Calculation of Conformational Properties of Oligomers of L-Proline[1]

Seiji Tanaka[2] and Harold A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14850.
Received January 24, 1974

ABSTRACT: Allowing for two conformations of the pyrrolidine ring and for cis–trans isomerization in the peptide group, the fractions of various conformers of oligomers of L-proline are calculated. Besides the conformations of the peptide group and the pyrrolidine ring, the conformational freedom of the backbone was confined to the rotation about the C^{α}–C′ bonds. In the calculations of the minimum-energy conformations of each oligomer, the interactions over the *whole* molecule were taken into account. The minimum-energy conformations were employed as the rotational isomeric states of the whole molecule. The statistical weights of these rotational isomeric states were calculated in two ways, both with and without the inclusion of the librational entropy (calculated by assuming that each conformer undergoes *small* fluctuations in conformation). Both sets of statistical weights were used to compute the fractions of conformers, which were then compared with experimental results. The calculated probabilities of occurrence of the trans conformation at the ith residue (in the *interior* of the chain) appear to be in good agreement with experimental values, when allowances are made for end-group effects. The conformational energy and the fractions of conformers are affected markedly by the bulkiness of the group at the C terminus of the chain, especially in short oligomers. The preference for the all-trans conformer in penta-L-proline arises only when the librational entropy is included, since the conformer ttctc is the more stable one *energetically*.

Conformational energy calculations have been developed and applied to the determination of the stable conformations of polypeptides and proteins and to the evaluation of thermodynamic quantities such as the conformational free energy of the polypeptide chain.[3] In this development, cognizance was always taken of the importance of checking the validity of the computational procedure by comparing the results with experimental facts wherever possible. In this paper, we compute the conformational properties of oligomers of L-proline and compare the results with experimental data, in order to assess the reliability of our procedures for treating proline residues, not only in oligopeptides but also in proteins. In this work, we make use of the geometry and energy parameters which we have deduced earlier[4,5] for the L-proline residue.

The conformational properties of the polypeptide chain molecule that have usually been studied experimentally are the fractions of various conformational states[6] (deduced from infrared and nuclear magnetic resonance measurements), the average unperturbed end-to-end distance,[7-9] and the dipole moment[10] and optical anisotropy.[11] Two models have been used in the statistical mechanical treatment of chain molecules, *viz.*, the (rotational isomeric) three-state model, which has been applied to the *n*-alkane chain,[8] and a continuous state model (without the inclusion of interactions between neighboring residues), which has been used to compute the properties of the randomly coiled polypeptide chain.[7-9] In the continuous-state model, the statistical weights for each conformational state have been evaluated[7-9] at discrete values (in appropriate increments) of the rotational angles. In this paper, we will use the rotational isomeric state model to calculate the distribution among cis and trans states of the peptide group for oligomers of L-proline. Besides the cis–trans isomerism about the peptide bond, account will be taken of the puckering of the pyrrolidine ring and rotation about the C^{α}–C′ bond (*i.e.*, the variation of ψ in the standard convention[12]).

There is abundant evidence for the existence of cis–trans isomerism about the peptide bond in poly(L-proline)[13] and in oligomers of L-proline;[6] cis peptide bonds have also been reported for cyclic compounds such as diketopiperazine,[14] cyclotri(L-prolyl),[15] and cyclo(Pro-Gly)$_3$,[16] for N-substituted polypeptide chains,[17,18] and for some peptide bonds in subtilisin BPN′[19] and ribonuclease S.[20]

Tonelli[21] has computed the fraction of cis and trans states in oligomers of L-proline which had been investigated experimentally by Deber, *et al.*;[6] he carried out conformational energy calculations in which he neglected longer range interactions beyond the second neighboring residue. In this paper, we calculate the fraction of cis and trans states in oligomers of L-proline by taking into account the interactions over the *whole* molecule and also by including the contribution from the librational entropy[22,23] to each minimum-energy conformation. In using the rotational isomeric state model, we do not consider the states of the individual residues (as in the three-state model[8]) but rather the discrete states of the whole molecule (which occur in very narrow regions about the local minima on the energy surface in the case of oligomers of L-proline). Since the conformations of oligomers may be affected by the terminal groups, especially at the C terminus, end effects are also examined in this paper. In comparing our results to those of Deber, *et al.*,[6] it should be kept in mind that solvent effects are not included in our computations.

Calculation Procedure

Geometry of Proline Residue. The pertinent atoms of an ith proline residue are shown in Figure 1. The bond lengths and bond angles are maintained fixed, and the peptide groups (either cis or trans) are kept planar. However, certain bond angles are interchanged between the cis and trans conformations,[4,24] *viz.*, for trans proline (ω_{i-1} = 180°), the bond angles are taken as $\tau(C_{i-1}'N_iC_i^{\alpha})$ = 121° and $\tau(C_{i-1}'N_iC_i^{\delta})$ = 126°, whereas, for cis proline (ω_{i-1} = 0°), these bond angles are interchanged to $\tau(C_{i-1}'N_iC_i^{\alpha})$ = 126° and $\tau(C_{i-1}'N_iC_i^{\delta})$ = 121°. For the C_{i-1}'–N_i peptide bond length, we use 1.36 Å which is somewhat longer than those[5] (1.32 Å) observed in peptides of ordinary amino acids. The larger value has been observed in X-ray crystallographic studies of *p*-bromocarbobenzoxyglycylprolylleucylglycine,[25] *cyclo*-(tetrasarcosyl),[26] and *N,N′*-dimethyldiketopiperazine.[27] All other bond lengths and bond angles are the same as those used in ref 4, 5, and 28. The puckering of the pyrrolidine ring (*i.e.*, the values of ϕ, χ_1, χ_2, χ_3, and χ_4) for the C^{γ} atom in the up (U) and down (D) positions is described by Momany, *et al.*[5] The existence of such puckering at the C^{γ} atom has been indicated by experimental results in the solid state[25,29-31] and in solution.[32] While

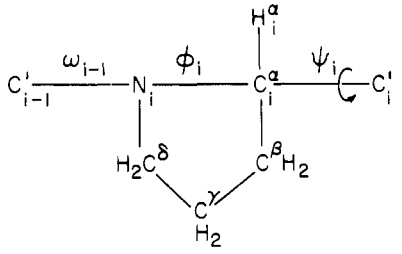

Figure 1. Nomenclature of pertinent atoms in a proline residue.

