The Role of Hydroxyproline in Collagen Folding: Conformational Energy Calculations on Oligopeptides Containing Proline and Hydroxyproline

M. BANSAL, *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India,* and V. S. ANANTHANARAYANAN,* *Department of Biochemistry, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X9*

Synopsis

We have observed that the rate of folding of the enzymatically hydroxylated form of poly(Gly-Pro-Pro) into the triple-helical conformation is considerably higher than that of the unhydroxylated polypeptide [R. K. Chopra and V. S. Ananthanarayanan (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7180–7184]. In this study, we examine a plausible kinetic pathway for triple-helix formation by selecting peptide models for the unhydroxylated collagen molecule, and computing their conformational energies before and after proline hydroxylation. Starting with the available data on the preferred conformations of proline- and hydroxyproline-containing peptide sequences, energy minimization was carried out on the following pairs of peptides: Gly-Ala-Pro-Gly-Ala and Gly-Ala-Hyp-Gly-Ala; Gly-Pro-Pro-Gly-Ala and Gly-Pro-Hyp-Gly-Ala; Gly-Ala-Pro-Gly-Ala-Pro and Gly-Ala-Hyp-Gly-Ala-Hyp. It was found that, with each pair of peptides, the energetically most favorable conformation (I) has an extended structure at the Gly-Ala or Gly-Pro segment and a β-bend at the Pro-Gly or Hyp-Gly segment. In the Hyp-containing peptides, this conformation is further stabilized by a $(Hyp_{i+2})OH \ldots OC(Gly_i)$ hydrogen bond. Conformation I is lower in energy by about 6–13 kcal/mol of the peptide than the fully extended conformations that resemble the single collagen polypeptide chain and contain no *intra*molecular hydrogen bond. In contrast to the proline counterpart, the hydroxyproline-containing peptides are found capable of adopting a partially extended conformation that does not contain the β-bend but retains the $(Hyp)OH \ldots OC(Gly)$ hydrogen bond. The energy of this conformation is intermediate between conformation I and the fully extended conformation. The continuation of the β-bend along the chain is restricted by stereochemical constraints that are more severe in the latter two pairs of peptides than in the first pair. Such a restriction may be considered to trigger the "unbending" of the minimum energy conformation leading to its straightening into the fully extended conformation; the latter, in turn, would lead to triple-helix formation through favorable *inter*chain interactions. We propose that the partially extended conformation in the Hyp-containing peptides could serve as a kinetic intermediate on the way to forming the fully extended conformation. Because of the $(Hyp_{i+2})OH \ldots OC(Gly_i)$ hydrogen bond, this conformation would also serve to lock the *trans* geometry at the Gly-Ala(Pro) and Ala(Pro)-Hyp peptide bonds, thereby enhancing the rate of their helix formation. A scheme for collagen folding in proposed on the basis of these results.

INTRODUCTION

It is well established from experimental and theoretical studies that 4-hydroxy-L-proline (Hyp) is essential for the thermodynamic stability of the

---

*To whom correspondence should be addressed.

triple-helical conformation of collagen.[1-4] However, the role of Hyp in the kinetics of triple helix formation has not been fully appreciated. Recently, while investigating the conformational consequence of the enzymic hydroxylation of proline residues in collagen, we made an observation that brought out the importance of Hyp in enhancing the rate of folding of the triple helix.[5,6] Using synthetic polypeptide substrates, we found that the hydroxylated form of poly(Gly-Pro-Pro) refolded into the triple-helical conformation from an initially unfolded state at a remarkably faster rate than the unhydroxylated polypeptide.[5] While similar kinetic data on the unhydroxylated and hydroxylated forms of the collagen molecule itself are lacking, it is worthwhile at this point to examine the conformational basis for the observed role of Hyp in triple-helix formation.

In this article, an attempt is made to describe a plausible kinetic pathway for triple-helix formation by examining peptide models for the unhydroxylated collagen polypeptide chain and following changes in their conformational energies after proline hydroxylation and subsequent chain assembly to form the triple-helical molecule. In selecting the appropriate peptide models for our purpose, the following considerations were made: First, the sequences containing the Hyp residue in collagen are known to be Gly-X-Hyp and Gly-Pro-Hyp, their proline counterparts being Gly-X-Pro and Gly-Pro-Pro. (X is any residue other than Pro or Hyp.) Studies on polytripeptide models with these sequences reveal that those containing the Hyp residue are thermodynamically more stable than their proline analogs.[4,5,7] Second, our recent studies[8,9] on the conformational requirement for proline hydroxylation in collagen by prolyhydroxylase have shown that the minimum sequences of the peptide substrate is Gly-X-Pro-Gly-Y, where X and Y can be any residue including Pro. We have, in addition, shown that for hydroxylation to take place the above pentapeptide sequence should be in the so-called PP-II + β-bend conformation[8,9] (which corresponds to an extended polyproline-II type structure at the Gly-X-Pro fragment followed by a β-bend in the X-Pro-Gly-Y fragment). The β-bend in this conformation would normally be stabilized by an *intra*molecular hydrogen bond between the amino group of the Y residue and the carbonyl group of X. Even though such a hydrogen bond is not feasible when Y is Pro (as in the X-Pro-Gly-Pro or Pro-Pro-Gly-Pro sequences), the β-bend conformation at the Pro-Gly segment is still favorable, as has been shown by us from theoretical calculations on N-acetyl-Pro-Pro-Gly-Pro-NHCH$_3$. In light of the above considerations, we have selected for our conformational energy calculations the following pairs of peptide sequences: Gly-Ala-Pro-Gly-Ala and Gly-Ala-Hyp-Gly-Ala; Gly-Pro-Pro-Gly-Ala and Gly-Pro-Hyp-Gly-Ala; Gly-Ala-Pro-Gly-Ala-Pro and Gly-Ala-Hyp-Gly-Ala-Hyp. As shown below, a study of the preferred conformations of these representative sequences reveals the role of the hydroxylation of the Pro residue in governing the ease of formation of the triple helix.

