Calculation of the Characteristic Ratio of Randomly Coiled Poly(L-proline)[1]

Seiji Tanaka[2a] and Harold A. Scheraga*[2b]

Department of Chemistry, Cornell University, Ithaca, New York 14853.
Received March 24, 1975

ABSTRACT: Allowing for rotation about the C^{α}–C′ bond (i.e., variation of ψ) and for some degree of freedom about the peptide bond (i.e., small variations of ω), the characteristic ratios, $\langle R^2\rangle_0/n\bar{l}^2$, of the form I (cis) and form II (trans) poly(L-proline) chain have been calculated by a Monte Carlo method in which the conformational energies were used as weighting factors. The Monte Carlo method enabled short-range interactions (beyond those involved in a single residue) to be taken into account. The effect of the presence of a small amount of one form (say cis in a trans-rich chain) on $\langle R^2\rangle_0/n\bar{l}^2$ was also investigated. The results for the trans-rich form are in good agreement with values observed experimentally in solvents in which the poly(L-proline) chain is predominantly in form II; the presence of a small amount of cis residues reduces the characteristic ratio of the trans-rich form of poly(L-proline) significantly.

Computations of stable conformations of macromolecules are being carried out to obtain an understanding of the nature and magnitudes of the interactions which determine the macromolecular conformation.[3] To assess the reliability of our procedures and energy parameters for treating the proline residue, we have calculated the fraction of conformers of oligomers of L-proline[4] and the thermodynamic parameters for the conformational transition between forms I and II of poly(L-proline)[5] and have compared these quantities with experiment. For this same purpose, we compare here some calculated and experimental conformational properties of randomly coiled[6] (or statistically averaged) poly(L-proline) under ideal conditions, i.e., at the θ point[7] where the conformational properties depend only on short-range interactions.[8–11]

Of the various conformational properties of a chain molecule in solution[12] that can serve our purpose, the unperturbed mean-square end-to-end distance, $\langle R^2\rangle_0$, is a convenient one, and reflects the properties of the whole conformational energy surface; the zero subscript implies that this quantity pertains to the θ point. It is customary to express the unperturbed chain dimension as a characteristic ratio, C, where

$$C = \langle R^2\rangle_0/n\bar{l}^2 \tag{1}$$

with n being the number of *identical* repeating units in the polymer chain, and $\bar{l}^2$ being the mean of the squared length of the bonds comprising the repeating unit; the quantity $n\bar{l}^2$ is the mean-square end-to-end distance of a freely jointed chain.[7] The definition of a repeating unit and of $\bar{l}^2$ adopted here will be given later (see eq 4). Several calculations of C for polypeptide chains have been reported,[13–21] and in general the calculated results have been in good agreement with experimental ones (see ref 17b for a comparison of calculated and experimental values of C). While calculations[20] and experimental results[22] have been reported for the characteristic ratio of poly(L-proline), we will offer an alternative treatment for the characteristic ratio of this homopolymer. In this paper, we present the calculation of C for poly(L-proline) by taking into account rotation about the C^{α}–C′ bond (variation of the dihedral angle[23] ψ), some freedom of rotation about the peptide bond (variation of the dihedral angle[23] ω), puckering of the pyrrolidine ring, and the presence of both cis and trans conformations of the peptide bond.

I. Calculation Procedure

In order to avoid the inclusion of conformations having high-energy short-range contacts in the Monte Carlo procedure used here, we eliminate such high-energy conformations from consideration at the outset by consideration of conformational energy maps of the chain structure of Figure 1A, which will be designated here as an L-Pro-L-Pro dipeptide. These maps will be used only for the random selection of L-Pro-L-Pro conformations, with equal weight being given (in the initial selection process) to all low-energy regions; the conformational energy is calculated (as a weighting factor for the whole chain) only after a whole chain has been generated.

A. Low-Energy Conformations of L-Pro-L-Pro Dipeptide. A description of the structural parameters and the conformational energy functions[24] (nonbonded, electrostatic, and torsional about the peptide bond) of the proline residue was given in our previous papers.[4,5,25] The IUPAC-IUB nomenclature and conventions[23] are used in this paper.

In order to estimate the energetically accessible range of the rotational states (i.e., of ψ_i and ω_i) of the L-proline dipeptide (to make the Monte Carlo calculation practical and effective), the conformational energy of the structure depicted in Figure 1A was computed. In contrast to our earlier computations,[4] where ω was held fixed, ω is allowed to vary here. Previously,[4,5] we allowed for two states of puckering at the C^{β} and C^{γ} atoms of the pyrrolidine ring,[24] one designated "down, or D", with the C^{γ} atom at $\phi = -75.0°$, $\chi^1 = +18.67°$, and the other designated "up, or U", with the C^{γ} atom at $\phi = -67.6°$, $\chi^1 = -6.11°$. However, in this paper, we use only D puckering because a chain consisting of 30 residues with U puckering, and chains of the same length with random D and U puckering, were found to be energetically less favorable than one with all-D puckering.[5] However, it should be noted that the existence of U puckering cannot be neglected for relatively short oligomers of L-proline or for the isolated L-proline residue.[4] The dihedral angle ψ_i is allowed to vary over the whole range, and the dihedral angle ω_i is allowed to vary to a small extent from its values for the planar trans (180°) and cis (0°) conformations.

In the dipeptide depicted in Figure 1A, there is another rotational degree of freedom, viz., ω_{i-1}, which can influence the conformation of this structure. Initially, the most probable values of ω_{i-1} were determined by minimizing the conformational energy of the structure of Figure 1A, allowing ω_{i-1}, ψ_i and ω_i to vary independently. The results of this energy minimization are summarized in Table I for the trans–trans, trans–cis, cis–trans, and cis–cis values of ω_{i-1} and ω_i, respectively. On the basis of these results for the L-proline dipeptide, the value of ω_{i-1} was fixed at that of ω_i

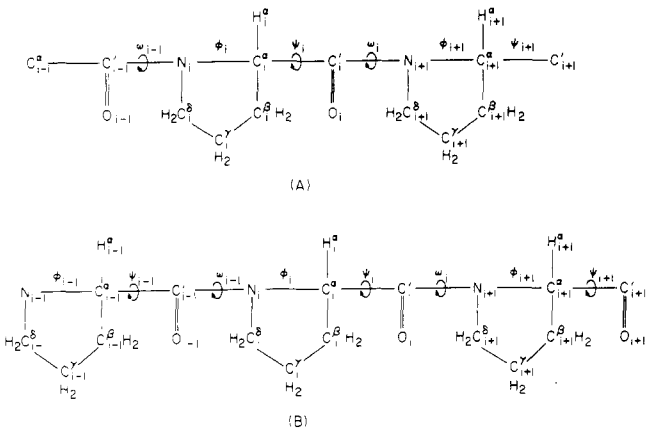

Figure 1. (A) Dipeptide structure of L-proline used in the computation of the conformational energy (shown in Figures 2A–2D) in the present paper. The IUPAC nomenclature and conventions[23] are employed. (B) Three residues in the interior of the poly(L-proline) chain. This structure is used to obtain the diagram of Table III.