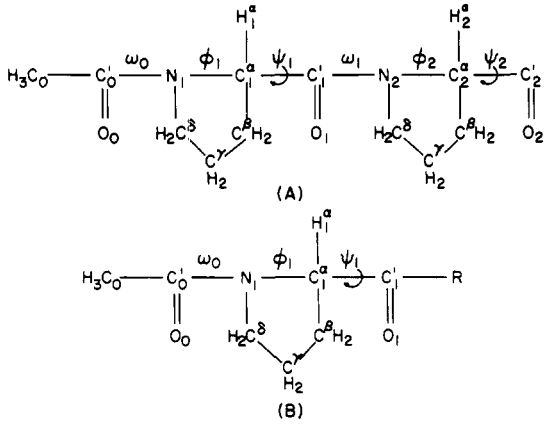

Figure 2. Representation of fragments treated initially. R designates either a methyl ester (OCH_3) or an *N,N'*-dimethylamide [$N(CH_3)_2$].

there are reports of puckering at the C^β atom, such puckering was not considered here[5] because, in our judgment, C^β puckering was deduced either from structures which did not have low reliability indices (*R* factor) or from high-reliability structures which were small and cyclic rather than linear. Restricting the geometry to only C^γ puckering, the designations D and U pertain to the endo ($\phi = -75.0°$) and exo ($\phi = -67.6°$) positions, respectively, of the C^γ atom.

Energy Functions. The energy functions are the recent ones of Momany, *et al.*[5] (described earlier by Lewis, *et al.*,[28] and Burgess, *et al.*[4]). They include nonbonded and electrostatic energy contributions.

Internal Energy. Since several different configurations (i.e., different bond angles for cis and trans) and conformations (U and D) are considered for the fixed part of each proline residue, we must take into account the (internal) energies of each of these configurations and conformations in comparing the relative stabilities of proline oligomers. The internal energies are calculated for the structure shown in Figure 1, with the up or down conformation for the pyrrolidine ring, and for the cis or trans conformation of the C_{i-1}'–N_i peptide bond, and the results are given in Table I.

Energies of Fragments. In order to save computer time in the calculations of the energies of the longer oligomers, two types of fragments were treated initially. These are shown in Figures 2A and 2B. By considering these fragments first, we can later avoid computing the energies of conformations of longer oligomers in which these fragments have very high energies. For these fragments, as well as for the later calculations with longer oligomers, the up–down and cis–trans characters were selected, and then the only independent variables were the ψ's, *i.e.*, the dihedral angles for rotation about C^α–C' bonds. In all computations reported in this paper, the N-terminal group was in the cis conformation, *i.e.*, with C_0–H cis to C_0'–O_0. In the structure of Figure 2B, the conformation of

Table I
The Internal Energy of the Fixed Part of a Proline Residue

Puckering conformation[a]	Peptide bond conformation	Internal energy,[b] kcal/mol
D	Cis	0.000[c]
U	Cis	0.202
D	Trans	0.336
U	Trans	0.550

[a] The designations down (D) and up (U) pertain to the endo ($\phi = -75.0°$) and exo ($\phi = -67.6°$) positions, respectively, of the C^γ atom. [b] The energy pertains to the unit shown in Figure 1. [c] All values are expressed as relative values with respect to the down-cis conformation by subtracting 12.355 kcal/mol from each of the computed energies.

the C-terminal R group was taken as follows. When R was an ester group (O–CH_3), both dihedral angles were taken as 180°, *i.e.*, with C_1^α–C_1' trans to O_{ester}–C_{methyl} and with C'–O_{ester} trans to C_{methyl}–H. When R was an *N,N'*-dimethylamide ($N(CH_3)_2$), the structure was planar, with C_1^α–C_1' cis to N–C_{methyl}. The methyl group at the N terminus in Figures 2A and 2B was included because it influences the energy when ψ_1 is varied (in general, the C_{i-1}^α atom affects the energy when ψ_i is varied). When the energy of the fragment in Figure 2A was computed by allowing ψ_1 and ψ_2 to vary independently of each other, it was found that the minimum-energy value of ψ_1 was independent of the value of ψ_2 for all possible conformations of U and D and cis and trans. Therefore, in all computations reported here (for only the structure of Figure 2A), the condition $\psi_1 = \psi_2$ was imposed. This condition was *not* imposed when the energies of the various oligomers were minimized.

The energies[33] of 16 possible combinations of configuration and conformation of the structure of Figure 2A are plotted as a function of ψ_1 in Figure 3, and those of Figure 2B are plotted in Figure 4 for a methyl ester C-terminal group and in Figure 5 for an *N,N'*-dimethylamide C-terminal group. The contributions of the internal energies of Table I have already been subtracted from the conformational energies plotted in Figures 3–5.

As can be seen in Figure 3, there is only one low-energy value of ψ_1, in a deep energy well with a minimum between 160 and 170°, and this position is not altered significantly by variation in the conformation of the peptide group or in the puckering of the pyrrolidine ring. There is also a minimum (not shown in Figure 3) near $\psi_1 \sim -60°$, but the energies are at least 80 kcal/mol higher than those near 160° for every combination of peptide conformation or pyrrolidine ring puckering. The high energy in the region of ψ_1 between −80 and 0° arises mainly from the strong repulsion between the atoms of the N-terminal pyrrolidine ring and the $C^\delta H_2$ group of the succeeding residue. Thus, when we later consider a fragment such as that of Figure 2A in a longer oligomer, we can neglect values of ψ_1 in the region near −60° and consider only those values in the lower energy region of 160°, since the energy values near $\psi_1 = -60°$ are too high to be compensated by long-range attractive forces in longer oligomers.

On the other hand, two minima (at $\psi_1 \approx -40°$ and 140°) are observed for *N*-acetyl-L-proline methyl ester (Figure 4). The minimum at $\psi_1 \sim -40°$ arises because there is no strong repulsion between the pyrrolidine ring and the C-terminal group, *if the latter is not bulky*. For a bulky C-terminal group, such as the *N,N'*-dimethylamide, however, the minimum at $\psi \approx -40°$ is not observed but only the one at 160° appears (see Figure 5).

Energies of Oligomers. The results of Figures 3–5 per-

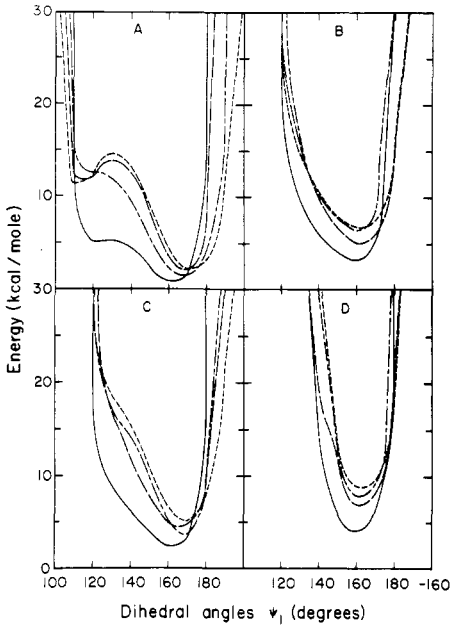

Figure 3. Dependence of conformational energy on ψ_1 ($=\psi_2$) for the structure depicted in Figure 2A. The peptide conformations (ω_0 and ω_1) are (A) trans–trans, (B) trans–cis, (C) cis–trans, and (D) cis–cis. The puckering conformations of the two pyrrolidine rings are (——) down–down, (— - —) down–up, (- - -) up–down, and (· · ·) up–up in each figure. In Figure 3A, the minimum near $\psi_1 \sim 120°$ is no longer a minimum, or is a relatively high-energy minimum of low statistical weight, if a C-terminal group is added to the fragment.

