CONFORMATIONAL ENERGY AND CONFIGURATIONAL STATISTICS OF POLY-L-PROLINE

By Paul R. Schimmel and Paul J. Flory

DEPARTMENT OF CHEMISTRY, STANFORD UNIVERSITY

Two distinctly different forms of poly-L-proline designated I and II, respectively, have been investigated in the solid state[1-3] and in solution.[4, 5] In form I the imide group is in the *cis* configuration;[3] in form II it is *trans*.[1, 2] Hydrogen bonds of the kind usually occurring in polypeptides are precluded by the absence of an amido hydrogen. Hence, the helical structures occurring in the solid state[1-3] must be dictated by other factors, predominantly steric in origin.

The conformational energies of internal alanyl and glycyl residues were estimated previously by use of semiempirical potential functions for intrinsic bond torsional potentials, for interactions between nonbonded atoms, and for dipole-dipole interactions.[6, 7] Energies were calculated for various values of the angles of rotation φ and ψ about the N—C^α and C^α—C bonds, respectively. The partition function, average chain dimensions, and other average quantities were obtained by assigning a Boltzmann factor, $e^{-E_i/RT}$, for suitably chosen values of φ_i and ψ_i for each residue i and summing over all "states" φ_i, ψ_i so defined, and over all residues.[6-8] Good agreement was established between theoretical and experimental values of the mean-square unperturbed end-to-end distance[6-9] and the mean-square dipole moment[10] for polypeptide chains of varying amino acid composition and sequence.[6-10]

The configuration of the poly-L-proline chain with all units *trans* is the main concern of the present paper. The rotation angle φ is taken to be fixed by the rigid geometry of the pyrrolidine ring; hence, the conformational energy of a residue depends on ψ alone. Further studies[11] dealing with poly-L-proline I and the cooperative I $\rightleftarrows$ II transition are reported in brief.

Calculation of the Conformational Energy and Chain Dimensions.—Two units of a poly-L-proline II chain are displayed in Figure 1. All bond lengths and bond angles are taken from Sasisekharan's X-ray and optical diffraction investigation of poly-L-proline II.[2] The imide group is assigned to the planar *trans* configuration. The distance between consecutive α-carbon atoms is then fixed at $l_u = 3.80$ Å. A poly-L-proline chain consisting of $x + 1$ residues may be treated, therefore, as a sequence of x virtual bonds of length l_u joining the consecutive α-carbon atoms of $x + 1$ residues. Conformational energies of a *trans* L-prolyl residue in a poly-L-proline II chain were calculated by summing over all van der Waals repulsive ($V_{R,jk}$) and London attractive ($V_{L,jk}$) interactions between atoms separated by one rotation angle ψ. The semiempirical potential functions and parameters obtained by Brant *et al.*[7] were used. Dipolar interactions between adjacent imide groups and the intrinsic torsional potential about the C^α—C bond were legitimately neglected; these contributions to the energy are of minor importance in this chain which is subject to stringent steric hindrances (cf. *Discussion*). The potential associated with rotations about the C^α—C bond, accordingly, is taken to be