obtained by the energy minimization, when computing the conformational energy of the ith residue in the *interior* of a poly(L-proline) chain as a function of ψ_i and ω_i. However, as can be seen in Table I, the value of ω_{i-1} had little effect on that of ω_i. For example, $\omega_i = 175.6°$ for ω_{i-1} in the trans state, and is 174.6° for ω_{i-1} in the cis state. Therefore, in order to save computer time, we fix the trans value of ω_{i-1} at 175.0° (i.e., as the average of 175.6 and 174.6) and the cis value of ω_{i-1} at −8.7° (i.e., as the average of −8.2 and −9.1°). Then, fixing ω_{i-1} at these values, we computed the conformational energy of the structure of Figure 1A as a function of ψ_i and ω_i for the trans–trans, trans–cis, cis–trans, and cis–cis conformations with respect to ω_{i-1} and ω_i, respectively. These results are shown in Figure 2A–D. The approximation, in which the average values of 175.0 and −8.7° for ω_i are assigned to ω_{i-1} to simulate the behavior of an *interior* residue, does not affect the calculations of the characteristic ratio because the energy maps of Figures 2A–D are used only to select single-residue conformations for the Monte Carlo computation, and *not* to compute the conformational energy. The latter is computed after each chain is generated by the Monte Carlo procedure

Table I
Conformations of Minimum Energy in the L-Proline Dipeptide[a]

Conformations of ω_{i-1} and ω_i	Starting[b] conformation, deg			Conformation of minimum energy, deg		
	ω_{i-1}	ψ_i	ω_i	ω_{i-1}	ψ_i	ω_i
Trans–trans	180	164.5	180	179.6	165.5	175.6
Trans–cis	180	160.5	0	178.6	163.5	−8.2
Cis–trans	0	164.2	180	0.3	166.3	174.6
Cis–cis	0	160.7	0	0.9	166.2	−9.1

[a] The minimization was carried out for the structure depicted in Figure 1A. [b] These values were chosen from the conformations of approximate minimum energy in the L-proline dipeptide, but having an acetyl end group as depicted in Figure 2A of ref 4. See the discussion in section IB for the conformation of the acetyl end group.

by inclusion of both short- and somewhat longer-range interactions.

B. Calculation of Characteristic Ratio. In the past, the peptide group was taken in the planar trans conformation, and the interactions between neighboring residues were neglected[12–18,20,21] when computing $\langle R^2 \rangle_0$; the computation was carried out by averaging the bond transformation matrix over all possible conformational states of a residue or by a Monte Carlo method. However, when the peptide group is in the cis conformation, the interactions between neighboring residues cannot be neglected;[10] because of this, and also because ψ and ω are allowed to vary continuously, the usual matrix method for polypeptide chains[13–17,20] cannot be used.[26] Therefore, in this paper, we will apply a Monte Carlo method[18,19,21] to compute the characteristic ratio of poly(L-proline). Since we are interested only in the unperturbed dimensions of the poly(L-proline) chain in this paper, it is not necessary to check for long-range interatomic overlaps. However, it has not yet been ascertained how far the "short-range" interactions extend in evaluating the unperturbed dimensions of polymer chains.[10] In the present computation, the short-range interactions are extended to include the neighboring residue, for the reason discussed in section IIIA. It should be noted that it has been shown that the value of $\langle R^2 \rangle_0$ calculated

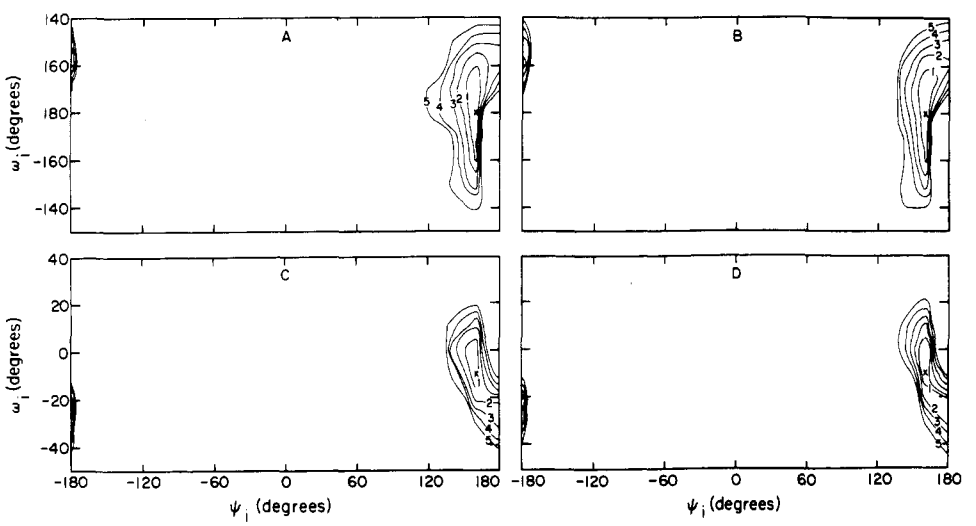

Figure 2. Conformational energy maps for the di-L-proline structure depicted in Figure 1A. In A and B, ω_i was restricted to the trans conformation (~180°), and in C and D to the cis conformation (~0°): (A) for $\omega_{i-1} = 175.0°$, (B) for $\omega_{i-1} = -8.7°$, (C) for $\omega_{i-1} = 175.0°$, and (D) for $\omega_{i-1} = -8.7°$. The approximate minima (approximate, because energy minimization was not carried out) are denoted by ×, and the contours are drawn at 1, 2, 3, . . . , etc., kcal/mol of dipeptide above the minima.

by a Monte Carlo method (i.e., from an appropriate finite number of chain conformations) converges to that computed by the exact matrix multiplication method.[19,21,27]

In order to carry out the calculation by a Monte Carlo method practically and effectively, it is inevitably necessary initially to eliminate conformations in which short-range interactions lead to high energies. For this purpose, we use the conformational energy maps of di-L-proline of Figure 2 and select only those conformations (in 10° increments) which lie within 5 kcal/mol of the minima. The actual numbers of conformational states for the ith proline residue, selected from the four maps of Figure 2, are 34, 20, 28, and 19 for the trans–trans, trans–cis, cis–trans, and cis–cis conformations of ω_{i-1} and ω_i, respectively.

In contrast to the structure of Figure 1A, the chains that are generated by the Monte Carlo procedure (from single-residue conformations selected from Figure 2) are terminated by real end groups, viz., acetyl at the N-terminus and methyl ester at the C-terminus, as in ref 4. We define the degree of polymerization, n', of the chain as including both end groups. According to the IUPAC-IUB nomenclature and conventions, the N-terminal end group and the atoms belonging to it are assigned the subscript 0 and the C-terminal end group is assigned the subscript $n' - 1$. The N-terminal end group was kept in the cis conformation, i.e., with C_0–H cis to C'_0–O_0, in Figure 2A of ref 4. The conformation of the C-terminal methyl ester group was taken (fixed) as follows. The dihedral angles for rotation about the $C^{\alpha}{}_{n'-2}$–$C'_{n'-2}$, $C'_{n'-2}$–$O^{ester}{}_{n'-1}$ and $O^{ester}{}_{n'-1}$–$C_{n'-1}H_3$ bonds are 147.4, 177.7, and 179.1° for $\omega_{n'-3}$ in the trans state (175.0°) and 167.7, 178.6, and −179.5° for $\omega_{n'-3}$ in the cis state (−8.7°), where $\omega_{n'-3}$ is the dihedral angle for rotation about the peptide bond, $C'_{n'-3}$–$N_{n'-2}$; these values yielded the minimum energy for the structure of Figure 2B of ref 4.