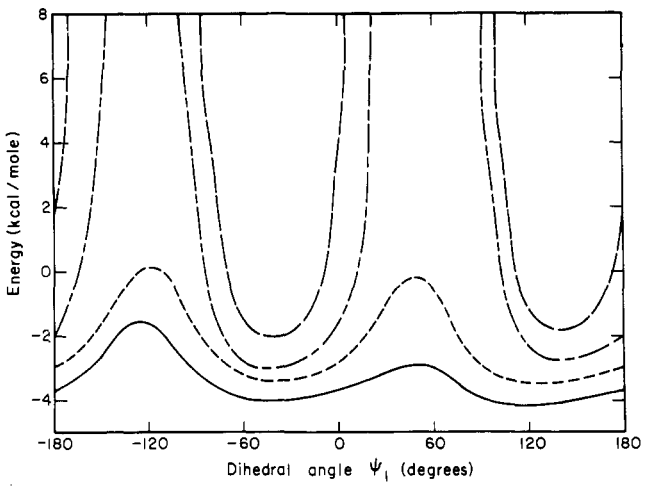

Figure 4. Dependence of conformational energy on ψ_1 for the sturcture depicted in Figure 2B, with R taken as a methyl ester. The peptide (ω_0) and pyrrolidine ring puckering conformations are (——) trans–down; (— · —) cis–down; (· · ·) trans–up; and (- -) cis–up.

mit us to treat a general oligomer, shown in Figure 6, more easily. In order to apply the rotational isomeric state model to such an oligomer, we must obtain all the low-energy minima of such a molecule. For such a minimization process, we select starting conformations of the oligomer as follows. The initial values of ψ_i (with $i = 1$ to $n - 1$, where n is the number of proline residues in the oligomer) are taken as those at the minima of Figure 3. For the initial values of ψ_n, *i.e.*, for the C_n^{α}–C_n' dihedral angle next to the C-terminal R group, we use the conformations of minimum energy in Figure 4 for a methyl ester terminal group and 160° (*i.e.*, the minimum-energy conformation in Figure 5) for an *N,N'*-dimethylamide end group. For any value of n, all possible combinations of these values of ψ_i are selected

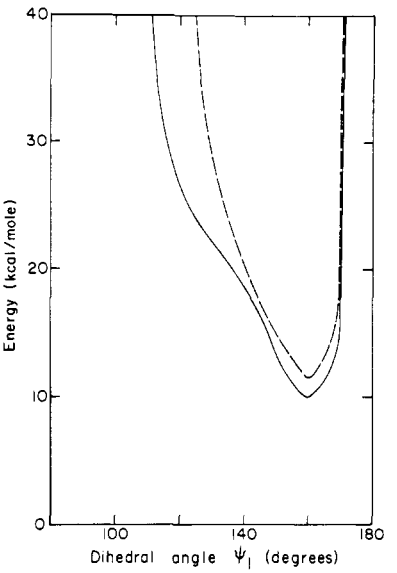

Figure 5. Same as Figure 4, but with R taken as an *N,N'*-dimethylamide: (——) trans–down; (- -) cis–down. Only D puckering was considered for this end group since the results were to be compared with those with the methyl ester end group, for which the all-D conformation was the most stable.

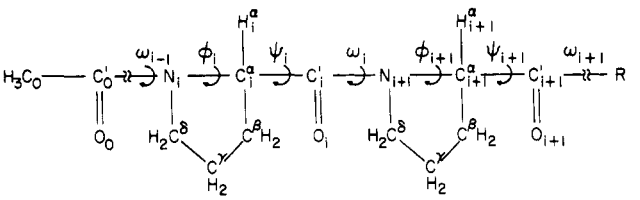

Figure 6. Schematic representation of an oligomer of L-proline. The C-terminal group R is the same as in Figure 2.

as starting conformations for energy minimization. It is reasonable to expect that this selection of starting conformations will lead to all stable conformations in the energy surface, and the rotational isomeric states of the whole molecule may be assumed to consist of the ensemble of these minimum-energy conformations.

The procedure used to minimize the conformational energy of the whole molecule with respect to all ψ_i's is the conjugate gradient method.[34,35] In all cases, the magnitudes of all first derivatives at the minimum were less than 10^{-4} (kcal mol^{-1})/radian.

Calculation of Librational Entropy and Statistical Weights. The relative statistical weights of conformations with *small* conformational fluctuations about the minimum can be computed[3,22,23,36] by

$$W = \exp[-E(q_0)/RT][(2\pi RT)^m/\det \mathbf{F}(q_0)]^{1/2} \quad (1)$$

where $E(q_0)$ is the value of the conformational energy at the minimum q_0 which designates the set of ψ_i's at the minimum, m is the number of variable dihedral angles in the molecule ($=n$ in this case), and the elements of the matrix $\mathbf{F}(q_0)$ are the second derivatives of the conformational energy at a minimum, *i.e.*, at q_0, *viz.*, $\partial^2 E/\partial\psi_i\partial\psi_j$. In using eq 1, it is assumed[3,22,23,36] that the kinetic energy of the molecule is independent of conformation and hence that det **F** may be regarded as the librational entropy.

In computing $\partial^2 E/\partial\psi_i\partial\psi_j$, we have used the method of Go, *et al.*,[37] which involves the differentiation of the energy function and of the bond transformation matrix. In order to make certain that the conformation under consideration is at an energy minimum, we also calculated the first derivative of the conformational energy using the same mathematical method.