$$V(\psi) = \sum_{j,k}(V_{R,jk} + V_{L,jk}), \tag{1}$$
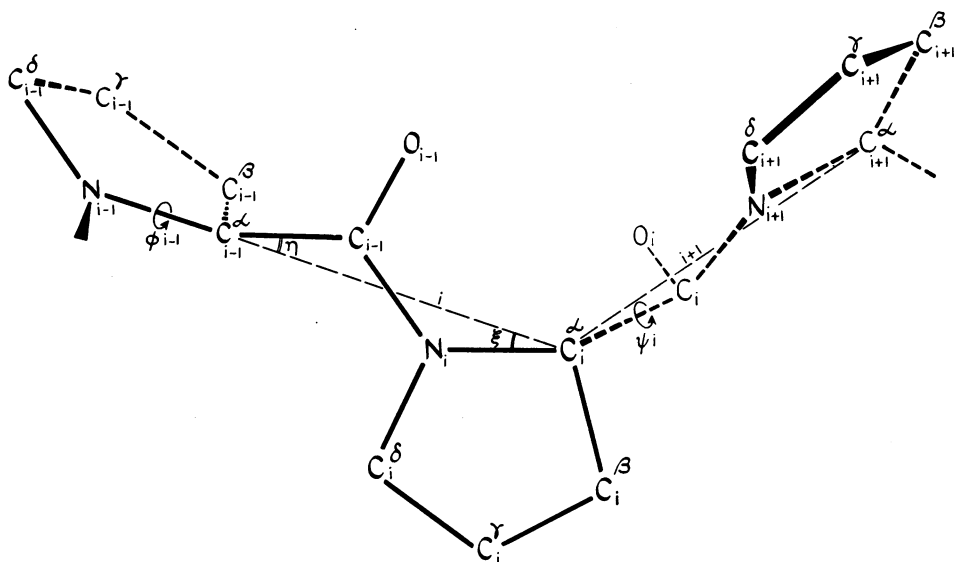
FIG. 1.—Schematic representation of the poly-L-proline II chain with the imide groups in the planar *trans* conformation and the rotation angle φ fixed (*ca.* 102°). Subscripts are serial indexes of proline residues. Virtual bonds i and $i + 1$ are shown by light dashed lines. The angle θ referred to in the text is the supplement of the N—C^α—C bond angle. A discussion of the standard notation used here is given in refs. 7 and 12.

where the summation is over all nonbonded atoms j and k separated by a single rotation angle ψ.

The conformational energy thus calculated is shown in Figure 2. Except in the range $\psi = 275$–$370°$ (or $10°$), steric repulsions give rise to energies more than 7 kcal mole^{-1} above the minimum which occurs at $304°$. With ψ restricted to 275–$370°$, rotations about each C^α—C bond are sensibly independent of rotations of adjacent bonds. Interactions dependent on two ψ angles are significant only when one (or both) of these ψ angles is well outside the sterically accessible range (275–$370°$). This fact, readily demonstrated by inspection of molecular scale models, greatly simplifies treatment of the configurational properties of the chain.

The mutual orientation of two successive virtual bonds, $l_{u,i+1}$ and $l_{u,i}$, is determined by the angle ψ_i, and the configuration of the chain as a whole is determined by specifying the value of every rotation angle ψ_i. Calculation of a configurationally dependent property, e.g., the mean-square unperturbed end-to-end distance, is readily carried out for a chain of any length x by application of the mathematical methods[6] which are summarized below.

A right-handed Cartesian coordinate system $X_iY_iZ_i$ is defined for each virtual bond, the axis X_i being taken in the direction of virtual bond i and Y_i in the plane of virtual bonds $i - 1$ and i. The transformation matrix T_i transforms a vector expressed in coordinate system $i + 1$ to its expression in coordinate system i. This transformation matrix is a function of rotation angles φ_i and ψ_i, and of the angles θ, η, and ξ defined in Figure 1 and its legend. These latter angles are fixed by the structure at $69.5°$, $22.2°$, and $13.2°$, respectively,[2] and the pyrrolidine ring fixes

φ_i at *ca.* 102°.[2] Hence, T_i is a function only of ψ_i apart from fixed geometrical parameters, i.e., η and ξ (for further details, see ref. 7).

The mean-square unperturbed end-to-end distance $\langle r^2 \rangle_0$ between terminal alpha atoms is given by

$$\langle r^2 \rangle_0 = \sum_{j=1}^{x} \sum_{i=1}^{x} \langle l_{u,i}{}^T \cdot l_{u,j} \rangle$$

$$= x l_u{}^2 + 2 \sum_{j=1}^{x} \sum_{i=1}^{j-1} l_{u,i}{}^T \langle T_i T_{i+1} \ldots T_{j-1} \rangle l_{u,j}, \tag{2}$$