METHODS

Preliminary energy calcalations were carried out using the standard residue geometries as given by Momany et al.[10] and the nonbonded parameters as given by Ramachandran and Sasisekharan.[11] The nonbonded interactions

were calculated using the 6–12 Lennard–Jones potential function while a monopole approximation was used for the electrostatic energies.[12] Three possible geometries were considered for the pyrrolidine ring, the "down" and "up" conformations as defined by Momany et al.[10] with ϕ_{Pro} values of $-75°$ and $-65°$, respectively, and the conformation as observed by neutron diffraction studies on the single crystal of 4-hydroxy-L-proline[13] where $\phi_{Pro} \sim -57°$. Various combinations of conformations that have earlier been shown to be preferred by the sequences Gly-Pro-X-Gly,[14] Gly-X-Pro-Gly[15] and Gly-Pro-Pro-Gly,[16,17] and Gly-Pro-Hyp-Gly,[18] as well as the reported conformations of poly(Hyp) from fiber diffraction studies,[19] were used as the starting conformations. These preliminary calculations clearly indicated that conformations containing a β-bend at the Pro(Hyp)-Gly segment were the most preferred and that considerable energy must be expended for the chain to unbend and form the triple helix. A detailed energy minimization was therefore carried out using 10 different starting conformations. Some of the conformations have a β-bend at the Pro-Gly sequence, while others are partially or fully extended. The minimization was carried out using the refinement program AMBER (Assisted Model Building with Energy Refinement) developed by Kollman and coworkers.[20,21] The latest all-atom force field potential function was used, in which the hydrogens are explicitly included and a 10–12 function is used to calculate the hydrogen-bond energy.[22] Since the charges for the Hyp residue were not available in the AMBER data base, they were calculated following the same procedure as was used for calculating the charges on the other amino acids[21,22] in which the charges were derived from an electrostatic potential fit to a point-charge model using a STO-3G basis set. A distance-dependent dielectric constant was used, consistent with the calculations of Kollman and co-workers[20] which indicate that this is most appropriate when solvent molecules are not included explicitly. An N-acetyl blocking group was attached at the beginning and an $NHCH_3$ group at the end of the peptide sequences being minimized. No condition was imposed regarding the tripeptide repeat in the structure; that is, the ϕ, ψ values, for example, of the Gly_1-X_2 segment, need not be the same as those of the Gly_4-Y_5 segment in Gly_1-X_2-$Pro(Hyp)_3$-Gly_4-Y_5. Although such an exact repeat may be expected to occur in a polymer chain on triple-helix formation, it is probable that local regions in the nascent (pro)collagen chain (prior to hydroxylation) would prefer quite different conformations. This is evident from the observed variations in the extent of enzymic proline hydroxylation in different parts of the collagen chain.[23]

RESULTS AND DISCUSSION

Gly-Ala-Pro-Gly-Ala and Gly-Ala-Hyp-Gly-Ala Peptides

Ten different starting conformations were selected (see the methods section) for these peptides and energy minimization was carried out on them. Those structures that, after minimization, differed by less than 0.2 kcal/mol in energy and by less than 2° in torsion angles were treated as identical and are not listed separately. For example, insignificant differences were found for the converged structures when the starting proline ring was in the up conforma-

TABLE I
Computed Conformations of Gly-Ala-Pro-Gly-Ala and Gly-Ala-Hyp-Gly-Ala[a]

Conformation	Residue[b]	Pro peptide			Residue	Hyp peptide		
		ϕ^c	ψ	ω		ϕ^c	ψ	ω
I	Gly	81	−51	173	Gly	81	−53	173
	Ala	−67	130	182	Ala	−40	−53	197
	Pro	−61(−34)	121	187	Hyp	−55(−32)	130	201
	Gly	109	−49	180	Gly	117	−59	173
	Ala	−78	64	180	Ala	−77	65	180
		$E_{tot} = 19.3$				$E_{tot} = -26.2$		
		$\Delta E = 0$				$\Delta E = 0$		
II	Gly	−76	59	172	Gly	−76	68	180
	Ala	−66	155	175	Ala	−36	99	187
	Pro	−46(−30)	116	184	Hyp	−60(−26)	123	186
	Gly	109	−40	181	Gly	113	−27	180
	Ala	−78	64	180	Ala	−75	67	179
		$E_{tot} = -17.8$				$E_{tot} = -25.0$		
		$\Delta E = 1.5$				$\Delta E = 1.2$		
III	Gly	−75	61	174	Gly	−75	61	174
	Ala	−64	149	178	Ala	−60	149	177
	Pro	−68(29)	80	190	Hyp	−67(29)	81	189
	Gly	156	−34	179	Gly	155	−36	179
	Ala	−76	64	180	Ala	−76	64	180
		$E_{tot} = -17.1$				$E_{tot} = -22.8$		
		$\Delta E = 2.2$				$\Delta E = 3.4$		
IV	Gly	−74	67	178	Gly	−75	65	177
	Ala	−53	117	177	Ala	−45	105	187
	Pro	−55(−27)	137	178	Hyp	−65(−25)	149	175
	Gly	−75	64	184	Gly	−73	69	183
	Ala	−76	63	179	Ala	−73	67	178
		$E_{tot} = -15.4$				$E_{tot} = -24.2$		
		$\Delta E = 3.9$				$\Delta E = 2.0$		
V	Gly	−75	65	179	Gly	−75	66	179
	Ala	−57	119	173	Ala	−42	104	184
	Pro	−58(−24)	156	170	Hyp	−67(−22)	165	173
	Gly	−63	150	183	Gly	−60	142	183
	Ala	−75	67	180	Ala	−75	66	180
		$E_{tot} = -12.1$				$E_{tot} = -21.2$		
		$\Delta E = 7.2$				$\Delta E = 5.0$		
VI	Gy	−76	60	181	Gly	−152	65	179
	Ala	−117	89	180	Ala	−113	85	184
	Pro	−75(31)	68	186	Pro	−77(33)	67	181
	Gly	158	63	185	Gly	160	56	181
	Ala	−75	67	179	Ala	−76	65	180
		$E_{tot} = -12.2$				$E_{tot} = -17.4$		
		$\Delta E = 7.1$				$\Delta E = 8.8$		
VII	Gly	−75	64	177	Gly	−75	63	176
	Ala	−62	128	180	Ala	−62	124	180
	Pro	−72(30)	151	173	Pro	−71(30)	155	174
	Gly	−66	136	181	Gly	−65	141	182
	Ala	−74	67	180	Ala	−75	66	180
		$E_{tot} = -11.5$				$E_{tot} = -18.8$		
		$\Delta E = 7.8$				$\Delta E = 7.4$		