Table II
Rotational Isomeric States, Conformational Energy, and Statistical Weights of Di-L-proline[a]

Puckering conformation[b]	Peptide bond conformation		Starting conformation,[c] deg		Min energy conformation, deg		Conformational[d] energy, ΔE, kcal/mol	Statistical weight[e]	
	ω_0	ω_1	ψ_1	ψ_2	ψ_1	ψ_2		From E	From $(E + L)$[f]
DD	t	t	160	120	164.46	137.54	0.0	1.0	0.453
			160	−40	164.45	−38.16	0.186	0.731	0.193
	t	c	160	140	160.54	133.34	1.07	0.165	0.022
			160	−50	159.66	−37.54	0.206	0.707	0.155
	c	t	160	120	164.25	134.56	1.34	0.104	0.049
			160	−40	164.27	−37.50	1.52	0.076	0.020
	c	c	160	140	160.67	142.43	1.96	0.037	0.006
			160	−50	161.11	−47.37	1.58	0.069	0.007
DU	t	t	170	130	169.80	137.49	1.17	0.139	0.038
			170	−40	169.81	−34.87	1.34	0.104	0.019
	t	c	160	140	163.83	137.53	3.43	0.003	0.0
			160	−40	163.01	−36.17	2.35	0.019	0.002
	c	t	170	130	169.53	135.44	2.58	0.013	0.004
			170	−40	169.55	−34.29	2.72	0.010	0.002
	c	c	160	140	163.88	142.19	4.21	0.001	0.0
			160	−40	165.59	−47.49	4.27	0.001	0.0
UD	t	t	170	120	167.89	136.40	0.337	0.566	0.317
			170	−40	167.79	−38.17	0.542	0.400	0.131
	t	c	160	140	160.89	133.66	1.98	0.035	0.005
			160	−50	159.74	−39.25	1.08	0.161	0.037
	c	t	160	120	164.54	133.14	2.56	0.013	0.007
			160	−40	164.57	−37.53	2.76	0.010	0.003
	c	c	160	140	159.33	140.55	3.36	0.003	0.001
			160	−50	160.18	−47.54	3.02	0.006	0.001
UU	t	t	170	130	173.59	137.27	1.12	0.151	0.051
			170	−40	173.57	−35.14	1.31	0.110	0.024
	t	c	160	140	166.11	137.55	4.05	0.001	0.0
			160	−40	165.05	−37.55	3.04	0.006	0.001
	c	t	170	130	169.72	134.36	3.97	0.001	0.0
			170	−40	169.74	−34.91	4.13	0.001	0.0
	c	c	160	140	163.75	141.09	5.58	0.0	0.0
			160	−40	165.73	−44.90	5.49	0.0	0.0

[a] With C-terminal methyl ester group and fixed values of ω's. [b] See footnote *a* of Table I. [c] The starting conformations were the minimum-energy values of ψ_1 of Figure 3 for ψ_1 and the minimum-energy values of ψ_1 of Figure 4 for ψ_2. [d] The reported energies are relative values, ΔE, with the lowest-energy conformer, *i.e.*, DD-tt, assigned an energy of 0.0. [e] Statistical weights are calculated by two methods, *i.e.*, by using only E(*i.e.*, by neglecting det **F**) and by taking the contribution from both the conformational energy and the librational entropy $(E + L)$ into account. [f] In the form of $\exp[-\Delta E/RT][(2\pi RT)^m/\det \mathbf{F}]^{1/2}$.

Calculation of Fraction of Conformers. The statistical weights of each of the conformers (minimum-energy conformations) of di-, tri-, tetra-, and penta-L-proline are calculated with eq 1. For di-L-proline, there are 32 conformers arising from all possible combinations of cis (c) and trans (t) peptides and up (U) and down (D) puckering of the pyrrolidine ring (*e.g.*, tt-DD, tt-DU, tt-UD, tt-UU, tc-DD, etc., or 16 in total) and, in addition, from two minimum-energy values of ψ_n when the C-terminal group is a methyl ester; when the C-terminal group is an *N,N'*-dimethylamide, there are only four conformers for di-L-proline (with only D puckering). Similarly, for tri-L-proline, there are 128 (*i.e.*, $4 \times 4 \times 4 \times 2$) conformers with a methyl ester C-terminal group. For oligomers larger than tri-L-proline, an assumption is introduced to reduce the otherwise excessive computer time that would be required, *viz.*, only the D conformation is taken for the pyrrolidine ring. This assumption can be justified by a number of observations. According to recent calculations of the conformational energy of poly(L-proline),[38] the chains with *regular* puckering of the pyrrolidine rings (*i.e.*, all up or all down) have lower energy than those with randomly distributed puckering conformations because of an unfavorable conflict between the pyrrolidine rings. Furthermore, as the chain becomes longer, the chain with the all-down puckering conformation becomes more stable than the one with the all-up puckering conformation.[38] In fact, our theoretical results (Table III) for tri-L-proline indicate that (for a given combination of cis and trans peptides) UUU is ~1.5 kcal/mol higher in energy than DDD. Also, as can be seen in Figure 3, DD combinations have the lowest energy (by at least 1 kcal/mol) for all combinations of cis and trans peptides. Because of this small energy difference, we cannot neglect the existence of U puckering in *short chains;* indeed (from Table V) it can be seen that the DD and DDD conformations are the dominant ones for di- and tri-L-proline, respectively, but that reasonable amounts of U occur in these two oligomers because of their short chain length. Finally, this assumption was checked by observing (Table VI) that the computed fraction of any conformer (tt, tc, . . . , ttt, ttc, . . . , etc.) was essentially the same in di- and tri-L-proline, respectively, when both U and D conformations were included and also when only D conformations were taken. On the basis of proton nmr data, Torchia[32] proposed that the pyrrolidine rings of poly(L-proline) rapidly interconvert between two conformations puckered at the C^γ atom. However, there are ambiguities in the interpretation of vicinal couplings. First, to justify the existence of two puckered conformations, Torchia cited the results of conformational energy calculations performed for an *isolated* proline residue, which may not be applicable to a polymer *chain.* Second, he did not examine the various possible sets of values of ϕ and χ_i (i = 1–4) but fixed ϕ at −60° and varied only the values of χ_i. Third, the rapid rocking motions of the methylene groups of the pyrrolidine ring (accompanied by the relatively slower motion of the ring around a single equilibrium puckered conformation) were not taken into consideration in his interpretation of the vicinal *proton*

Table III
Energetically Favorable Confomers,[a] Conformational Energy, and Statistical Weights of Tri-L-proline[b]

