes significantly to the stabilization of the triple-helical structure, in addition to any other possible functional role that may be played by this residue in collagen. Even a small side chain, like that of Ser, can form intra- or interstrand hydrogen bonds. Thus, side-chain interactions can make significant contributions to the stabilization of the collagen triple helix.

ACKNOWLEDGMENT

We thank Drs. G. Avitabile and G. Paterlini for useful discussions and for help with some of the computations. Some of the computations were carried out using the Cornell National Supercomputer Facility, a resource of the Center for Theory and Simulation in Science and Engineering at Cornell University, which is funded in part by the National Science Foundation, New York State, the IBM Corp., and members of its Corporate Research Institute. Facilities made available by CE.IN.GE. (Napoli) were used for part of the computations.

REFERENCES

Åqvist, J. (1986) *Comput. Chem. 10*, 97–99.

Bender, E., Silver, F. H., Hayashi, K., & Trelstad, R. L. (1982) *J. Biol. Chem. 257*, 9653–9657.

Bhatnagar, R. S., & Rapaka, R. S. (1976) in *Biochemistry of Collagen* (Ramachandran, G. N., & Reddi, A. H., Eds.) pp 479–523, Plenum Press, New York.

Chou, K.-C., Némethy, G., & Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 3161–3170.
Fossey, S. A., Némethy, G., Gibson, K. D., & Scheraga, H. A. (1991) *Biopolymers 31*, 1529–1541.
Gay, D. M. (1983) *ACM Trans. Math. Software* 9, 503–524.
Hofmann, H., Fietzek, P. P., & Kühn, K. (1978) *J. Mol. Biol. 125*, 137–165.
Hofmann, H., Fietzek, P. P., & Kühn, K. (1980) *J. Mol. Biol. 141*, 293–314.
Howard, J. C., Momany, F. A., Andreatta, R. H., & Scheraga, H. A. (1973) *Macromolecules 6*, 535–541.
Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A., & Woodhead-Galloway, J. (1973) *J. Mol. Biol. 79*, 137–148.
Katz, E. P., & David, C. W. (1990) *Biopolymers 29*, 791–798.
Katz, E. P., & David, C. W. (1992) *J. Mol. Biol. 228*, 963–969.
Miller, M. H., & Scheraga, H. A. (1976) *J. Polym Sci., Polym. Symp. 54*, 171–200.
Miller, M. H., Némethy, G., & Scheraga, H. A. (1980a) *Macromolecules 13*, 470–478.
Miller, M. H., Némethy, G., & Scheraga, H. A. (1980b) *Macromolecules 13*, 910–913.
Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361–2381.
Némethy, G. (1981) *Biochimie 63*, 125–130.
Némethy, G. (1988) in *Collagen* (Nimni, M. E., Ed.) Vol. I, pp 79–94, CRC Press, Boca Raton, FL.
Némethy, G., & Scheraga, H. A. (1982) *Biopolymers 21*, 1535–1555.
Némethy, G., & Scheraga, H. A. (1989) *Bull. Inst. Chem. Res., Kyoto Univ. 66*, 398–408.
Némethy, G., Miller, M. H., & Scheraga, H. A. (1980) *Macromolecules 13*, 914–919.
Némethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S., & Scheraga, H. A. (1992) *J. Phys. Chem. 96*, 6472–6484.
Piez, K. A. (1984) in *Extracellular Matrix Biochemistry* (Piez, K. A., & Reddi, A. H., Eds.) Chapter 1, Elsevier, New York.
Piez, K. A., & Trus, B. L. (1978) *J. Mol. Biol. 122*, 419–432.
Ramachandran, G. N., & Kartha, G. (1955) *Nature 176*, 593–595.
Rich, A., & Crick, F. H. C. (1955) *Nature 176*, 915–916.
Traub, W., & Fietzek, P. P. (1976) *FEBS Lett. 68*, 245–249.
Traub, W., Yonath, A., & Segal, D. M. (1969) *Nature 221*, 914–917.
Vásquez, M., Némethy, G., & Scheraga, H. A. (1983) *Macromolecules 16*, 1043–1049.
Vitagliano, L. (1993) Doctorate Thesis, Università di Napoli, Italy (in preparation).
Zimmerman, S. S., Pottle, M. S., Némethy, G., & Scheraga, H. A. (1977) *Macromolecules 10*, 1–9.