TABLE I   *Continued*

Conformation	Residue[b]	Pro peptide			Residue	Hyp peptide		
		ϕ[c]	ψ	ω		ϕ[c]	ψ	ω
VIII	Gly	−65	142	176	Gly	−63	134	176
	Ala	−61	143	178	Ala	−60	142	179
	Pro	−68(28)	156	177	Pro	−68(29)	154	176
	Gly	−66	149	175	Gly	−64	143	176
	Ala	−64	144	180	Ala	−62	136	181
		$E_{tot} = -9.8$				$E_{tot} = -16.6$		
		$\Delta E = 9.5$				$\Delta E = 9.4$		

[a] The dihedral angles are in degrees and the energies in kcal per mole of the given peptide.

[b] The residues are listed in the order of their positions in the peptide sequence.

[c] The values in parentheses are for the torsion angle χ (N—C^α—C^β—C^γ) for the Pro or Hyp residues. Positive and negative values correspond, respectively, to the down and up conformations of the pyrrolidine ring.[10]

tion with $\phi_{Pro} = -65°$ or when ϕ_{Pro} was selected as $-57°$. The resulting eight different conformations obtained are shown in Table I. (The energy is expressed throughout in kcal/mol of the peptide sequence under consideration.) The energy contributions due to the terminal acetyl and N-methyl groups have been subtracted from the computed total energy (E_{tot}) of each conformation listed in the table. The energy difference (ΔE) between a given conformation and the lowest energy conformation (I) is also shown in the table. The following comments may be made about the data shown in Table I:

(a) In conformations I, II, IV, V, the pyrrolidine ring in Pro and Hyp is found in the up conformation [as indicated by negative values for the ring torsional angle χ'(N—C^α—C^β—C^γ)], while in III, VI, VII, and VIII it is in the down conformation (χ' positive).

(b) Conformations I, II and III contain the β-bend in the Pro-Gly or Hyp-Gly region, while IV–VIII are extended structures and do not contain the bend. In the former category, I and II with the pyrrolidine ring in the up position were found to form the "ideal" β-bend having the $(i + 3) \rightarrow i$ NH...OC hydrogen bond as defined recently by Paul and Ramakrishnan.[24] If the ring conformation is down, as in III, then the hydrogen-bonding interaction was found to be relatively weak, but the pentapeptides, nevertheless, were found to prefer the folded conformation. This would seem to indicate that the presence of the $(i + 3) \rightarrow i$ type of NH...OC hydrogen bond is not very essential for the stability of the β-bend. In fact, conformation III was found to be stabilized mainly by a relatively weaker $(i + 2) \rightarrow i$ type of NH...OC hydrogen bond.[24] This is also found to be true for the β-bend conformation observed by Lee et al.[16] in Pro-Pro-Gly-Pro sequences [where an $i + 3 \rightarrow i$ bond is not possible with Pro as the $(i + 3)$ residue], although these authors have not noted this point explicitly.

(c) The conformation for the tripeptide fragment Gly-Ala-Pro preceding the β-bend region does not appear rigidly fixed. However, when Hyp occurs in the place of Pro, then a favorable OH...OC interaction between the hydroxyl group of Hyp and the carbonyl oxygen of Gly is found in the Gly-Ala-Hyp

fragment when the Hyp ring is in the up conformation. This intramolecular hydrogen-bonding arrangement is similar to that proposed by us earlier[19] for the B-form of poly(Hyp). This hydrogen bond is found not only when the pentapeptide contains a β-bend (at the Hyp-Gly region), as in conformations I and II, but also when the chain is extended, as in IV and V. The latter two conformations may be called "partially extended," in contrast to the fully extended structures VI–VIII. (The difference between IV and V lies in the "curling" of the C-terminal end in conformation IV while this is avoided in V).

(d) It is found from conformation I that two successive β-bends can occur if the sequence were extended further, with ϕ_{i+3} and ψ_{i+3} equal to ϕ_i and ψ_i, respectively. However, model-building studies indicated that a third successive β-bend will cause significant stereochemical problems, particularly when Ala is replaced by a bulkier residue (see below).