Puckering conformation[c]	Peptide bond conformation			Min. energy conformation,[d] deg			Conformational energy, ΔE,[e] kcal/mol	Statistical weight[f]	
	ω_0	ω_1	ω_2	ψ_1	ψ_2	ψ_3		From E	From $(E + L)$[g] (×10)
DDD	t	t	t	164.45	164.52	139.61	0.0	1.0	0.813
				164.45	164.49	−39.65	0.143	0.785	0.363
	t	t	c	164.46	160.20	132.74	0.722	0.296	0.071
				164.56	159.44	−32.82	0.070	0.888	0.414
	t	c	t	161.18	163.89	136.61	1.44	0.088	0.075
				161.23	163.96	−33.02	1.76	0.051	0.29
	c	t	t	164.22	164.51	141.90	1.36	0.100	0.076
				164.20	164.47	−39.61	1.56	0.072	0.034
	c	t	c	164.28	159.25	−35.70	1.34	0.104	0.046
	c	c	t	161.89	163.65	137.27	1.89	0.041	0.030
DDU	t	t	t	164.46	169.85	139.04	1.15	0.142	0.070
				164.44	169.83	−35.73	1.29	0.114	0.036
DUD	t	t	t	169.80	167.92	138.85	0.860	0.234	0.188
				169.79	167.84	−39.71	1.03	0.176	0.081
	t	t	c	169.89	159.28	−34.52	1.42	0.090	0.036
UDD	t	t	t	167.87	164.55	139.85	0.347	0.557	0.557
				167.80	164.48	−39.57	0.493	0.435	0.251
	t	t	c	167.84	160.06	132.48	1.02	0.179	0.528
				168.04	159.33	−31.40	0.361	0.544	0.326
DUU	t	t	t	169.78	173.67	138.29	1.63	0.064	0.032
				169.79	173.58	−36.20	1.79	0.048	0.015
UDD	t	t	t	167.82	169.84	138.97	1.50	0.079	0.049
				167.81	169.82	−36.05	1.64	0.063	0.025
UUD	t	t	t	173.56	167.87	139.28	0.829	0.247	0.249
				173.58	167.80	−39.65	1.00	0.184	0.107
	t	t	c	173.70	159.01	−33.05	1.21	0.129	0.069
UUU	t	t	t	173.54	173.65	138.48	1.60	0.067	0.042
				173.53	173.56	−36.17	1.77	0.050	0.020

[a] Only those conformers with energy less than 2 kcal/mol are listed. There are 17 other conformers in the energy range between 2 and 3 kcal/mol and 21 conformers between 3 and 4 kcal/mol. The total number of conformers under consideration is 128. [b] With C-terminal methyl ester group and fixed values of ω's. [c] See footnote *a* of Table I. [d] The starting conformations were the minimum-energy values of ψ_1 of Figure 3 for ψ_1 aud ψ_2 and the minimum-energy values of ψ_1 of Figure 4 for ψ_3. [e] The reported energies are relative values, ΔE, with the lowest-energy conformer, *i.e.*, DDD-ttt, assigned an energy of 0.0. [f] See footnote *e* of Table II. [g] See footnote *f* of Table II.

couplings. For these reasons, it does not appear to us that the nmr data rule out the existence of a single puckered conformation in a long poly(L-proline) chain. With the assumption that only D conformations can occur in oligomers longer than tri-L-proline, we have 32 (*i.e.*, $2^4 \times 2$) and 64 (*i.e.*, $2^5 \times 2$) conformers for tetra- and penta-L-proline, respectively, with methyl ester C-terminal groups.

In the rotational isomeric state model, the partition function Z for an n-mer is given as a sum over the statistical weights W of all conformers, *i.e.*

$$Z(n) = \sum_{\alpha} \sum_{\beta} W(n, \alpha, \beta) \tag{2}$$

where α and β are summation indices for the conformational states of the pyrrolidine ring and peptide group, respectively (α appearing only for di- and tri-L-proline). The fractions of conformers f_α and f_β are then readily calculated from the partition function; *e.g.*, f_β (where, say for di-L-proline, β refers to tt, tc, etc.) is given by

$$f_\beta(n) = Z(n)^{-1} \sum_{\alpha} W(n, \alpha, \beta) \tag{3}$$

The probability, $P_{i;t}$, of finding the ith peptide bond in the trans conformation may also be calculated from the partition function as

$$P_{i;t}(n) = Z(n)^{-1} \sum_{\alpha} \sum_{\beta=\beta_{i=t}} W(n, \alpha, \beta) \tag{4}$$

where $\Sigma_{\beta=\beta_{i=t}}$ means that the sum is taken over the statistical weights of only those conformers in which the ith peptide bond is trans. In order to investigate the role of the librational entropy, we compute the statistical weights W in two ways, both with and without inclusion of the term det **F** of eq 1. All values of fractions of conformers and probabilities are computed at a temperature of 25° for later comparison with nmr data[6] in chloroform at this temperature.

Results and Discussion

The rotational isomeric states, conformational energies, and statistical weights of the local minimum-energy conformers of di-, tri-, tetra-, and penta-L-proline, with N-terminal acetyl and C-terminal methyl ester groups, are given in Tables II–IV. For di-L-proline (Table II), the results for all 32 conformers under consideration here are listed. For the remaining oligomers, only the energetically favorable conformers [with relative energies of less than 2.0 kcal/mol (Table III) or less than 1.5 kcal/mol (Table IV)] and some high-energy conformers are given for illustrative purposes.

The fractions of conformers with respect to pyrrolidine ring puckering, f_α, in di- and tri-L-proline were calculated by an analog of eq 3 and are given in Table V. From the calculations, it appears that the D conformations are more preferred than the U ones in these two oligomers. No experimental results, with which these calculations can be compared, are available at the present time.

The calculated values of the fractions of the dyad and triad conformers of di- and tri-L-proline, respectively, are given in Table VI. While experimental results are not available for these oligomers, having the end groups used in our calculations, Deber, *et al.*,[6] obtained these quantities experimentally for oligomers of L-proline with an N-terminal *tert*-butyl group and a C-terminal benzyl group; these experimental values are given in the last column of Table VI.

Table IV
Energetically Favorable Conformers,[a] Conformational Energy, and Statistical Weights of Tetra- and Penta-L-proline[b]

Oligomer	Peptide bond conformation					Min. energy conformation,[c] deg					Conformational energy, ΔE,[d] kcal/mol	Statistical weight[e]	
	ω_0	ω_1	ω_2	ω_3	ω_4	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5		From E	From $(E + L)$[f] ($\times 10^2$)
Tetra-L-proline	t	t	t	t		164.54	164.60	164.59	138.28		0.0	1.0	1.45
						164.49	164.53	164.52	−39.48		0.136	0.794	0.644
	t	t	t	c		164.52	164.57	160.02	132.97		0.666	0.325	0.141
						164.49	164.66	159.02	−31.38		0.026	0.957	0.854
	t	t	c	t		164.37	160.85	163.82	138.21		1.18	0.136	0.210
	t	c	t	t		161.09	163.81	164.40	143.34		1.41	0.092	0.139
	c	t	t	t		164.24	164.53	164.60	140.34		1.34	0.104	0.148
						164.25	164.52	164.53	−40.21		1.47	0.084	0.068
	t	c	t	c		161.31	163.21	159.79	129.51		1.12	0.152	0.041
						160.82	164.32	158.85	−24.73		0.327	0.576	0.338
	c	t	t	c		164.26	164.65	159.10	−31.57		1.36	0.101	0.090
	c	t	t	c		162.14	163.20	159.95	−46.51		1.21	0.131	0.033
Penta-L-proline	t	t	t	t	t	164.47	164.55	164.54	164.52	139.14	0.529	0.409	0.103
						164.48	164.50	164.48	164.52	−39.48	0.683	0.316	0.046
	t	t	t	t	c	164.46	164.59	164.58	160.02	132.94	1.24	0.123	0.009
						164.46	164.59	164.65	159.01	−32.11	0.546	0.397	0.061
	t	t	c	t	c	164.74	162.43	164.47	160.86	133.19	0.0	1.0	0.020
						164.47	158.77	164.47	159.66	−21.60	0.074	0.883	0.091
	t	c	t	t	c	161.08	163.91	164.61	159.11	−30.18	0.925	0.210	0.039
	c	t	c	t	c	163.02	160.94	162.89	159.81	148.04	1.49	0.081	0.002
						164.82	157.63	165.02	158.98	−34.11	1.22	0.128	0.015
	c	c	c	c	c	154.78	162.07	166.65	168.23	154.09	11.4[g]	0.0	0.0
						154.72	162.07	166.89	169.02	−74.99	13.5[g]	0.0	0.0