According to Brant and Flory,[13] a virtual bond, l', is defined by the constant distance between two successive C^{α} atoms when the peptide bond between them is in a planar trans conformation. Since the virtual bond length is not constant when rotation about the peptide bond is allowed, it was necessary to use the actual bond vectors $\mathbf{b}_{C^{\alpha}C'}$, $\mathbf{b}_{C'N}$, and $\mathbf{b}_{NC^{\alpha}}$; the magnitudes of these vectors are 1.530, 1.360, and 1.453 Å, respectively. For a chain of degree of polymerization n', there are $(n' - 1)$ virtual bonds, including the methyls of the terminal acetyl and methyl ester groups. However, since the geometry of the end groups is not the same as that used for the amino acid residues in the interior of the chain,[24] we define the end-to-end distance of the kth chain of degree of polymerization n' by means of the distance from $C^{\alpha}{}_1$ to $C^{\alpha}{}_{n'-2}$, neglecting the fragments at both termini (however, it should be noted that the acetyl and methyl ester end groups were included when the conformational energy was computed); the end-to-end distance of the kth chain of degree of polymerization n' is computed from the vectors connecting $C^{\alpha}{}_1$ and $C^{\alpha}{}_{n'-2}$, i.e., as

$$\mathbf{R}_k = \sum_{i=1}^{n} \mathbf{l}^*{}_i \tag{2a}$$

and

$$\mathbf{l}^*{}_i = \mathbf{b}_{C^{\alpha}{}_iC'{}_i} + \mathbf{b}_{C'{}_iN_{i+1}} + \mathbf{b}_{N_{i+1}C^{\alpha}{}_{i+1}} \tag{2b}$$

where n is the number of bond vectors of length $\mathbf{l}^*{}_i$ (which is the same as the number of repeating units, not counting the end groups) in the polymer chain and $n = n' - 3$. If the peptide bonds are maintained at $\omega = 180°$, the virtual bond length defined by Brant and Flory[13] is then given by $l'_i = |\mathbf{l}^*{}_i \cdot \mathbf{l}^*{}_i|^{1/2}$ and is constant (~3.8 Å). However, if ω varies, then the length of the bond vector $\mathbf{l}^*{}_i$ is not constant.

The conformations of the kth chain were generated by the following method. The assignment of the conformational state of the ith residue was made by the following two steps. First, the conformational state (trans or cis) of the ith *peptide bond* was determined by a pseudo-random number,[28] assuming a random occurrence of trans and cis conformations,[29] with the random numbers assigned so as to reproduce the desired fraction of trans and cis; in this same process, the conformational state of the $(i - 1)$th *peptide bond* was chosen randomly. Thus, a random choice was made as to which of the four conformational energy maps of Figure 2 was to be used in the second step. Second, the conformational states of the ith *residue*, i.e., the values of a set of ψ_i and ω_i, were then chosen randomly from the allowed area of one of the four conformational energy maps selected in the first step. These procedures were repeated until i reached n, and then $\mathbf{R}_k$ was computed by eq 2. No energies were calculated at this stage, these preliminary steps 1 and 2 (and the energy maps of Figure 2) serving *only* to select the values of ϕ_i and ψ_i for each residue.

One hundred chains (i.e., $N = 100$) were generated independently by this procedure, and then the mean-square end-to-end distance for this set was computed by

$$\langle R^2 \rangle_0 = \frac{\sum_{k=1}^{N} R_k{}^2 \exp(-E_k/RT)}{\sum_{k=1}^{N} \exp(-E_k/RT)} \tag{3}$$

where T and R are the absolute temperature and the gas constant, respectively. For use in eq 3, the conformational energy E_k of the kth chain was computed by taking into account the interaction energies within the ith residue, between the ith and $(i + 1)$th, and between the ith and $(i + 2)$th, and summing over i. In all cases, the conformations of the N-terminal methyl group and C-terminal methyl ester group were fixed, as indicated above. Substituting the values of $\langle R^2 \rangle_0$ and $\overline{l^2}$ defined by eq 3 and 4, respectively, into eq 1, the characteristic ratio, C, was computed.

$$\overline{l^2} = (b_{NC^{\alpha}}{}^2 + b_{C^{\alpha}C'}{}^2 + b_{C'N}{}^2)/3 \tag{4}$$

with the numerical value of $(\overline{l^2})^{1/2}$ being 1.449 Å.

II. Results

The characteristic ratios were computed by the Monte Carlo procedure for chains with $n = 100$ at a temperature of 30°, and the results are presented in Table II. The data in parentheses are the arithmetic averages of the four independent Monte Carlo calculations in each group. It has previously been shown[21] that the value of C for poly(L-lysine), obtained by a Monte Carlo method with $n = 40$, was essentially the same as that for infinite chain length; also, the values of C for polyglycine and poly(L-alanine) with $n = 30$ were only a few percent smaller than those for infinite chain length.[27] Thus, the choice of $n = 100$ in the present calculation seems to be adequate to reproduce the characteristic ratio for a sufficiently long chain. Indeed, in preliminary computations (not shown here), the calculated characteristic ratios for $n = 60$ fell within the ranges of values given for $n = 100$ in Table II, with similar scatter of the results.

The values of C in Table II, pertaining to poly(L-proline), are given for polymers with all-trans and all-cis peptide groups, and for polymers with predominantly trans peptide groups but containing 5 and 10% cis peptide groups, respectively. Values of C for higher cis content (except the 100% cis polymer) were not calculated because there are no experimental data available for such polymers.

Table II
Characteristic Ratios

Polymer chain		$\langle R^2\rangle_0/n\overline{l^2}$ [a] Calcd value ($\epsilon = 2$)	Calcd value ($\epsilon = \infty$)	Obsd value	Ref
Poly(L-proline)	All-trans	277	277		Present work
		193	193		
		249	249		
		112	89.7		
		(208)[b]	(202)[b]		
				153 ± 14[c]	22
				135 ± 11[d]	
				103 ± 7[e]	
				95 ± 7[f]	
	5% cis[g]	213	147		Present work
		154	155		
		117	91.9		
		70.5	81.2		
		(139)[b]	(119)[b]		
	10% cis[g]	46.4	47.1		Present work
		3.32	10.1		
		8.27	8.26		
		115	32.2		
		(43.2)[b,h]	(24.6)[b,h]		
	All-cis	80.2	80.2		Present work
		77.7	77.7		
		73.2	73.2		
		59.0	59.0		
		(72.5)[b,h]	(72.5)[b,h]		
	All-trans		799[i]		16
	All-trans	208[i]			20
		200[i]			
Polyglycine		14.9			14
Poly(L-alanine)		63.7			14
Poly(L-lysine)		59.8			21
Polysarcosine		20.4		12.4 ● 1.4[j]	17

[a] All the characteristic ratios in this table are expressed in terms of the present definition for the average square of the bond length, $\overline{l^2} = \frac{1}{3}(b_{NC_\alpha}^2 + b_{C_\alpha C'}^2 + b_{C'N}^2)/3$, instead of by the virtual bond scheme used in the original papers (ref 12-21) in which a virtual bond was defined by the constant distance between two adjacent α carbons (i.e., 3.8 Å) in the trans peptide conformation. [b] Arithmetic average values of each set of four computations, each of which consists of 100 Monte Carlo chains. [c] Measured in trifluoroethanol and in propionic acid at 5°. [d] Measured in trifluoroethanol propionic acid, and in acetic acid at 30°. [e] Measured in water at 5°. [f] Measured in water at 30°. [g] These values of "5%" and "10%" are average values for the 100 chains. Since a weighting factor is applied to each chain conformation in the computation of C, the relative energies of each conformation are automatically taken into account. For example, if the all-trans chain had an overwhelmingly lower energy than one with 5 or 10% cis, then this amount of cis would not have affected the computed value of C. [h] Addition of a small amount of cis to an all-trans (or of trans to an all-cis) chain will lower the value of C. That is why the value of C for the chain with 10% cis is lower than that for the all-cis chain. [i] These two values correspond to two different puckering conformations of the pyrrolidine ring. [j] This value was evaluated in ref 17a using the experimental data reported by Fessler and Ogston.[30]

To illustrate the effect of electrostatic interactions (i.e., to compare the values of C in solvents of high and low dielectric constant, ϵ), and also to compare the present results with those reported by Schimmel and Flory,[16] the calculations that were originally carried out with $\epsilon = 2$[4,5] were performed with $\epsilon = \infty$, and are also reported in Table II. In this computation, the same set of generated chains was used, and only the value of ϵ was changed in computing the energy.