(e) The fully extended conformations VI–VIII for Gly-Ala-Pro-(Hyp)Gly-Ala do not contain any intramolecular hydrogen bond. VI is found be close to the conformation obtained by Nemethy et al.[15] as the most preferred for a single chain of poly(Gly-Ala-Pro). VII is similar to the conformation observed in poly(Hyp) in the A-form[19] and in polyproline-II.[2] VIII is identical to the best conformation for parallel chains in poly(Gly-Ala-Pro).[15] It is interesting to note that these extended conformations are higher in energy by as much as about 7–9 kcal/mol than the minimum energy conformation (I) for the pentapeptides studied here (Table I). This is due to the latter containing the β-bend (stabilized by $4 \rightarrow 1$ hydrogen bond) and, in the case of the Hyp peptide, the $(Hyp_3)OH \ldots OC(Gly_1)$ hydrogen bond.

(f) It may be pointed out that, while the extended conformation VI is quite favorable for the proline peptide (Table I and Ref. 15) and is energetically close to V and VII ($E_{tot} = -12$ kcal/mol), it is found to be less stable ($E_{tot} = -17.4$ kcal/mol) for the Hyp-containing peptide when compared to conformation V ($E_{tot} = -21.2$ kcal/mol) or VII ($E_{tot} = -18.8$ kcal/mol). Thus, the extended conformation most favored for the proline peptide is seen to be destabilized in the Hyp peptide due to unfavorable electrostatic interactions, particularly with the succeeding Gly residue.

(g) It is worth noting that the fully extended conformations are about 5 kcal/mol higher in energy than the partially extended ones in the case of both the Pro- and Hyp-containing peptides.

From the above observations, it follows that the formation of the triple helix starting from the minimum energy conformations of the pentapeptide sequences Gly-Ala-Pro-Gly-Ala and Gly-Ala-Hyp-Gly-Ala should involve the straightening of the initial folded conformations into that corresponding to that of the individual strands of the triple helix, which would then be followed by association of the three strands. While the first step, viz., the straightening process, is seen to involve an increase in E_{tot} (Table I), the subsequent chain association step will lead to a substantial decrease in E_{tot} due to *inter*chain interactions.[17] The partially extended conformations IV and V can serve as an intermediate in the straightening process. This is elaborated further below.

Gly-Pro-Pro-Gly-Ala and Gly-Pro-Hyp-Gly-Ala Peptides

The computed conformational energy values for these pentapeptides are given, in ascending order, in Table II for the 8 conformations from I to VIII.

TABLE II
Computed Conformations of Gly-Pro-Pro-Gly-Ala and
Gly-Pro-Hyp-Gly-Ala[a]

Conformation	Residue[b]	Pro-peptide			Residue	Hyp-peptide		
		ϕ^c	ψ	ω		ϕ^c	ψ	ω
I	Gly	169	−88	184	Gly	164	−77	180
	Pro	−62(23)	154	187	Pro	−66(25)	162	179
	Pro	−65(26)	84	185	Hyp	−56(24)	106	190
	Gly	158	−47	171	Gly	177	−61	177
	Ala	−77	70	181	Ala	−75	63	180
		$E_{tot} = -6.6$				$E_{tot} = -13.7$		
		$\Delta E = 0.0$				$\Delta E = 0.0$		
II	Gly	−58	119	178	Gly	−56	119	181
	Pro	−52(−29)	122	176	Pro	−42(−34)	102	188
	Pro	−42(−30)	113	185	Hyp	−60(−27)	114	189
	Gly	119	−36	182	Gly	129	−38	181
	Ala	−76	66	179	Ala	−76	65	179
		$E_{tot} = -2.6$				$E_{tot} = -11.0$		
		$\Delta E = 4.0$				$\Delta E = 2.7$		
III	Gly	−62	128	181	Gly	−57	120	176
	Pro	−67(27)	149	183	Pro	−52(−28)	135	184
	Pro	−68(28)	81	189	Hyp	−68(30)	84	188
	Gly	164	−38	171	Gly	155	−35	174
	Ala	−67	137	180	Ala	−67	139	180
		$E_{tot} = -3.3$				$E_{tot} = -9.9$		
		$\Delta E = 3.3$				$\Delta E = 3.8$		
IV	Gly	−58	124	177	Gly	−60	123	175
	Pro	−50(−30)	122	175	Pro	−42(−33)	106	188
	Pro	−54(−28)	132	180	Hyp	−64(−25)	141	180
	Gly	−74	65	181	Gly	−75	64	181
	Ala	−74	65	180	Ala	−75	66	180
		$E_{tot} = -0.9$				$E_{tot} = -8.9$		
		$\Delta E = 5.7$				$\Delta E = 4.8$		
V	Gly	130	−87	170	Gly	128	−89	172
	Pro	−66(30)	161	181	Pro	−67(30)	160	181
	Pro	−75(31)	68	174	Hyp	−74(32)	70	175
	Gly	161	−70	184	Gly	159	−72	183
	Ala	−71	154	180	Ala	−70	156	180
		$E_{tot} = 1.0$				$E_{tot} = -7.3$		
		$\Delta E = 5.6$				$\Delta E = 6.4$		
VI	Gly	−58	120	177	Gly	−60	139	176
	Pro	−50(−30)	121	174	Pro	−40(−32)	104	184
	Pro	−56(−26)	154	174	Hyp	−67(−24)	163	177
	Gly	−59	142	183	Gly	−58	119	183
	Ala	−74	68	179	Ala	−75	65	179
		$E_{tot} = 3.8$				$E_{tot} = -6.5$		
		$\Delta E = 10.4$				$\Delta E = 7.2$		
VII	Gly	−58	119	176	Gly	−58	119	176
	Pro	−56(−26)	145	176	Pro	−50(−29)	122	176
	Pro	−69(30)	154	174	Hyp	−54(−28)	143	178
	Gly	−63	139	181	Gly	−60	140	182
	Ala	−74	68	180	Ala	−75	69	180
		$E_{tot} = 3.1$				$E_{tot} = -5.8$		
		$\Delta E = 9.7$				$\Delta E = 7.9$		