[a] Only those conformers with energy less than 1.5 kcal/mol are listed. There are ten conformers of tetra-L-proline and 23 conformers of penta-L-proline in the energy range between 1.5 and 3.0 kcal/mol. [b] With C-terminal methyl ester group and fixed values of ω's. [c] The starting conformations were the minimum-energy values of ψ_1 of Figure 3 for ψ_1 to ψ_{n-1} and the minimum-energy values of ψ_1 of Figure 4 for ψ_n. [d] The reported energies are relative values, ΔE, with the lowest energy conformers, *i.e.*, tttt for tetra-L-proline and ttctc for penta-L-proline, assigned an energy of 0.0. [e] See footnote *e* of Table II. [f] See footnote *f* of Table II. [g] These high-energy conformers are given only for illustrative purposes.

As shown experimentally by Madison and Schellman,[39] the fractions of conformers in oligomers of proline are very sensitive to the bulkiness of the C-terminal group (and not so sensitive to the bulkiness of the N-terminal group). These authors[39] also pointed out that the solvent plays a less important role in determining the conformations of oligomers of derivatives of L-proline when the C-terminal group is bulky. Therefore, to examine the effect of the bulkier C-terminal group, the calculations were repeated by replacing the methyl ester (ME) by an *N,N'*-dimethylamide (DMA) end group, and the results are also presented in Table VI. It can be seen that the calculated values are indeed markedly affected by the bulkiness of the C-terminal group. In the case of di-L-proline, ct becomes more favorable than tc, and in the case of tri-L-proline, ctt becomes more favorable than ttc, when DMA is substituted for the ME end group, and these results are in accord with the qualitative trend of the experimental results.[21] For di- and tri-L-proline, with an ME end group, the result that f_{tc} and f_{ttc} are larger than f_{ct} and f_{ctt}, respectively, is due mainly to the low energies of the DD-tc conformer (ψ_1 = 159.7° and ψ_2 = −37.5°; Table II) and the DDD-ttc conformer (ψ_1 = 164.6°, ψ_2 = 159.4°, and ψ_3 = −32.8°; Table III); these values of the terminal ψ (~−35°) can appear only at the end of the chain. Since $\psi_n \sim -35°$ is a high-energy conformation for a DMA end group, f_{ct} and f_{ctt} become larger than f_{tc} and f_{ttc} when DMA is substituted for ME. Thus, these results are not in conflict with the indication[40] from nmr studies that it is preferable for a trans sequence to follow a cis sequence rather than *vice versa* in the cis–trans isomerization of poly(L-proline). Rifkind and Applequist[41] also pointed out this fact, based on the inspection of space-filling molecular models. While the discrepancies between theory and experiment in Table VI are thus probably due to differences in

Table V
The Fractions of Conformers with Respect to Pyrrolidine Ring Puckering in Di- and Tri-L-proline

Oligomer	Conformation of pyrrolidine ring[a]	Fraction of conformer	
		From E[b]	From $(E + L)$[c]
Di-L-proline	DD	0.62	0.58
	DU	0.06	0.04
	UD	0.26	0.32
	UU	0.06	0.05
Tri-L-proline	DDD	0.48	0.35
	DDU	0.05	0.03
	DUD	0.08	0.20
	UDD	0.25	0.31
	DUU	0.02	0.01
	UDU	0.02	0.02
	UUD	0.08	0.08
	UUU	0.02	0.01

[a] See footnote *a* of Table I. [b] Values calculated from the contribution of only the conformational energy to the statistical weight. [c] Values calculated from the contribution of both the conformational energy and the librational entropy to the statistical weight.

C-terminal end groups, we place more emphasis in this paper on the conclusions about the conformational states of a proline residue in the *interior* of a longer oligomer than di- and tri-L-proline.

The calculated probabilities of finding the ith peptide group in the trans conformation, $P_{i;t}$, are presented in Table VII. The values of $P_{i;t}$ for residues in the middle of the chain are less affected by the values of the dihedral angles at both termini (ω_0 and ψ_n) and by the bulkiness of the terminal groups than are the dyad and triad fractions of

Table VI
The Dyad and Triad Fractions in Conformers of Di- and Tri-L-Proline

		Fraction of conformer						
		Calculated value						
		Methyl ester end group				*N,N'*-Dimethylamide end group		
Oligomer	Conformer	From $E^{a,b}$	From $(E + L)^{b,c}$	From $E^{a,d}$	From $(E + L)^{c,d}$	From $E^{a,d}$	From $(E + L)^{c,d}$	Exptl value[e]
Di-L-proline	tt	0.69	0.79	0.60	0.71	0.86	0.86	0.40
	tc	0.24	0.14	0.30	0.20	0.05	0.05	0.20
	ct	0.05	0.06	0.06	0.08	0.08	0.08	0.40
	cc	0.03	0.01	0.04	0.01	0.02	0.02	0.0
Tri-L-proline	ttt	0.59	0.66	0.51	0.60	0.75	0.74	0.30
	ttc	0.31	0.26	0.34	0.25	0.06	0.07	0.10
	tct	0.03	0.03	0.04	0.05	0.07	0.08	0.16
	ctt	0.04	0.03	0.05	0.06	0.08	0.08	0.30
	tcc	0.0	0.0	0.01	0.0	0.0	0.0	0.04
	ctc	0.02	0.01	0.04	0.03	0.01	0.01	0.10
	cct	0.01	0.01	0.02	0.02	0.03	0.03	0.0
	ccc	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[a] See footnote *b* of Table V. [b] Including both D and U puckering of the pyrrolidine ring. [c] See footnote *c* of Table V. [d] Only D puckering of the pyrrolidine ring was allowed (see legend of Figure 5). [e] From Table II of ref 21.