where the angular brackets denote statistical mechanical averages and $l_{u,i}$ is the column vector representation of virtual bond i, with elements l_u, 0, 0; $l_{u,i}{}^T$ is the transpose or row form of this vector. For the homopolymers consisting exclusively of *trans* L-prolyl residues the matrix T_i is identical for all residues. Hence, the serial index may be omitted. Dependence of each transformation T_i exclusively on ψ_i permits the averaged product of T matrices to be replaced by the product of averaged matrices $\langle T \rangle$. Equation (2) for the characteristic ratio $\langle r^2 \rangle_0 / x l_u{}^2$ may therefore be written

$$\langle r^2 \rangle_0 / x l_u{}^2 = 1 + 2 \sum_{k=1}^{x-1} (1 - k/x)(\langle T \rangle^k)_{11}, \tag{3}$$

where $(\langle T \rangle^k)_{11}$ denotes the 1,1 element of the matrix $\langle T \rangle^k$. The summation in equation (3) is readily performed with the result

$$\langle r^2 \rangle_0 / x l_u{}^2 = \{ (E + \langle T \rangle)(E - \langle T \rangle)^{-1} - 2/x(\langle T \rangle)(E - \langle T \rangle^x)(E - \langle T \rangle)^{-2} \}_{11}, \tag{4}$$

where E is the identity matrix of order 3.

The averaged matrix $\langle T \rangle$ is given by

$$\langle T \rangle = \frac{\displaystyle\int_0^{2\pi} T(\theta, \xi, \eta, \varphi, \psi) \exp\left[\frac{-V(\psi)}{RT}\right] d\psi}{\displaystyle\int_0^{2\pi} \exp\left[\frac{-V(\psi)}{RT}\right] d\psi}. \tag{5}$$

The integrals were evaluated numerically by summing over discrete rotational states from 275° to 370° in increments of 5°. The result is[13]

$$\langle T \rangle = \begin{bmatrix} 0.44 & 0.30 & 0.81 \\ -0.54 & -0.62 & 0.52 \\ 0.71 & -0.69 & -0.11 \end{bmatrix}. \tag{6}$$

The characteristic ratio calculated from equations (4) and (6) is displayed in Figure 3 as a function of chain length. For comparison, the characteristic ratios obtained previously[6, 7, 9] for homopolymers of glycine and of alanine are shown also.

The foregoing estimation of the conformational energy as represented in Figure 2 is subject to inaccuracies in detail. Consequences of systematic alterations of the shape of the potential function were therefore investigated by replacing the curve in Figure 2 by a series of parabolic potentials centered at $\psi_0 = 310°$, the angle about which the lower portion of the curve in Figure 2 is approximately symmetric.

Sterically unfavorable values of ψ (10–275°) were disallowed. The resulting potential function is

$$V(\psi) = k(\psi - \psi_0)^2, \quad 275° \leqslant \psi \leqslant 370°;$$
$$= \infty, \quad 10° < \psi < 275°; \tag{7}$$

where k is an adjustable parameter. The sensitivity of the chain dimensions to the rotational potential was investigated by calculating $\langle T \rangle$ and $\langle r^2 \rangle_0 / x l_u^2$ according to equations (4) and (5) for various k's. Through trial calculations $k = 0.00148$ kcal mole^{-1} deg^{-2} was found to reproduce within 1.0 per cent the characteristic ratios calculated from equation (1). Selected parabolic potential functions are shown in Figure 4. The limiting value of $\langle r^2 \rangle_0 / x l_u^2$ for $x = \infty$ is indicated at the right-hand terminus of each curve; the value of this ratio similarly calculated for a degree of polymerization $x = 255$ is given in parentheses.

The rotational hindrance potential may be approximated alternatively as a square well of width 2Δ centered about ψ_0:

$$V(\psi) = 0, \quad \psi_0 - \Delta \leqslant \psi \leqslant \psi_0 + \Delta;$$
$$= \infty, \quad \psi_0 + \Delta < \psi < \psi_0 - \Delta. \tag{8}$$

A value of $\Delta = 24$ degrees reproduces the calculations of $\langle r^2 \rangle_0 / x l_u^2$ obtained with the potential function of equation (1) within 1.0 per cent at all chain lengths. The effect of varying Δ on $\langle r^2 \rangle_0 / x l_u^2$ for $x = 255$ and $x = \infty$ is shown in Figure 5.