TABLE II   *Continued*

Conformation	Residue[b]	Pro-peptide			Residue	Hyp-peptide		
		ϕ^c	ψ	ω		ϕ^c	ψ	ω
VIII	Gly	−64	140	179	Gly	−62	136	181
	Pro	−68(29)	155	176	Pro	−68(28)	150	177
	Pro	−69(30)	156	172	Hyp	−69(31)	155	173
	Gly	−63	142	177	Gly	−64	134	176
	Ala	−62	134	180	Ala	−62	138	180
		$E_{tot} = 3.5$				$E_{tot} = -5.4$		
		$\Delta E = 10.1$				$\Delta E = 8.3$		

[a] The dihedral angles are in degrees and the energies in kcal per mole of the given peptide.
[b] The residues are listed in the order of their positions in the peptide sequence.
[c] The values in parentheses are for the torsion angle χ' ($N-C^\alpha-C^\beta-C^\gamma$) for the Pro or Hyp residues. Positive and negative values correspond, respectively, to the down and up conformations of the pyrrolidine ring.[10]

These results are very similar to those for the earlier set of pentapeptides described above. Thus, conformations containing the β-bend (I–III) are more favorable than the extended conformations (V–VIII). The E_{tot} values for all the conformations in Table II are seen to be higher than their counterparts in Table I, due to the additional rigidity imposed by the imino acid in the second position in the former. In the Hyp-containing peptide, conformations II and III as well as the partially extended conformations IV and VI contain the $(Hyp_3)OH \ldots OC(Gly_1)$ hydrogen bond. Although the minimum energy conformation I for either the Pro- or Hyp- pentapeptide contains the β-bend (Table II), the bend cannot continue along the chain (i.e., $\phi_1, \psi_1 \neq \phi_4, \psi_4$) due to more severe stereochemical constraints in these peptides than in the earlier class [viz. Gly-Ala-Pro(Hyp)-Gly-Ala]. We also note that the energy differences between the partially extended conformation VI and the fully extended conformation VIII are smaller in these peptides than in Gly-Ala-Pro(Hyp)-Gly-Ala peptides (Tables I and II), indicating that the Gly-Pro-Pro(Hyp) sequences prefer the extended structures more readily. The Hyp peptides have a lower overall energy difference between the bent and extended conformations when compared to the Pro peptide analogs (Table II), showing that the Hyp residue more readily prefers the conformations required for extended structure formation.

Gly-Ala-Pro-Gly-Ala-Pro and Gly-Ala-Hyp-Gly-Ala-Hyp

In order to see whether the presence of a second imino acid residue at the C-terminal end of the peptide sequence would affect the observations made on the results of the pentapeptides, the energy calculations were extended to the hexapeptide sequences shown. For this purpose, 6 starting conformations were used (only 1 of which contained the β-bend) and their conformational energies arrived at by minimization. As before, the β-bend-containing conformation (I) is found to have the lowest energy for both the Pro- and Hyp peptides (Table III). This conformation is shown in Fig. 1 for the Hyp peptide indicating

TABLE III
Computed Conformations of Gly-Ala-Pro-Gly-Ala-Pro and Gly-Ala-Hyp-Gly-Ala-Hyp[a]

Conformation	Residue[b]	Pro-peptide			Residue	Hyp-peptide		
		ϕ^c	ψ	ω		ϕ^c	ψ	ω
I	Gly	−75	64	176	Gly	−74	69	180
	Ala	−60	128	176	Ala	−34	95	191
	Pro	−47(−30)	114	186	Hyp	−58(−28)	120	186
	Gly	123	−35	179	Gly	118	−39	181
	Ala	−55	120	175	Ala	−55	109	185
	Pro	−56(−26)	106	181	Hyp	−65(−23)	144	180
		$E_{tot} = -12.4$				$E_{tot} = -28.5$		
		$\Delta E = 0.0$				$\Delta E = 0.0$		
II	Gly	−75	63	178	Gly	−75	68	178
	Ala	−62	122	173	Ala	−41	102	186
	Pro	−56(−26)	151	172	Hyp	−65(−23)	154	178
	Gly	−64	149	178	Gly	−59	136	179
	Ala	−54	121	176	Ala	−40	102	186
	Pro	−51(−29)	121	181	Hyp	−64(−24)	141	180
		$E_{tot} = -7.2$				$E_{tot} = -25.9$		
		$\Delta E = 5.2$				$\Delta E = 2.6$		
III	Gly	−62	141	178	Gly	−59	130	178
	Ala	−54	120	175	Ala	−44	106	187
	Pro	−56(−26)	152	169	Hyp	−63(−27)	154	179
	Gly	−61	144	183	Gly	−56	133	179
	Ala	−51	116	175	Ala	−42	104	187
	Pro	−55(−26)	133	180	Hyp	−64(−25)	143	180
		$E_{tot} = -4.6$				$E_{tot} = -23.6$		
		$\Delta E = 7.8$				$\Delta E = 4.9$		
IV	Gly	−74	65	176	Gly	−74	68	178
	Ala	−61	131	178	Ala	−56	132	182
	Pro	−72(31)	157	171	Hyp*	−70(30)	154	171
	Gly	−65	137	177	Gly	−64	136	178
	Ala	−60	138	182	Ala	−59	138	179
	Pro	−76(32)	137	181	Hyp*	−66(28)	138	180
		$E_{tot} = -6.2$				$E_{tot} = -20.6$		
		$\Delta E = 6.2$				$\Delta E = 7.9$		
V	Gly	−65	142	176	Gly	−63	132	177
	Ala	−61	142	178	Ala	−58	139	179
	Pro	−68(28)	156	178	Hyp	−66(29)	153	176
	Gly	−67	151	175	Gly	−63	145	176
	Ala	−62	142	177	Ala	−60	140	178
	Pro	−69(29)	153	180	Hyp	−75(32)	71	180
		$E_{tot} = -5.0$				$E_{tot} = -19.3$		
		$\Delta E = 7.4$				$\Delta E = 9.2$		
VI	Gly	−168	68	181	Gly	−167	71	180
	Ala	−117	76	181	Ala	−116	76	181
	Pro	−76(32)	65	183	Hyp	−75(32)	66	184
	Gly	164	62	181	Gly	161	78	181
	Ala	−115	81	181	Ala	−116	62	181
	Pro	−75(32)	69	180	Hyp	−75(32)	71	180
		$E_{tot} = -6.0$				$E_{tot} = -15.5$		
		$\Delta E = 6.4$				$\Delta E = 13.0$		