Table VII
The Probabilities of Finding the *i*th Peptide Group in the Trans Conformation for Oligomers of L-proline with a C-Terminal Methyl Ester Group

		Calculated value		
Oligomer	$P_{i;t}$	From E^a	From $(E + L)^b$	Exptl value[c]
Di-L-proline[d]	$P_{0;t}$	0.93	0.94	0.60
	$P_{1;t}$	0.74	0.85	0.80
Tri-L-proline[d]	$P_{0;t}$	0.68	0.79	0.60
	$P_{1;t}$	0.83	0.82	0.80
	$P_{2;t}$	0.66	0.78	0.76
Tetra-L-proline[e]	$P_{0;t}$	0.88	0.88	0.60
	$P_{1;t}$	0.76	0.84	0.80[f]
	$P_{2;t}$	0.95	0.93	0.80
	$P_{3;t}$	0.52	0.66	0.80
Penta-L-proline[e]	$P_{0;t}$	0.89	0.88	0.60
	$P_{1;t}$	p.89	0.83	0.99[f]
	$P_{2;t}$	0.47	0.68	0.99[f]
	$P_{3;t}$	0.94	0.93	0.99
	$P_{4;t}$	0.31	0.51	0.99

[a] See footnote *b* of Table V. [b] See footnote *c* of Table V. [c] From Table III of ref 21. [d] Including both D and U puckering of the pyrrolidine ring. [e] Only D puckering of the pyrrolidine ring was allowed. [f] These values were assumed because of an experimental ambiguity in ref 6 and 21.

Table VI. Thus, the calculated values for peptide groups in the middle of the chain are in better agreement (except for $P_{2;t}$ in penta-L-proline) with the experimental values than are the values for the terminal peptide groups; in some cases, the agreement is improved by the inclusion of the contribution from the librational entropy. The low value of $P_{2;t}$ in penta-L-proline arises because of the energetic preference for ttctc (resulting mainly from the difference in internal energy of the cis and trans forms, shown in Table I) in this all-D oligomer; however, there is no point in discussing the disagreement between the experimental and calculated values of $P_{2;t}$ because the experimental value of 0.99 was assumed because of an experimental ambiguity.[21] In the calculation of Tonelli,[21] the values of $P_{i;t}$ for the interior of the chain had already reached 0.99 at the stage of the tripeptide, while those found by experiment (Table VII) did not reach 0.99 until the pentapeptide. However, in the present calculations, the cis conformation can occur in the middle of the chain, for a chain as long as the tetrapeptide, which is in ageement with the experimental results, although our calculated values did not reach 0.99 even for the pentapeptide.

The experimentally observed[6] onset of the all-trans helical conformation (form II) at the pentapeptide stage cannot be accounted for either by the calculations of Tonelli[21] or by our calculations. However, we have found that the conformation of the (dominant) all-trans oligomer is regular (i.e., that the ψ_i's are similar), and this may contribute to an understanding of the experimental fact mentioned above. From Tables II, III, and IV, we see that the all-trans conformer is the lowest energy one for di- to tetra-L-proline, and, from Table IV, we see that the ttctc conformer is the lowest energy one for penta-L-proline. However, the ttttt conformer becomes the most stable one when the contribution from the librational entropy is taken into account (see last column of Table IV). Furthermore, the ttttt conformer has a *regular structure* as far as the ψ_i's are concerned (*i.e.*, $\psi_i = 164.5°$ for $i = 1$–4), which resulted when the ψ_i's were varied *independently* in the energy minimization. This structure is essentially 1⅔ turns of a left-handed helix which presumably is a long enough section of regular helix to give rise to the negative Cotton effect which is observed in the optical rotatory dispersion curve that is characteristic of the form II helix of poly(L-proline) (see Figure 7 of ref 6). At the same time, as can be seen in Table IV, the tttt conformation of tetra-L-proline has similar values of ψ_i (for $i = 1$–3); this is essentially 1⅓ turns of helix. While it is difficult to decide whether this size of regular helix in tetra-L-proline is large enough to show the characteristic Cotton effect of poly(L-proline), we cite the experimental results of Okabayashi, *et al.*,[42] who observed the appearance of the poly(L-proline) helix properties (optical rotatory dispersion and ultraviolet absorption spectra) at the tetrapeptide stage and their completeness in the pentapeptide. Aside from the question as to whether the overall shape of the molecule is helical or not, the values of $\psi_i \simeq 164.5°$ obtained for the tetra- and pentapeptide are also found for the di- and tripeptides (see ψ_1 for tt in Table II and ψ_1 and ψ_2 for ttt in Table III); this regularity found in short oligopeptides is in agreement with X-ray results of Kartha and his colleagues[43] on *t*-BOC-(L-pro)$_3$ (and also on the tetrapeptide), in which the values of ψ in the tri- and tetrapeptides were not very far from those of poly(L-proline) II. The presence of ttctc as a low-energy conformer

suggests that the nucleation of the trans ⇄ cis conversion may occur not only at the terminus but also in the middle of the chain and also may account for the random occurrence of cis residues within the poly(L-proline) chain.[44] It is of interest to examine the relative difficulty of nucleation of the cis conformation at or near the terminus and within the chain. This question has to be considered in terms of the cooperativity of the transition, taking into account the conformations of the neighboring groups at the termini or within the chain. This point will be discussed (using parameters for nucleation both at the termini and within the chain) in a forthcoming paper.[38]

In the present paper, we assumed that the molecules exist as a Boltzmann distribution of conformations about the minimum-energy one. In the model used here, long-range interactions were included; these seem to contribute to the relative stabilities of the various conformations which differ relatively little in energy and to the tendency toward the formation of helix-like structures. On the other hand, no account was taken of the entropy arising from libration around the N-C^{α} bond; in addition, we have not satisfactorily eliminated the end effect in the very short oligomers, as indicated by the discrepancies between the calculated and experimental results. In order to examine these two points, we are extending our investigation to calculate the characteristic ratio of poly(L-proline) as a useful probe of the conformational energy surface of a proline residue.[45] The point to emphasize is that the calculations presented here are in good agreement with the experimental results for the conformational behavior of proline residues *within* the polypeptide chain.

Acknowledgment. Helpful discussions with Drs. A. W. Burgess, F. A. Momany, and P. K. Ponnuswamy are gratefully acknowledged. We are also indebted to Mrs. Shirley Rumsey for assistance with the computer programming.

Note Added in Proof. Our choice of values of ϕ for U and D puckering agrees reasonably well with recently reported computations of Venkatachalam, *et al.*[46] Our values of −67.6 and −75.0° for ϕ, for U and D puckering, respectively, fall within the lowest energy regions of Figures 6a and 6d of ref 46 for trans and cis proline, respectively, even when ψ is allowed to vary over the accessible range around 160° (see Figure 3). For this comparison, our values of ϕ convert to values of their θ of −7.6° and −15.0°, using their relation: $\phi \approx \theta - 60°$.