For illustrative purposes, values of C computed by other authors for poly(L-proline),[16,20] polyglycine,[14] poly(L-alanine),[14] poly(L-lysine),[21] and polysarcosine[17] are given in Table II, with the values reported in the original papers being converted from the virtual bond scheme to the average-bond-length scheme used here by applying the relation given in ref 31. Available experimental values of C are also listed in Table II.

III. Discussion

There are three differing views about the conformational state of poly(L-proline) in solution.[32,33] (1) From X-ray and spectroscopic observations in the solid state, a regular helical structure was proposed (see, for example, ref 32). (2) Hydrodynamic and spectroscopic measurements were also interpreted in terms of an interrupted helical structure with flexible joints between regular helical sequences.[32,33] (3) Finally, from measurements of molecular weight, second virial coefficients, and intrinsic viscosity, it was proposed[22] that *high* molecular weight poly(L-proline) behaves as a random coil in trifluoroethanol, in propionic acid, in acetic acid, and in water. The characteristic ratio can be computed for each of these three models. For the regular helical conformation

$$\langle R^2\rangle_0/n\overline{l^2} = n(d^2/\overline{l^2}) \quad (5)$$

where d is the translation along the helix axis, per residue; i.e., C is proportional to the chain length. For a 100-residue chain of poly(L-proline), in its minimum-energy all-trans conformation[5] (form II), the value of C is 384. For an interrupted helical conformation, C may be calculated by making use of the theory for the form I $\rightleftarrows$ form II interconversion, in the manner used to compute the chain dimensions in the helix–coil transition, with a matrix multiplication method[12] in the absence of excluded volume effects or with a Monte Carlo method[19b] in the presence of excluded volume effects. For the third model, the random coil, the calculation of C may be carried out by the procedure described in this paper. Thus, before comparing the calculated results of any model with experiment, it is necessary to establish which model best describes the experimental results. In this respect, since the choice of appropriate model may not yet be a settled one, we confine ourselves here to a discussion of the calculations, and compare them to experimental data based on the random coil model, without committing ourselves as to the validity of this model[6] (see, also, section IIIC).

As can be seen in Table II, the results of the Monte Carlo calculation show some scatter from run to run. However, despite this, it can be seen that the values calculated for the all-trans conformation lie close to the experimental range.[22] Actually, the values of C computed for the polymer with 5% cis peptide groups agree better with the experimental values. Although poly(L-proline) is thought to exist in the all trans form (i.e., form II helix) in trifluoroethanol, in propionic acid, in acetic acid, and in water, it is not possible to rule out the possible presence of up to 5% cis peptide groups by currently used spectroscopic methods.

In any event, the data in Table II indicate that the introduction of a small amount of cis peptide groups into a predominantly trans chain of poly(L-proline) can influence its characteristic ratio drastically. This behavior arises mainly from the fact that, in a trans-rich chain, the end-to-end distance is decreased by the abrupt onset of the cis conformation (i.e., by abrupt bending in the interior of the chain) in between long sequences of trans groups. The higher the trans content (in a trans-rich chain) or the cis content (in a cis-rich chain), especially for the trans-rich chain (where a cis residue causes a chain reversal), the more drastic is the effect of introducing the other conformation. This is similar to the behavior of C in stereoirregular vinyl polymers[34] and in stereoirregular 1,4-polybutadiene[35] and 1,4-polyisoprene,[35] where the presence of a small amount of stereoirregularity reduces the values of C.

We have not computed the temperature coefficient of the characteristic ratio (which might have served as an additional test of our potential functions) for two reasons, viz., (a) such a computation would have required a knowledge of the temperature dependence of the empirical potentials (and such information is not available), and (b) there is scatter in the values of C computed here for a temperature of 30°. Thus, for these reasons it would not be valid to attribute any significance to a computed value of $d \ln C/dT$.

The scatter in the values of C, observed here for poly(L-proline), and elsewhere for stereoirregular polymers,[36] does not appear in similar calculations for polymers such as polyethylene[37] (where longer-range interactions are not taken into account). As indicated in the figure cited in ref 36, the mean value of a Monte Carlo calculation of C has significance despite the large scatter in the computed values. Thus, despite the scatter in the calculated value of C in Table II, the conclusions drawn here appear to be valid.

The scatter in the values of C, computed by the Monte Carlo technique, can be reduced (a) by taking a larger number, N, of chains in the sample, and (b) by increasing the efficiency of the Monte Carlo technique by introducing somewhat longer-range interactions in the initial procedure in which conformations are selected, thereby reducing the ultimately calculated energies (or increasing the statistical weights) of the chains included in the Monte Carlo sample. In the procedure used here, the conformations were selected from the maps of Figure 2, based on interactions in di-L-proline and, when the energy E_k was calculated, interactions between the ith and $(i + 1)$th and between the ith and $(i + 2)$th residues were taken into account. However, if these longer-range interactions had been taken into account earlier, i.e., in the procedure by which conformations were selected, as will be discussed in section IIIA, then more of the generated chains would have had a high statistical weight in the computation of C. The interactions between the oxygen atom of the carbonyl group of the $(i + 1)$th residue and either the H^{α}_{i-1} atom or the atoms of the pyrrolidine ring of the $(i - 1)$th residue (see Figure 1B) are especially important in determining the values of ψ_{i+1}, ω_i, ψ_i, ω_{i-1}, and ψ_{i-1} when ω_i for the cis conformation is preceded by ω_{i-1} of another cis conformation.

Table III
Allowed Conformations of *cis*-Proline Preceded by *cis*-Proline[a]

Conformation of (i-1)th Residue	Conformation of (i)th Residue — Conformational Energy (kcal/mole)[b]										
	-10	-8	-6	-4	-2	0	2	4	6	8	10
I					VI	VII V		IX	VIII	X	
II[c]		VI	IX VII	V		VIII	X				
III						VI	IX VII	V X		VIII	
IV							VI	IX VII	V	X II	VIII
V[c]						V		IX		X	
VI[c]	VI	IX VII	V X	VIII							
VII[c]		VI	IX	VII	V	VIII		I			
VIII						IX VII		V X VI	V VII	I	
IX[c]	VI		IX VII	V X		VIII					
X[c]			X VI	IX VII		V		I VIII		IV	

[a] Only those conformations having energies less than 10 kcal/mol per mol of tripeptide are tabulated. [b] The conformational energy is a function of ψ_{i-1}, ω_i, ψ_i, ω_{i-1}, and ψ_{i-1}. The values listed are the lowest of those calculated for three values of ψ_{i-1} (viz., 140, 160, and 180°). [c] If residue $i - 2$ is trans, and $i - 1$ and i are both cis, then the energies are too high to appear in this table if residue $i - 1$ has the conformation II, V, VI, VII, IX, or X. The values listed in the table (for residue $i - 1$ in conformation II, V, VI, VII, IX, or X) pertain only to the case where residue $i - 2$ is cis. A trans is possible for residue $i - 2$ for all other conformations listed in the table.