[a] The dihedral angles are in degrees and the energies in kcal per mole of the given peptide.
[b] The residues are listed in the order of their positions in the peptide sequence.
[c] The values in parentheses are for the torsion angle $\chi'(N—C^\alpha—C^\beta—C^\gamma)$ for the Pro or Hyp residues. Positive and negative values correspond, respectively, to the down and up conformations of the pyrrolidine ring.[10]
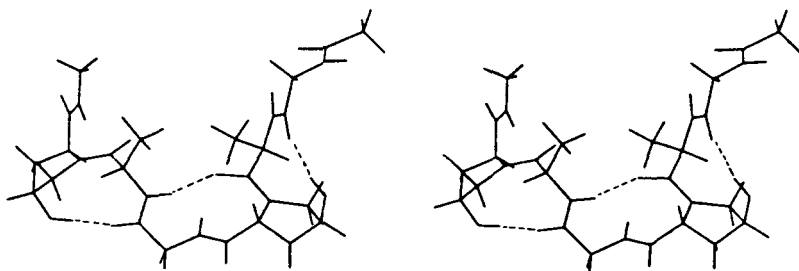
Fig. 1. Stereodrawing of the hexapeptide fragment Gly-Ala-Hyp-Gly-Ala-Hyp corresponding to conformation I in Table III. Hydrogen bonds, $N_5H_5 \ldots O_2$, $(Hyp_3)OH \ldots OC(Gly_1)$, and $(Hyp_6)OH \ldots OC(Gly_4)$ are present.
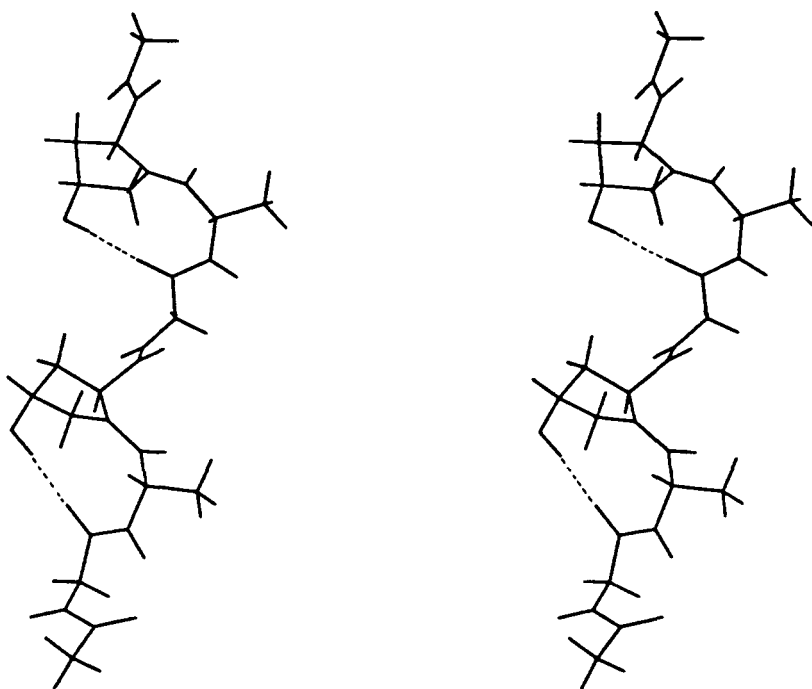

Fig. 2. Stereodrawing of conformation III in Table III. Only the $(Hyp)OH \ldots OC(Gly)$ hydrogen bonds are possible in this case.