Further, the stable trans conformation found here is the lower *energy* one up to, and including, the hexamer.[38] For the heptamer and longer chains, the cis conformation has the lower energy[38] under vacuum. Since the cis form is more compact, the long-range nonbonded attractive interactions stabilize the cis form as the chain length increases.[38]

References and Notes

(1) This work was supported by research grants from the National Science Foundation (GB-28469X3), from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312), and from Walter and George Todd.
(2) From Kyoto University, 1972–1974.
(3) For a recent review, see H. A. Scheraga, *Chem. Rev.*, **71**, 195 (1971).
(4) A. W. Burgess, F. A. Momany, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **70**, 1456 (1973).
(5) F. A. Momany, R. F. McGuire, and H. A. Scheraga, *J. Phys. Chem.*, to be submitted for publication.
(6) C. M. Deber, F. A. Bovey, J. P. Carver, and E. R. Blout, *J. Amer. Chem. Soc.*, **92**, 6191 (1970).
(7) D. A. Brant and P. J. Flory, *J. Amer. Chem. Soc.*, **87**, 2791 (1965).
(8) P. J. Flory, "Statistical Mechanics of Chain Molecules," Interscience, New York, N. Y., 1969.
(9) S. Tanaka and A. Nakajima, *Polym. J.*, **2**, 717 (1971).
(10) P. J. Flory and P. R. Schimmel, *J. Amer. Chem. Soc.*, **89**, 6807 (1967).
(11) R. T. Ingwall and P. J. Flory, *Biopolymers*, **11**, 1527 (1972); **12**, 1123, 1137 (1973).
(12) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).
(13) For review of the conformational properties of proline polymers, see (a) E. Katchalski, A. Berger, and J. Kurtz in "Aspects of Protein Structure," G. N. Ramachandran, Ed., Academic Press, New York, N. Y., 1963, p 205; (b) J. P. Carver and E. R. Blout in "Treatise on Collagen," Vol. 1, G. N. Ramachandran, Ed., Academic Press, New York, N. Y., 1967, p 441; (c) L. Mandelkern in "Biological Macromolecules," Vol. 1, G. D. Fasman, Ed., Marcel Dekker, New York, N. Y., 1967, p 675.
(14) See, for example, I. L. Karle, *J. Amer. Chem. Soc.*, **94**, 81 (1972).
(15) C. M. Deber, D. A. Torchia, and E. R. Blout, *J. Amer. Chem. Soc.*, **93**, 4893 (1971).
(16) C. M. Deber, D. A. Torchia, S. C. K. Wong, and E. R. Blout, *Proc. Nat. Acad. Sci. U. S.*, **69**, 1825 (1972).
(17) F. A. Bovey, J. J. Ryan, and F. P. Hood, *Macromolecules*, **1**, 305 (1968).
(18) J. H. Howard, F. A. Momany, R. H. Andreatta, and H. A. Scheraga, *Macromolecules*, **6**, 535 (1973).
(19) C. S. Wright, R. A. Alden, and J. Kraut, *Nature (London)*, **221**, 235 (1969).
(20) H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagami, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, **242**, 3984 (1967).
(21) A. E. Tonelli, *J. Amer. Chem. Soc.*, **92**, 6187 (1970).
(22) N. Go and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969).
(23) N. Go, P. N. Lewis, and H. A. Scheraga, *Macromolecules*, **3**, 628 (1970).
(24) G. N. Ramachandran and V. Sasisekharan, *Advan. Protein Chem.*, **23**, 283 (1968); G. N. Ramachandran and C. M. Venkatachalam, *Biopolymers*, **6**, 1255 (1968).
(25) T. Ueki, T. Ashida, M. Kakudo, Y. Sasada, and Y. Katsube, *Acta Crystallogr., Sect. B*, **25**, 1840 (1969).
(26) P. Groth, *Acta Chem. Scand.*, **24**, 780 (1970).
(27) P. Groth, *Acta Chem. Scand.*, **23**, 3155 (1969).
(28) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Isr. J. Chem.*, **11**, 121 (1973).
(29) A. McL. Mathieson and H. K. Welsh, *Acta Crystallogr.*, **5**, 599 (1952).
(30) Y. C. Leung and R. E. Marsh, *Acta Crystallogr.*, **11**, 17 (1958).
(31) A. Rich and F. H. C. Crick, *J. Mol. Biol.*, **3**, 483 (1961).
(32) D. A. Torchia, *Macromolecules*, **4**, 440 (1971).
(33) The fragment of Figure 2A contains only a part (*i.e.*, C=O) of a peptide group at the C terminus and therefore an unbalanced electrostatic charge. However, the effect of this unbalanced charge on the minimum-energy value of ψ_1 is small, as can be seen by comparing the minimum-energy value of ψ_1 for pro–pro (with a methyl ester end group) (164.46° for down–down and trans–trans; see Table II) with that for the corresponding combination in the fragment of Figure 2A (~160°; see Figure 3A).
(34) M. J. D. Powell, *Comput. J.*, **7**, 155 (1964).
(35) W. I. Zangwill, *Comput. J.*, **10**, 293 (1967).
(36) N. Go, M. Go, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **59**, 1030 (1968).
(37) M. Go, N. Go, and H. A. Scheraga, *J. Chem. Phys.*, **52**, 2060 (1970).
(38) S. Tanaka and H. A. Scheraga, manuscript in preparation.
(39) V. Madison and J. Schellman, *Biopolymers*, **9**, 511 (1970).
(40) D. A. Torchia and F. A. Bovey, *Macromolecules*, **4**, 246 (1971).
(41) J. M. Rifkind and J. Applequist, *J. Amer. Chem. Soc.*, **90**, 3650 (1968).
(42) H. Okabayashi, T. Isemura, and S. Sakakibara, *Biopolymers*, **6**, 323 (1968).
(43) Private communication from Dr. G. Kartha. In crystals of cyclotriproline, he and his colleagues also observed the longer peptide bond length of 1.36 Å, which has been employed in this paper. These results recently appeared in G. Kartha, G. Ambady, and P. V. Shankar, *Nature (London)*, **247**, 204 (1974); and in G. Kartha, T. Ashida, and M. Kakudo, *Acta Crystallogr., Sect. B*, **30**, 1861 (1974).
(44) D. E. Dorman, D. A. Torchia, and F. A. Bovey, *Macromolecules*, **6**, 80 (1973).
(45) S. Tanaka and H. A. Scheraga, work in progress.
(46) C. M. Venkatachalam, B. J. Price, and S. Krimm, *Macromolecules*, **7**, 212 (1974).