Discussion.—Steric overlaps in poly-L-proline II cause the conformational energy to exceed its minimum value prohibitively (i.e., by more than *ca.* 4 kcal mole^{-1}) for over three fourths of the range of ψ. The energy rises rapidly at 280° due to unfavorably small distances between O_{i-1}—C_{i+1}^δ and C_i^β—O_i; its rise near 0° is due to contacts between C_i^β—C_{i+1}^δ, C_i^β—N_{i+1}, N_i—O_i, O_{i-1}—O_i, and C_{i-1}—O_i.

Conformations in the neighborhood of $\psi = 120$–150° for an L-proline dipeptide were represented by Leach, Nemethy, and Scheraga[14] as being permitted, although
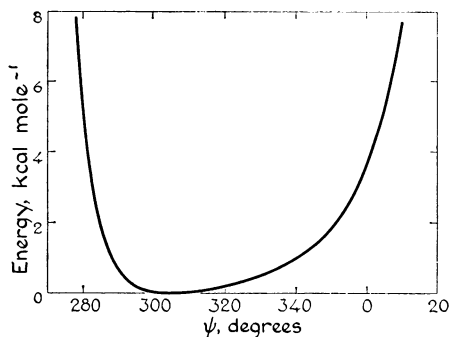
FIG. 2.—Relative conformational energy as a function of ψ, calculated according to eq. (1).
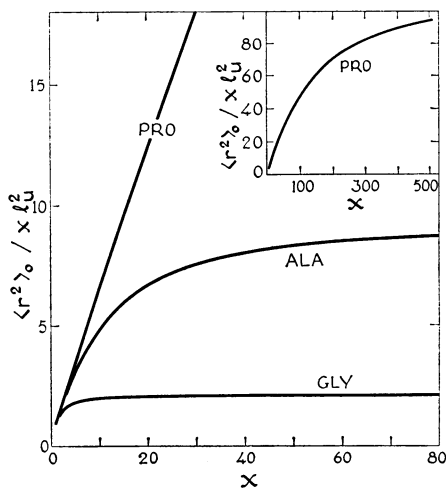
FIG. 3.—$\langle r^2 \rangle_0 / x l_u^2$ plotted against the number of units x for homopolymers of glycine (GLY), alanine (ALA), and proline (PRO).
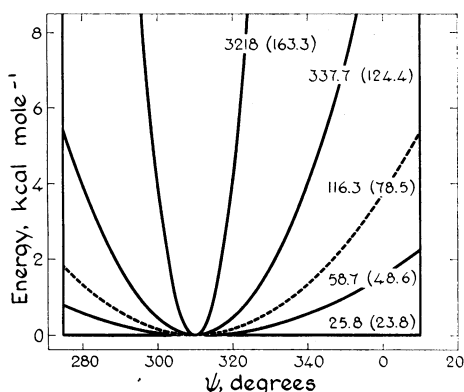
FIG. 4.—Parabolic rotational potential functions calculated according to equation (7). The value of $<r^2>_0/xl_u^2$ for $x = \infty$ is given at the right-hand terminus of each curve; the value of this ratio for $x = 255$ is given in parentheses. The dashed curve reproduces $<r^2>_0/xl_u^2$ for all x within 1% of the characteristic ratios calculated with the energy function of Fig. 2.