A. Inclusion of Longer-Range Interactions. In order to see the effect of longer-range (i.e., next nearest neighbor) interactions, especially for a cis–cis sequence, additional calculations were carried out for the tripeptide structure of Figure 1B. The initial values of ψ_{i+1}, ω_i, ψ_i, ω_{i-1}, and ψ_{i-1} were chosen in 20° intervals from the low-energy range (within 5 kcal/mol of the minimum) of the dipeptide energy map of Figure 2D (cis–cis). The actual values selected are indicated by the Roman numerals I to X in Figure 3. The conformational energies of the tripeptide of Figure 1B were computed for all possible combinations of states I to X in (ψ_i, ω_i) and (ψ_{i-1}, ω_{i-1}), and of 140, 160, and 180° in ψ_{i+1}. The results are diagrammed in Table III, in which the rows represent the conformational states of the $(i - 1)$th residue, and the columns represent those of the ith residue; the positions of the Roman numerals within the body of the table designate the conformational energy for the combination of the conformational states of the $(i - 1)$th and ith residues. The lowest-energy conformation has an energy of −10 kcal/mol. Those conformations with energies higher than −4 kcal/mol (i.e., greater than that of the low-

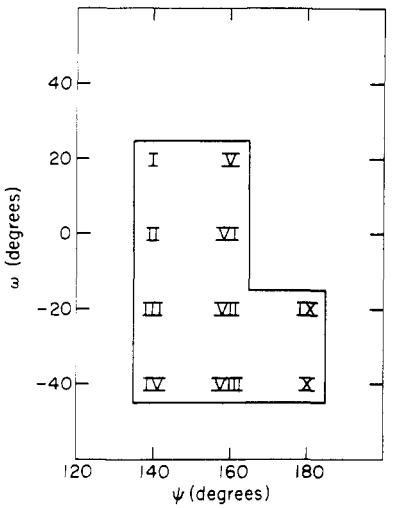

Figure 3. Designation of the ten conformational states (ψ, ω), selected from Figure 2D, for computations of the energy of the structure of Figure 1B for all-cis peptide groups.

est one by at least 6 kcal/mol) would not be expected to contribute to the conformational state of the whole chain. Hence, it can be seen that the conformational states of the combination of residues $(i - 1)$ and i are extremely restricted for a cis–cis combination. Thus, previously used methods[13-17,20] to compute C, in which the interresidue interactions are neglected, are not applicable to poly(L-proline) when cis residues are involved. This conclusion also shows that, if (instead of using the maps of Figure 2) such correlations of conformational states for $(\omega_{i-1}, \psi_{i-1})$ and (ω_i, ψ_i) had been taken into account at the stage of chain *generation* (notice that this does not mean that the conformational energy is calculated at this stage), a Monte Carlo calculation would have been more effective (i.e., most of the generated chains would have had low energy since such atomic overlaps would have been eliminated at the outset for the cis–cis combination); this might lead to less scatter in the computed values of C when cis residues are present. In the generation method used in this paper, the conformational state of the ith residue was chosen independently within the dipeptide energy maps (given in Figure 2), without any correlation between $(\omega_{i-1}, \psi_{i-1})$ and (ω_i, ψ_i), but only between ω_{i-1} and ω_i. In this respect, one can expect that the richer the cis content becomes, the higher will be the probability with which such unfavorable high energy conformations might occur, as seen in Table III, i.e., the less reliable the results become (because most of the generated chains have high energy, and thus contribute very little to the computed values of C). The values of C for the all-cis polymer, as seen in Table II, are not as scattered as those found for the all-trans, and for the 5 and 10% cis polymers. This is due mainly to the fact that the generated chains for the all-cis polymer have a narrow distribution with respect to end-to-end distance, i.e., a small variation of end-to-end distance among the generated chains. This behavior arises for cis–cis, because the directions of the bond vectors in a residue are restricted to the narrower area allowed in the dipeptide energy map (see Figure 2D and Table III) than that allowed for trans–trans, trans–cis, or cis–trans (see Figure 2A–C), in addition to the absence of an abrupt bending conformation. Thus, there is less scatter in the values of C (for all cis) in spite of the fact that most of the generated chains might have high energy, because the bond vectors in Figure 2D are restricted to a very small region.

The effect of electrostatic interactions on C of poly(L-proline) is detectable, but not very large as seen in Table II. This results mainly from the fact that the variation of the electrostatic energy is small within the highly restricted conformational space of poly(L-proline). Therefore, the relative stabilities among the generated chains for $\epsilon = \infty$ do not differ significantly from those for $\epsilon = 2$.

B. Comparison with Other Results. In comparing the results obtained here with those of others, it should be noted that, when ω_i is held fixed, the variation of the energy with ψ_i will depend on the geometry of the residue and on the potential functions used in the computations.[25] With the geometry and potential functions used here, there is only one low-energy minimum (near $\psi_i \approx 160°$)[4] for the structure of Figure 1A; a minimum near $\psi_i \approx -60°$ has an energy at least 80 kcal/mol higher[4] than that near $\psi_i \approx 160°$. In addition, there is a very high energy barrier (>10^5 kcal/mol) between these two minima. The existence of a high-energy minimum near $\psi_i \approx -60°$ and of a very high energy barrier between the two minima is not affected very much if ω_i is varied by as much as ±40° around the positions $\omega_i = 180°$ and $\omega_i = 0°$.

If there is indeed a very high energy barrier between the two minima, then the two conformational states cannot interconvert, and the system has to be treated as a copolymer, i.e., as a mixture of two species whose relative amounts depend on the mechanism of polymerization when the chains are synthesized. While the minimum near $\psi_i \approx -60°$ is accessible in the isolated proline monomer and in monomeric derivatives such as *N*-acetyl L-proline-*N'*-methyl ester,[4,5] there are no X-ray data to indicate its occurrence in long chains of poly(L-proline). Its lack of occurrence in long chains of poly(L-proline) cannot be attributed to an augmentation of its energy by its neighbors in the chain because the augmentation would not be high enough to overcome the extremely high barrier between the two minima; instead, its absence is likely due to the high energy at this minimum. If, indeed, the system is a copolymer, i.e., a mixture of two noninterconverting species, then the transformation matrices of the virtual bond scheme[12,13] must be averaged *separately* over each minimum instead of over the *entire* range of ψ_i, and then combined according to the relative probability of occurrence of each conformation, in order to compute C.

With these comments in mind, we may examine previous calculations of C for poly(L-proline). The computation of Schimmel and Flory,[16] based on a flat pyrrolidine ring and $\omega = 180°$, led to only one low-lying energy minimum in ψ_i (near $\psi_i = 160°$) and a value of C which, according to Mattice and Mandelkern,[22] was too large to account for the experimental values (see Table II). Mattice et al.[20] then computed C with several models in which both ϕ_i and ψ_i were varied, and two puckered conformations of the pyrrolidine ring were considered. Most of their models included *two* accessible minima in ψ_i separated by a high barrier (and with ω held fixed) and involved averaging over the *whole* ψ_i space. For the reasons cited in the previous paragraph, it appears to us that these models may not be applicable. However, in one of their models, they calculated C by neglecting the minimum near $\psi_i = -60°$ (which arises from *their* geometry and potential function[25]) and averaging only over the minimum near $\psi_i = 160°$. For the reason cited above, these values of C should be theoretically reasonable, and hence are shown in Table II.