$(Ala_2)CO \ldots NH(Ala_5)$ and the $(Hyp_3)OH \ldots OC(Gly_1)$ and $(Hyp_6)OH \ldots OC(Gly_4)$ hydrogen bonds. The partially extended structure containing the $(Hyp_3)OH \ldots OC(Gly_1)$ and $(Hyp_6)OH \ldots OC(Gly_4)$ hydrogen bonds is shown in Fig. 2. Conversion of this structure to the fully extended one (conformation V or VI in Table III) requires breaking of these two hydrogen bonds and is therefore separated by a larger energy difference than in the other two cases of peptides (Tables I and II). The fully extended structure of
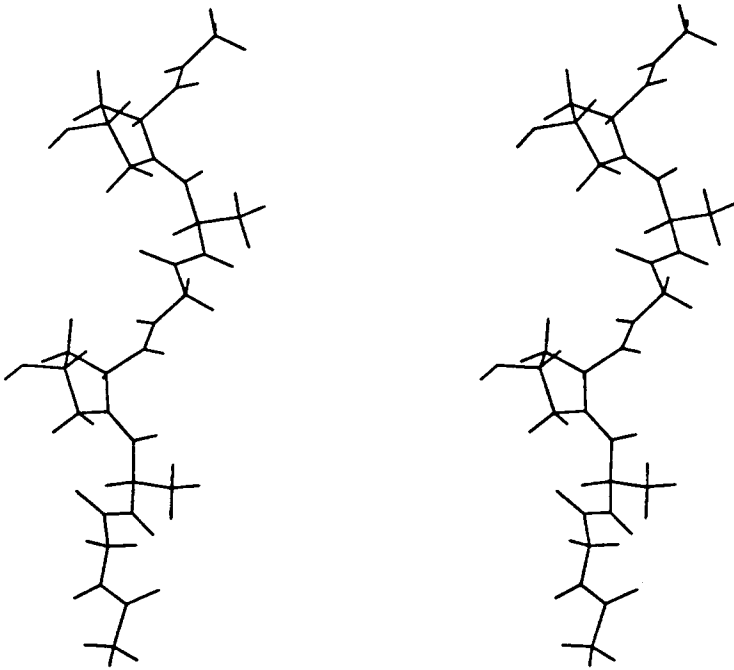
Fig. 3.   Stereodrawing of conformation VI of Table III. This is the optimum conformation for triple-helix formation for the sequence poly(Gly-Pro-Hyp).[18]

the Hyp peptides corresponding to conformation VI in Table III is depicted in Fig. 3. It may be noted that, for the Hyp peptide, the conformation containing the OH...OC hydrogen bonds (Fig. 2) and the fully extended collagenlike structure (Fig. 3) are closely similar. The distance between every C_i^α and C_{i+3}^α atom is found to be about 9.0 ± 0.1 Å in both cases. The orientations of the amino and carboxyl groups are also similar, and are favorable for the formation of interchain hydrogen bonds. Thus, the intermediate conformation in the Hyp peptide is almost as conducive to the formation of the interchain hydrogen bonds in the triple helix as the fully extended conformations.
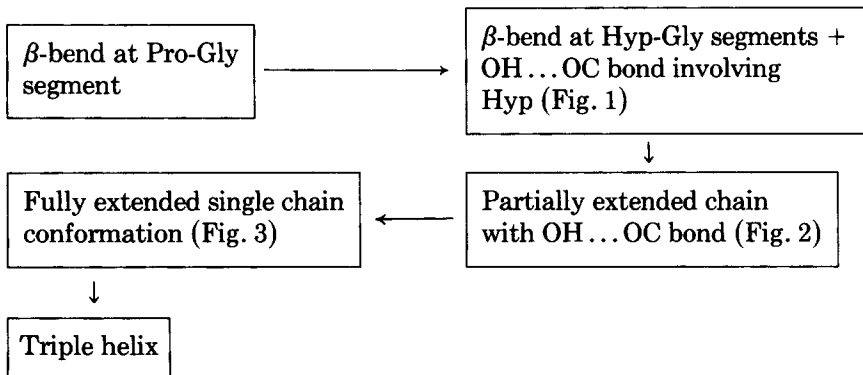
Kinetic Pathway for Triple-Helix Formation

It is known that the triple helix is the most stable conformation for polypeptides containing the collagenlike Gly-X-Y repeat sequences.[2] The major source of stability of the triple helix is the interchain noncovalent interactions including the hydrogen bond. An interesting result of the present study is that the fully extended single-chain conformation, which is a prerequisite for triple-helix formation, is *not* the minimum energy conformation for the pentapeptide models of collagen; a conformation containing a β-bend at the Pro-Gly or Hyp-Gly region is found to be substantially more stable than the extended one (Tables I–III). Such information is not available from the earlier theoretical studies by others on polytripeptide models of

collagen,[15, 17, 18, 24] which dealt with regularly repeating structures for the tripeptide units. It is pertinent to recall here our observation that the β-bend cannot continue along the polypeptide chain because of steric constraints, which are found to be particularly severe in peptides containing the Gly-Pro-Pro and Gly-Pro-Hyp sequences. These constraints would therefore be expected to be responsible for straightening the initially folded conformation into the fully extended one, similar to those found in the polytripeptides[15, 17, 25] as well as in polyproline-II[2] and poly(Hyp).[19]

There is, however, an intermediate conformation seen only in the Hyp-containing peptides, which is somewhat less extended than the final (single-chain) conformation and retains the OH...OC hydrogen bond found in the starting conformation. The presence of this hydrogen bond would tend to "lock" the *trans* conformation around the Gly-X and X-Hyp peptide bonds (X=Ala or Pro in the peptides selected), thereby minimizing the probability of their isomerization into the *cis* form. This will greatly facilitate triple-helix formation, the rate of which (in collagen) is known to be retarded significantly by the presence of the *cis* isomer.[26] In contrast, the *trans* peptide bonds in the intermediate conformations of the proline peptide counterparts would exist in equilibrium with those containing the *cis* isomer, one of which has actually been shown to be the minimum energy conformation for the single chain of poly(Gly-Pro-Pro) by Miller and Scheraga.[17]