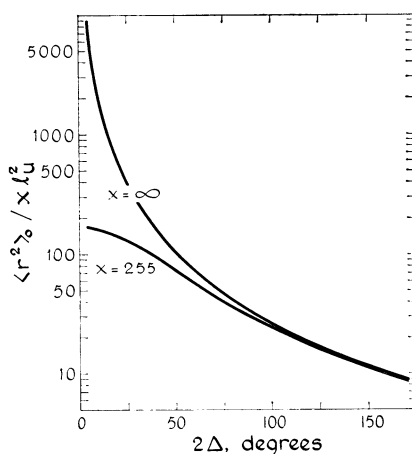FIG. 5.—Values of $<r^2>_0/xl_u^2$ as a function of the width 2Δ of a square-well potential centered about ψ_0 (see eq. 8) for $x = 255$ and $x = \infty$.

disallowed by steric repulsions if adopted by a pair of successive ψ angles for a sequence of three or more *trans* L-prolyl residues. We find, however, that even for a pair of such residues (related by a single ψ angle) the energy for a conformation in this range is prohibitive, being in excess of 60 kcal mole^{-1} due to overlaps between $C_{i+1}{}^\delta$ and atoms of the ith proline ring. Leach *et al.*[14] used coordinates for the prolyl residue differing slightly from ours. However, the marked difference in result can scarcely be resolved on this basis alone. The steric repulsions cited here are readily confirmed by inspection of Corey-Pauling-Koltun (CPK) molecular scale models. We have confirmed the findings of Leach *et al.* that unfavorable contacts between residues separated by two virtual bonds occur when both intervening angles assume values near $\psi = 120°$.[14] The α-helical conformation is therefore disallowed.

A somewhat different conformational energy calculation of poly-L-proline II, confined to rigid helices with exclusion of all irregular forms, has been reported by De Santis *et al.*[15] The angles about each C^α—C bond were varied *in concert* and the energy was calculated from the repulsive van der Waals interactions only. We have found that interactions dependent on only a single ψ angle are sensibly independent of the rotational states of adjacent residues when all ψ angles are restricted to the 275–370° domain. For this reason, the shape of our potential energy curve and theirs, calculated for one of a sequence of residues in identical conformations, is in good agreement within this range of ψ. Such discrepancies as occur are due to minor differences in the residue coordinates, to a difference in the repulsive potentials used, and to our having taken account of attractive (London dispersion) energies.

In the poly-L-proline II crystal, ψ is *ca.* 327°.[2] The energy calculated for this angle is only 0.4 kcal mole^{-1} above the minimum. *Intermolecular* interactions in

the crystal may easily affect the position of the minimum to the extent indicated (*ca.* 20°).

Omission of the electrostatic energy of interaction between dipoles of adjacent *imide* groups is justified by the fact that only one region of the energy map (i.e., energy function[6, 7] of φ and ψ) is allowed. Variation of the electrostatic energy in the permitted range is not great.[16] Inclusion of the dipole-dipole energy, which is attractive in most of this range, would have the effect of narrowing slightly the energy function shown in Figure 2.[17]

The average chain dimensions calculated for the poly-L-proline II chain, as expressed in the ratio $\langle r^2 \rangle_0 / x l_u^2$, are large and the convergence of this ratio with x is slow. These features are consequences of the form of the conformational energy function in Figure 2, in conjunction with the fact that the rotation angles about the N—C^α and C—N bonds are fixed. The contrasting behavior of this ratio for the polyglycine and the poly-L-alanine chains in which rotations are permitted about N—C^α bond is apparent in Figure 3.

The highly extended nature of poly-L-proline II is further demonstrated by comparison of $\langle r^2 \rangle_0$ in the stiff random coil to r_h^2 in the Cowan-McGavin helix. For chains of 16 and 128 residues, $\langle r^2 \rangle_0$ is 94 per cent and 64 per cent, respectively, of r_h^2. This behavior agrees with the experimental viscosity results (in acetic acid) of Steinberg *et al.*[18] who found that the dimensions of poly-L-proline II, deduced on the basis of an ellipsoidal model, agree with those calculated for the Cowan-McGavin helix for a chain length of *ca.* 50 residues.[18] The intrinsic viscosities of polymers of higher molecular weight[18] indicate departures from the dimensions of the rigid helix, and these departures increase with chain length in qualitative accord with the predictions above. However, the usual asymptotic equations for treating the hydrodynamic behavior of random coils should not be applied to a chain having the rigidity evident in poly-L-proline, even at chain lengths of 1000 units.