A similar problem, involving two minima separated by a large barrier, was encountered by Tanaka and Nakajima[17a] in the calculation of C of poly(*N*-methyl-L-alanine) and polysarcosine. They found that the conformation of poly(*N*-methyl-L-alanine) was extremely restricted to two narrow areas with an intervening high-energy barrier as in

poly(L-proline), while that of polysarcosine, in which the methyl group of poly(*N*-methyl-L-alanine) is replaced by a hydrogen atom, did not involve such a high-energy barrier. Recent calculations of energy maps for those homopolymers by Mattice[38] confirmed that poly(*N*-methyl-L-alanine) has two minima with an intervening high-energy barrier, independent of the rotation of the methyl on the nitrogen or of the geometric parameters. By preventing the conformations of poly(*N*-methyl-L-alanine) from converting from one potential well to another in the single-residue energy map, Tanaka and Nakajima[17b] selected the conformation of the *N*-methylalanine residue at one of the local minima, i.e., the lower-energy minimum at the left-handed α helix, to calculate C for a copolypeptide of *N*-methylalanine residues with various other amino acid residues.[17b]

In summary, when there is a high barrier between two potential wells, as in the case of the energy map of Mattice et al.[20] for di-L-proline, the bond vectors should not be averaged over the entire conformational space of a residue. In addition, the effect of next-nearest-neighbor interactions may play an important role in determining the value of C, as pointed out in section IIIA. By allowing for some degree of freedom of rotation about the peptide bond, and by including a small amount of the cis conformation in a trans-rich chain, as was done in the computations reported here, it appears to be possible to account for the observed values of C of poly(L-proline) with the geometry and potential functions used earlier.[4,5,24,25] While we have used only one puckered conformation of the pyrrolidine ring in these computations, it may also be possible to account for the observed values of C by introducing flexibility into the pyrrolidine ring; this might lower the barrier between the two minima in ψ_i, or might possibly broaden the low-lying potential well. However, any freedom that is allowed to the pyrrolidine ring should be consistent with X-ray and NMR data.

C. Concluding Remarks. From the discussion of section IIIB, it appears that the values of $\langle R^2\rangle_0/nl^2$ for poly(L-proline) have been accounted for reasonably well, under the assumption that this polymer exists in the random-coil conformation in such solvents as trifluoroethanol, propionic acid, acetic acid, and water. As described in the first paragraph of section III, it is of importance to prove *experimentally* whether poly(L-proline) exists as a random-coil conformation, or as some other possible structure such as the form I and form II helices, or as the interrupted helix, in these solvents. Mattice and Mandelkern[22] have assumed that the random-coil model applies (based on the departure of a log $[\eta]$ vs. M plot from linearity at high molecular weight M). They have, therefore, used the standard procedures to calculate the characteristic ratio, with results that are in the range calculated here, using this same model. However, the possibility remains that poly(L-proline) may exist as a broken rod, i.e., an interrupted helix; this possibility is based on the fact that the intrinsic viscosity $[\eta]$ of a rod-like molecule with a *single* break is about 15% less than that of an unbroken rod,[39,40] and a rod with more than one break probably has an even lower intrinsic viscosity.

As a final conclusion, we have calculated the characteristic ratio under the hypothesis that poly(L-proline) may be treated as a random coil. A reasonable account has been offered for the observed vatues of $\langle R^2\rangle_0/n\bar{l}^2$ of poly(L-proline) by introducing some degree of rotational freedom about the peptide bond, and a small amount of cis in a trans-rich chain.

Addendum

Just before submission of this manuscript, Dr. Leo Mandelkern informed us that he and his colleagues (C. C. Wu and R. A. Komoroski) had obtained Fourier transform 270 MHz proton NMR evidence for 2–3% of cis peptide bonds in a trans chain of poly(L-proline) in D_2O. These observations support our conclusion that approximately 5% cis is required to account for the value of C of a (predominantly trans) chain of poly(L-proline); i.e., it is reasonable for us to have assumed that a predominantly trans chain of poly(L-proline) may have a small percentage of cis residues.

Appendix

Consideration of Matrix Method for Evaluating C for Poly(L-proline). We consider here the applicability of the matrix method to the evaluation of C for randomly coiling poly(L-proline), and show that this procedure is not practical for this homopolymer.

In general, the conformations of chains of poly(L-proline) of N residues can be described by a set of rotational states ϕ_i, ψ_i, and ω_i (see Figure 1B) and by a set of puckering conformations of the pyrrolidine rings. We designate the conformational states related to ϕ_i, ψ_i, and ω_i by Ω and those related to the puckering conformations by Ω'. The sets Ω and Ω' may consist of discrete states (rotational isomeric states) or continuously varying states. In particular, it is not necessary to restrict ϕ_i and the χ_i's to specific puckering states, such as U and D.

As described in the text, in treating poly(L-proline), it is necessary to take into account inter-residue interactions: those between residues $i-1$ and i (first-neighbor interactions), and those between residues $i-1$ and $i+1$ (second-neighbor interactions). Therefore, in order to obtain the partition function of randomly coiled poly(L-proline), it is necessary to correlate the conformational states of the three consecutive residues $i-1$, i, and $i+1$ (see Figure 1B). Of course, one can easily extend this treatment to the nearest-neighbor model, and also to a model that takes into account interactions with residues beyond second neighbors.

As seen in Figure 1B, the conformation of a residue can be described by the three dihedral angles ϕ, ψ, and ω, and by the puckering conformation of the pyrrolidine ring. We define the rotational states of the three consecutive residues by $\Omega_{i-1}(\phi_{i-1},\psi_{i-1},\omega_{i-1},\phi_i,\psi_i,\omega_i)$ and $\Omega_i(\phi_i,\psi_i,\omega_i,\phi_{i+1},\psi_{i+1},\omega_{i+1})$, and designate the number of conformational states to be employed for the ith residue as $\eta'(\phi_i)$, $\eta'(\psi_i)$, and $\eta''(\omega_i)$. Then, the total number of allowed conformational states for Ω_i is

$$N(\Omega_i) = \eta'(\phi_i)\eta'(\psi_i)\eta''(\omega_i)\eta'(\phi_{i+1})\eta'(\psi_{i+1})\eta''(\omega_{i+1}) \quad \text{(A-1)}$$

In $\eta''(\omega_i)$, and correspondingly for $\eta''(\omega_{i+1})$, we may limit the rotational states only to those around the trans ($\omega = 180°$) and cis ($\omega = 0°$) conformations because of the partial double bond character of the peptide bond. Then,

$$\eta''(\omega_i) = \eta''_t(\omega_i) + \eta''_c(\omega_i) \quad \text{(A-2)}$$

[and similarly for $\eta''(\omega_{i+1})$], where $\eta''_t(\omega_i)$ and $\eta''_c(\omega_i)$ are the number of rotational states of ω_i in the small region around the trans and cis conformations, respectively. The total number of allowed conformational states for Ω'_i is designated as $N(\Omega'_i)$.

The conformational energy of three successive residues can then be described in terms of the set Ω_i, Ω_{i+1} related to the backbone conformations and Ω'_{i-1}, Ω'_i, and Ω'_{i+1} related to the puckering conformations of the pyrrolidine rings of residues $i-1$, i, and $i+1$. For the sake of convenience, we will start by considering only those conformational

states of the backbone which correspond to certain specific states of the pyrrolidine ring; then, we will allow for all possible states of puckering. For specific puckering states, Ω'_{i-1}, Ω'_i, and Ω'_{i+1}, the conformational energy may be calculated as a function of Ω_{i-1} and Ω_i, i.e., as $E(\Omega_{i-1},\Omega_i)$. We may then construct the statistical weight matrix for the ith residue as

$$\mathbf{u}(\Omega'_{i-1},\Omega'_i,\Omega'_{i+1}) = [\exp\{-E(\Omega_{i-1},\Omega_i)\}/RT] \quad \text{(A-3)}$$

for a specific puckering state Ω'_{i-1}, Ω'_i, and Ω'_{i+1}. In eq A-3, the element (Ω_{i-1},Ω_i) is given by the Boltzmann factor of the energy, which is determined by the states $\Omega_{i-1}(\phi_{i-1},\psi_{i-1},\omega_{i-1},\phi_i,\psi_i,\omega_i)$ and $\Omega_i(\phi_i,\psi_i,\omega_i,\phi_{i+1},\psi_{i+1},\omega_{i+1})$. Therefore, the order of the matrix in eq A-3 is $N(\Omega_{i-1}) \times N(\Omega_i)$.