Based on the results of our energy calculations, we postulate the following kinetic pathway for triple-helix formation in collagen, which brings out the important role of Hyp in this process: The nascent procollagen molecule would take up a folded conformation containing the β-bend at the Pro-Gly regions (and the PP-II conformation in the X-Pro regions[8, 9]). This would act as the conformation necessary for the hydroxylation of the Pro residue by prolylhydroxylase as proposed[23] and verified[5, 8, 9] in our earlier studies. The hydroxyl group of the resulting Hyp residue would form a hydrogen bond with the carbonyl oxygen of the Gly residue (in the sequence Gly-X-Hyp or Gly-Pro-Hyp), and this favorable stabilizing interaction is preserved in the subsequent step in the formation of the triple helix, namely, the straightening (or "unbending") of the folded β-bend conformation. The resulting partially extended conformation (Fig. 2) containing the OH...OC hydrogen bond is found to be about 3–5 kcal/mol lower in energy as compared to the fully extended polyproline-II type conformation, which is about 8–12 kcal/mol higher in energy with respect to the initial folded conformation (Tables I–III). Thus, the partially extended conformation containing the OH...OC hydrogen bond could act as a kinetic intermediate for the formation of the fully extended single-chain conformation (such as the last 2 conformations in Tables I–III and represented in Fig. 3), which would subsequently associate intermolecularly to form the stable triple helix. Such an intermediate conformation will not be available for the unhydroxylated peptide chain. This, in conjunction with the additional important consequence of the OH...OC bond formation, namely, the locking of the *trans* isomer around the Pro- and Hyp-involving peptide bonds, would account for the observed enhancement of the folding rate in the Hyp-containing peptides in comparison with the proline counterparts. The proposed kinetic pathway for collagen folding is

shown in following scheme:

```
┌─────────────────────┐              ┌──────────────────────────────┐
│ β-bend at Pro-Gly   │ ───────────> │ β-bend at Hyp-Gly segments + │
│ segment             │              │ OH ... OC bond involving     │
│                     │              │ Hyp (Fig. 1)                 │
└─────────────────────┘              └──────────────────────────────┘
                                                   │
                                                   ↓
┌─────────────────────────┐          ┌──────────────────────────────┐
│ Fully extended single chain│ <───── │ Partially extended chain     │
│ conformation (Fig. 3)   │          │ with OH ... OC bond (Fig. 2)  │
└─────────────────────────┘          └──────────────────────────────┘
         │
         ↓
┌─────────────────┐
│ Triple helix    │
└─────────────────┘
```

In conclusion, we have attempted to show that, based on energetic consider-ations, peptides and polypeptides containing the Hyp residue could fold into the triple-helical conformation faster than their proline analogs, thanks to the conformations that are uniquely available to the former in the course of the folding process.

References

1. Ramachandran, G. N. Bansal, M. & Bhatnagar, R. S. (1973) *Biochem. Biophys. Acta* **322**, 166–169.

2. Ramachandran, G. N. & Ramakrishnan, C. (1976) in *Biochemistry of Collagen*, Ramachandran, G. N. & Reddi, A. H., Eds., Plenum, New York, pp. 45–85.

3. Berg, R. & Prokop, D. J. (1973) *Biochem. Biophys. Res. Commun.* **52**, 115–120.

4. Sakakibara, S., Inoyue, K., Shudo, K., Kishida, Y., Kobayashi, Y. & Prokop, D. J. (1973) *Biochem. Biophys. Acta* **303**, 198–202.

5. Chopra, R. K. & Ananthanarayanan, V. S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7180–7184.

6. Ananthanarayanan, V. S. (1983) in *Conformation in Biology*, Srinivasan, R. & Sarma, R. H., Eds., Adenine, New York, pp. 99–111.

7. Rao, N. V. & Adams, E. (1979) *Biochem. Biophys. Res. Commun.* **86**, 654–660.

8. Ananthanarayanan, V. S. (1985) in *Proceedings of the 16th FEBS Congress*, Ovchinnikov, Yu. A., Ed., VNU Science, The Netherlands, Part A, PP. 23–30.

9. Ananthanarayanan, V. S., Attah-Poku, S. K., Mukkamala, P. L. & Rehse, P. H. (1985) *Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci.* **8**, 209–221.

10. Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975) *J. Phys. Chem.* **79**, 2361–2381.

11. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–438.

12. Poland, D. & Scheraga, H. A. (1967) *Biochemistry* **6**, 3791–3800.

13. Koetzel, T. F., Lehmann, M. S. & Hamilton, W. C. (1973) *Acta Cryst.* **829**, 231–236.

14. Nemethy, G. Miller, M. & Scheraga, H. A. (1980) *Macromolecules* **13**, 910–913.

15. Nemethy, G., Miller, M. & Scheraga, H. A. (1980) *Macromolecules* **13**, 914–919.

16. Lee, E., Nemethy, G., Scheraga, H. A. & Ananthanarayanan, V. S. (1984) *Biopolymers* **23**, 1193–1206.

17. Miller, M. & Scheraga, H. A. (1976) *J. Polymer Sci.* **54**, 171–200.

18. Miller, M., Nemethy, G. & Scheraga, H. A. (1980) *Macromolecules* **13**, 470–478.

19. Bansal, M., Brahmachari, S. K. & Sasisekharan, V. (1979) *Macromolecules* **12**, 15–22.

20. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.

21. Singh, U. C. & Kollman, P. A. (1984) *J. Comp. Chem.* **5**, 129–143.

22. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7**, 230–252.

23. Brahmachari, S. K. & Ananthanarayanan, V. S. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5119–5124.

24. Paul, P. K. C. & Ramakrishnan, C. R. (1985) *J. Biomol. Struct. Dynam.* **2**, 879–898.

25. Tumanayan, V. G., Abagyan, R. A. & Episova, N. G. (1984) *Biopolymers* **23**, 1499–1512.

26. Bachinger, H. P., Bruckner, P., Timpl, R., Prokop, D. J. & Engel, J. (1980) *Eur. J. Biochem.* **106**, 619–632.