Since the conformation over short sequences of units in the stiff random coil and in the helix are comparable, the optical rotatory properties of the dissolved molecule are probably similar to those of a rigid helix. Pysh has shown recently that the rotational strengths per residue in the Cowan-McGavin helix are virtually independent of chain length above five residues.[19] Inasmuch as mutual orientations of the residues in the stiff random coil should approximate the rigid helix within sequences of this length, the rotational strengths should be comparable to the value for the helix.

The dimensions of poly-L-proline II in solution are known to be very sensitive to the solvent medium. For example, Steinberg *et al.* have shown that the intrinsic viscosity (mol wt, 19,000) drops *ca.* 40 per cent when the polymer is transferred from acetic acid to water.[18] The reduction of the intrinsic viscosity may have been brought about by isomerization of an occasional imide bond to the *cis* form. Alternatively, or additionally, it may reflect solvent-induced alterations in the rotational hindrance potential about the C^α—C bond. The results of our calculations summarized in Figures 4 and 5 demonstrate that the chain dimensions should be extremely sensitive to relatively small changes in the rotational potential. The large influence of the rotational hindrance potential on the chain dimensions is due to the fact that the rotations are restricted, sterically, to a small domain. When all rotation angles are fixed at the same value (i.e., for a helix), $\langle r^2 \rangle_0 / x l_u^2$ rises without

bound with increasing number of residues. Small deviations in the rotation angles from their fixed values rapidly reduce $\langle r^2 \rangle_0 / x l_u^2$ and cause it to converge to a finite limit. In poly-L-proline II an analogous situation occurs when the effective accessible domain for rotations about the C^α—C bond is increased. The effect of variations in the hindrance potential on the chain dimensions is naturally greatest for long chains.

The comparative insensitivity of optical rotatory power, for reasons given above, is reflected in the observation that large changes in intrinsic viscosity are not always accompanied by comparable changes in optical rotation.[18] Lithium bromide, however, causes a change in both intrinsic viscosity and optical rotation which possibly results from large alterations in the rotational potential function.[18, 20, 21] Some penetration into the sterically disallowed domain and departures from rigidity at the imide bond may occur since LiBr binds to the polymer.[21]

The preceding analysis of poly-L-proline II is in accord with the experimental observations and provides a rational interpretation of the available data. A meaningful quantitative comparison of experimental and theoretical values for the chain dimensions is not possible, however. A relatively small error in the representation of the rotational hindrance potential would lead to a large discrepancy with experiment. Additional discrepancies would arise if a few residues were to adopt the *cis* configuration. Even a very small fraction of the units in the *cis* configuration would markedly lower the chain dimensions. The occurrence of a *cis* unit in a mainly *trans* chain can be shown to be opposed by strong steric overlaps. Hence, *cis* units should be of rare occurrence, except under conditions which approach those required for the conversion II → I.

A small proportion of D-prolyl units would also depress the chain dimensions. This fact should be borne in mind in investigations on polyproline: a high degree of stereoregularity (*ca.* 99%) of the asymmetric centers of the prolyl units may be required for significant quantitative experiments on the chain dimensions.

The *cis* residues comprising poly-L-proline I are considerably more hindered than those of poly-L-proline II. In contrast to poly-L-proline II, rotations within a residue of I are strongly dependent upon the rotations of its neighbors for all values of ψ; significant interactions dependent on three rotation angles occur. The residues in poly-L-proline I are much more tightly packed than in poly-L-proline II. The conformation I is preferred in poor solvents[18] wherein relatively favorable intrachain interactions supplant less favorable polymer-solvent interactions. The occurrence of I as the stable form under any circumstances seems explicable only if the groups forced into close contact in this form interact attractively. Additionally, the realization of this favorable energy by a given *cis* residue requires that several of its neighbors (on both sides) adopt the *cis* configuration. This circumstance imposes the condition of simultaneous formation of sequences of residues in the *cis* configuration and hence renders the I ⇌ II transformation cooperative, as evidenced by its occurrence within a narrow interval of solvent composition.[22, 23] The effect of chain length on the location of the transition shows the stability of the all *cis* chain to be enhanced by an increase in chain length.[23] This observation offers further evidence for the importance of intrachain interactions between neighboring units as a factor favoring form I.