We next vary the puckering conformations Ω'_{i-1}, Ω'_i, and Ω'_{i+1} of the pyrrolidine ring. Using the submatrix of eq A-3, we will construct the statistical weight matrix so that all possible conformations of the pyrrolidine rings of residues $i-1$, i, and $i+1$ are allowed. The order of this matrix will be $[N(\Omega'_i)^2 \times N(\Omega_i)] \times [N(\Omega'_i)^2 \times N(\Omega_i)]$. For example, when two puckering conformations, such as U and D, are used [i.e. when $N(\Omega'_i) = 2$], the statistical weight matrix is given as follows:

$\mathbf{U}_i =$

	$i+1$	U	D	U	D
$i-1$	i	U	U	D	D
U	U	u(UUU)	u(UUD)	0	0
U	D	0	0	u(UDU)	u(UDD)
D	U	u(DUU)	u(DUD)	0	0
D	D	0	0	u(DDU)	u(DDD)

(A-4)

where the indices of the columns and rows designate the puckering conformations of three successive residues, Ω'_{i-1}, Ω'_i, Ω'_{i+1}. By using eq A-3, the elemental matrix $\mathbf{u}(\Omega'_{i-1},\Omega'_i,\Omega'_{i+1})$ for any conformational states Ω'_{i-1}, Ω'_i, and Ω'_{i+1} is replaced by $\mathbf{u}$(UUU), $\mathbf{u}$(UUD), etc., in eq A-4. The **0** in eq A-4 is the null matrix of order $N(\Omega) \times N(\Omega)$. Neglecting the special interactions at the ends of the chain, the statistical weight matrices for residues 1 and N at the ends can be written as

$$\mathbf{U}_1 = \begin{bmatrix} \mathbf{u}(UU) & \mathbf{u}(UD) & 0 & 0 \\ 0 & 0 & \mathbf{u}(DU) & \mathbf{u}(DD) \\ \mathbf{u}(UU) & \mathbf{u}(UD) & 0 & 0 \\ 0 & 0 & \mathbf{u}(DU) & \mathbf{u}(DD) \end{bmatrix} \quad \text{(A-5)}$$

and

$$\mathbf{U}_N = \begin{bmatrix} \mathbf{u}(U) & \mathbf{u}(U) & 0 & 0 \\ 0 & 0 & \mathbf{u}(D) & \mathbf{u}(D) \\ \mathbf{u}(U) & \mathbf{u}(U) & 0 & 0 \\ 0 & 0 & \mathbf{u}(D) & \mathbf{u}(D) \end{bmatrix} \quad \text{(A-6)}$$

We then calculate the partition function as

$$Z = \boldsymbol{\epsilon}\mathbf{U}_1\left(\prod_{i=2}^{N-1}\mathbf{U}_i\right)\mathbf{U}_N\boldsymbol{\epsilon}^* \quad \text{(A-7)}$$

where

$$\boldsymbol{\epsilon} = (\mathbf{e}_1,\mathbf{e}_1,\mathbf{e}_1,\mathbf{e}_1) \quad \text{(A-8)}$$

and

$$\boldsymbol{\epsilon}^* = \begin{pmatrix} \mathbf{e}^*_N \\ \mathbf{e}^*_N \\ \mathbf{e}^*_N \\ \mathbf{e}^*_N \end{pmatrix} \quad \text{(A-9)}$$

where $\mathbf{e}_1$ is the unit row vector of order $N(\Omega)$ and $\mathbf{e}^*_N$ is the unit column vector of order $N(\Omega)$, i.e.,

$$\mathbf{e}_1 = (1, 1, \ldots, 1) \quad \text{(A-10)}$$

and

$$\mathbf{e}^*_N = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{(A-11)}$$

For poly(L-proline) with a specific conformation of the pyrrolidine rings (say, all D conformations), the case considered in the text, $N(\Omega'_i) = 1$. Hence, the partition function can be calculated, using eq A-3, as

$$Z = \mathbf{e}_1\mathbf{u}_1\left[\prod_{i=2}^{N-1}\mathbf{u}_i(\Omega'_{i-1},\Omega'_i,\Omega'_{i+1})\right]\mathbf{u}_N\mathbf{e}^*_N \quad \text{(A-12)}$$

where $\mathbf{u}_1$ and $\mathbf{u}_N$ are given by the corresponding elements (submatrices) of eq 5 and 6 that correspond to the specific D conformational states of the pyrrolidine rings when $\Omega'_{i-1},\Omega'_i,\Omega'_{i+1}$ = DDD.

Using the statistical weight matrices and the partition function, one can formulate the equations to calculate the characteristic ratio C by using Flory's method.[12] However, we are not concerned with this formulation here, but only in ascertaining the order of the statistical weight matrix.

We now consider the order of matrices of eq A-3 or A-4 that are required to calculate the partition function of eq A-7 or A-12 for poly(L-proline). If we allow only one conformation for the pyrrolidine ring (e.g., D puckering and fixed ϕ), then $N(\Omega') = 1$ and $\eta'(\phi_i) = 1$. Taking six rotational states for each of the trans and cis regions for the peptide bond, $\eta''_t(\omega_i) = \eta''_c(\omega_i) = 6$, eq A-2 yields $\eta''(\omega_i) = 12$. Furthermore, if we choose $\eta'(\psi_i) = 12$, eq A-1 gives $N(\Omega_i) = 1 \times 12 \times 12 \times 1 \times 12 \times 12 = 12^4$. Thus, in eq A-12, we would have to use $12^4 \times 12^4$ order matrices for $\mathbf{u}_i$, $\mathbf{u}_1$, and $\mathbf{u}_N$; i.e., the conformational energy would have to be calculated for $12^4 \times 12^4$ states of three consecutive residues. In addition to these extensive energy calculations, the need to multiply $12^4 \times 12^4$ order matrices would make it impossible to calculate the partition function from eq A-12. Of course, if one reduced the number of allowed rotational states, the computation might become feasible.

If two puckering conformations were allowed for the pyrrolidine ring (say, U and D), and the number of rotational states of the backbone were the same as computed above, the order of the matrices of eq A-4, A-5, and A-6 would be $(16 \times 12^4) \times (16 \times 12^4)$ since $N(\Omega') = \eta'(\phi_i) = 2$.

Thus, in conclusion, too much computational effort would be required in order to apply the matrix method to the present problem.

Note Added in Proof. While this paper was in press, a paper by C. M. Venkatachalam, B. J. Price, and S. Krimm, *Biopolymers*, **14**, 1121 (1975), on the conformational energy of a pro-pro dipeptide (similar to that of our Figure 1A), appeared. They found *three* stable conformations which

they called cis ($\omega = 0°$, $\psi = 160°$), trans (or trans′) ($\omega = 180°$, $\psi = 160°$), and cis′ ($\omega = 180°$, $\psi = -40°$), with a cis′–trans′ barrier of the same order of magnitude (20 kcal/mol) as the cis–trans barrier. First of all, it should be noted that, as emphasized here and earlier,[4,5] the values of ω and ψ for stable conformations depend very much on the bond lengths, bond angles, and potential functions used in the computations; their different conformations arise from the use of different geometry and potential functions than those used here. In any event, with such a high barrier, all three conformations should be observable by NMR in proline chains longer than the dipeptide but, as they point out, D. E. Dorman, D. A. Torchia, and F. A. Bovey, *Macromolecules*, **6**, 80 (1973), found only two separate resonances for poly(L-proline), which does not support the existence of three conformations. Conceivably, special circumstances involving the rates of interconversion of these conformations could have prevented the appearance of three *separate* resonances. In summary, on the basis of our geometry for the puckered pyrrolidine ring[4,5,24] (and our potential functions,[24] that were refined by computations on crystals) and the NMR evidence cited above, we do not believe that the cis′ conformation is a low-energy one.