Note added in proof: Recent energy transfer experiments (Stryer, L., and R. Haugland, these PROCEEDINGS, in press), employing oligomers of poly-L-proline II as spacers between energy donor and energy acceptor, confirm our calculation of the dimensions of the short chains.

* This work was supported by the Directorate of Chemical Sciences, Air Force Office of Scientific Research contract no. AF49(638)1341.

[1] Cowan, P. M., and S. McGavin, *Nature*, **176**, 501 (1955).

[2] Sasisekharan, V., *Acta Cryst.*, **12**, 897 (1959).

[3] Traub, W., and U. Shmueli, *Nature*, **198**, 1165 (1963).

[4] For a recent review, see Carver, J. P., and E. R. Blout, "Polypeptide models for collagen," a chapter to be published in *Treatise on Collagen: Chemistry of Collagen*, ed. G. N. Ramachandran (London: Academic Press, 1967), vol. 1.

[5] A review of the earlier work is given by Harrington, W. F., and P. von Hippel, *Advan. Protein Chem.*, **16**, 1 (1961).

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

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

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

[9] Brant, D. A., and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2788 (1965).

[10] Schimmel, P. R., and P. J. Flory, in preparation.

[11] Schimmel, P. R., and P. J. Flory, in preparation.

[12] Edsall, J. T., P. J. Flory, J. C. Kendrew, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *Biopolymers*, **4**, 121 (1966).

[13] Integration over $\psi = 10$–$275°$ was legitimately neglected because the relative energy is greater than 7 kcal mole^{-1} in this range. Summing in increments of $1°$ did not alter ($\pm 0.01\%$) the results obtained with the energy function of eq. (1). Smaller increments of ψ were used in some of the subsequent calculations with the square well potential function (eq. 8); the results in this case show a small but inconsequential variation with the size of the increment used. A temperature of $310°K$ was assumed in all calculations.

[14] Leach, S. J., G. Nemethy, and H. A. Scheraga, *Biopolymers*, **4**, 369 (1966).

[15] DeSantis, P., E. Giglio, A. M. Liquori, and A. Ripamonti, *Nature*, **206**, 456 (1965).

[16] The dipole–dipole energy was estimated by representing the electrical asymmetry of the imide group by a point dipole assumed identical to the dipole moment of the amide group. Details of the calculation and estimates of the dipole–dipole energy, which is attractive for most of the sterically allowed range of ψ, are given in refs. 6 and 7.

[17] The important role of the dipole energy in poly*amino* acids previously demonstrated[6,7] is a consequence of the presence of *two*, competing conformations well-separated on the energy map. One of these includes the preferred conformation for a chain of L-prolyl residues; the other approximates the residue conformation in the α-helix. Dipolar interactions affect these competing regions quite differently. Exclusion of one of them by the restriction on ψ for poly-L-proline obviates assessment of the electrostatic contribution to the conformational energy.

[18] Steinberg, I. Z., W. F. Harrington, A. Berger, M. Sela, and E. Katchalski, *J. Am. Chem. Soc.*, **82**, 5263 (1960).

[19] Pysh, E. S., *J. Mol. Biol.*, **23**, 587 (1967).

[20] Harrington, W. F., and M. Sela, *Biochim. Biophys. Acta*, **27**, 24 (1958).

[21] Kurtz, J., and W. F. Harrington, *J. Mol. Biol.*, **17**, 440 (1966).

[22] Gornick, F., L. Mandelkern, A. F. Diorio, D. E. Roberts, *J. Am. Chem. Soc.*, **86**, 2549 (1964).

[23] Engel, J., *Biopolymers*, **4**, 945 (1966).