Acknowledgment. We are indebted to Dr. Leo Mandelkern for helpful discussion of this paper, and for providing the information in the Addendum.

References and Notes

(1) (a) This work was supported by research grants from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312), from the National Science Foundation (BMS71-00872 A04), and from Walter and George Todd. (b) Part of this paper was presented at a symposium on "Morphology of Polymers at the Molecular Level", Stanford University, Palo Alto, Calif., August 1974.

(2) (a) From Kyoto University, 1972–1975; (b) address correspondence to this author.

(3) (a) For reviews, see H. A. Scheraga, *Chem. Rev.*, **71**, 195 (1971); (b) "Peptides, Polypeptides, and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Ed., Wiley, New York, N.Y., 1974, p 49.

(4) S. Tanaka and H. A. Scheraga, *Macromolecules*, **7**, 698 (1974).

(5) S. Tanaka and H. A. Scheraga, *Macromolecules*, **8**, 494, 504, 516 (1975).

(6) We present here a treatment of a random-coil model for poly(L-proline) without committing ourselves as to the validity of this model for this homopolymer.

(7) P. J. Flory, "Principles of Polymer Chemistry", Cornell University Press, Ithaca, N.Y., 1963, Chapters X and XII.

(8) It has been shown elsewhere[9] that, while proteins in aqueous solution do not meet the conditions for ideal behavior, nevertheless their conformations are determined in large measure (though not exclusively) by short-range interactions.

(9) H. A. Scheraga, *Pure Appl. Chem.*, **36**, 1 (1973).

(10) The question as to how far the short-range interactions extend, at the θ point, is still an open one; i.e., they may extend beyond the single residue, to include interactions with one or two neighboring residues. See ref 11 for a discussion of this point.

(11) M. Go, F. T. Hesselink, N. Go, and H. A. Scheraga, *Macromolecules*, **7**, 459 (1974).

(12) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969, Chapter VII.

(13) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2791 (1965).

(14) D. A. Brant, W. G. Miller, and P. J. Flory, *J. Mol. Biol.*, **23**, 47 (1967).

(15) W. G. Miller, D. A. Brant, and P. J. Flory, *J. Mol. Biol.*, **23**, 67 (1967.

(16) P. R. Schimmel and P. J. Flory, *Proc. Natl. Acad. Sci. U.S.*, **58**, 52 (1967).

(17) (a) S. Tanaka and A. Nakajima: *Polym. J.*, **2**, 717 (1971); (b) 725 (1971).

(18) H. E. Warvari and R. A. Scott, *J. Chem. Phys.*, **57**, 1146 (1972).

(19) (a) S. Tanaka and A. Nakajima, *Macromolecules*, **5**, 708 (1972); (b) 714 (1972).

(20) W. L. Mattice, K. Nishikawa, and T. Ooi, *Macromolecules*, **6**, 443 (1973).

(21) F. T. Hesselink, T. Ooi, and H. A. Scheraga, *Macromolecules*, **6**, 541 (1973).

(22) W. L. Mattice and L. Mandelkern, *J. Am. Chem. Soc.*, **93**, 1769 (1971).

(23) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).

(24) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, in press.

(25) We emphasize that the geometry adopted[24] for the proline residue was based on an analysis of all available X-ray data. The choice was made, using the following criteria. (a) The reliability index (R factor) had to be small. (b) The analysis of the X-ray data should not have been biased by the introduction of a geometric model (such as that of Pauling and Corey for the peptide group). (c) The structure had to be free of structural features (such as ionized end groups, metal complexes, small strained rings, etc.) which would make the data inapplicable for our purpose. With these selected (X-ray) bond lengths and bond angles, the imposition of the requirement that the pyrrolidine ring be closed led to two ring-puckering conformations designated as "up" and "down", where these terms refer to the relative positions of the C^γ and C^β atoms. The values for ϕ and χ^1 for the "up" and "down" conformations are given in section IA. In *both* the "up" and "down" forms, both the C^β and C^γ atoms lie out of the plane defined by the C^α, N, and C^δ atoms; i.e., ring puckering occurs at both C^γ and C^β, but predominantly at C^γ. In addition, based on X-ray data, the bond angles around the nitrogen atom in cis and trans X-Pro bonds differ.[24] Finally, as pointed out on pp 701 and 702 of ref 4, we argued that NMR data do not rule out the existence of a single puckered conformation of the pyrrolidine ring in a *long* chain of poly(L-proline). Since the computed potential energy of a proline oligomer is sensitive to the geometry, it is important to realize that the adopted geometry[24] was selected only after an exhaustive and thorough analysis of the X-ray data, applying the criteria described above.

(26) In treating polypeptides, the matrix method has usually been applied to chains with planar trans peptide groups.[13–17,20] In principle, one can formally formulate this method to include cis peptide bonds, various states of puckering, and situations where longer-range interactions are involved (see Appendix). However, the matrix method cannot be used here for reasons cited in the Appendix.

(27) S. Premilat and J. Hermans, Jr., *J. Chem. Phys.*, **59**, 2602 (1973).

(28) Uniform pseudo-random numbers were generated by using RANDU of the IBM Scientific Subroutine Package.

(29) By allowing for a random occurrence of trans and cis conformations, we are not contradicting the results of our earlier treatment[5] of the form I $\rightleftharpoons$ form II interconversion as a cooperative process. If randomly occurring cis residues (in a trans-rich chain) were to have a lower statistical weight than if the cis residues occur in a block, the Monte Carlo procedure will automatically recognize this when the statistical weight of each generated chain conformation is computed. Thus, the Monte Carlo procedure allows for all possible distributions of cis residues, and the energy calculation gives an appropriate weighting factor to each conformation.

(30) J. H. Fessler and A. G. Ogston, *Trans. Faraday Soc.*, **47**, 667 (1951).

(31) The characteristic ratios calculated from the virtual bond scheme were converted to the average-bond-length scheme used in this paper by means of the following relation: $\langle R^2\rangle_0/n\bar{l}^2 = [(n'-1)l'^2/(n'-3)\bar{l}^2]C' \simeq (l'^2/\bar{l}^2)C'$ at large n', where C' is the value of $\langle R^2\rangle_0/nl'^2$ computed by the virtual bond scheme with a virtual bond length l', and $\bar{l}^2$ is defined in eq 4.

(32) E. Katchalski, A. Berger, and J. Kurtz, "Aspects of Protein Structure", G. N. Ramachandran, Ed., Academic Press, New York, N.Y., 1963, p 205.

(33) L. Mandelkern, "Poly-α-amino Acids," G. D. Fasman, Ed., Marcel Dekker, New York, N.Y., 1967, p 675.

(34) See Figure 11 on p 232 of ref 12.

(35) S. Tanaka and A. Nakajima, *Polym. J.*, **3**, 500 (1972).

(36) See especially Figure 13 of Chapter VI of ref 12, where the scatter becomes large in the range of a low fraction of one component ($w_{iso} \sim 0.9$ in the reference cited).

(37) D. K. Carpenter, G. Santiago, and B. Zierenberg, *Polym. Prepr., Am. Chem. Soc., Div. Polym. Chem.*, **14**, 204 (1973).

(38) W. L. Mattice, *Macromolecules*, **6**, 855 (1973).

(39) H. Yu and W. H. Stockmayer, *J. Chem. Phys.*, **47**, 1369 (1967).

(40) A. Teramoto, T. Yamashita, and H. Fujita, *J. Chem. Phys.*, **46**, 1919 (1967).