Monte Carlo Calculation on *trans/cis*-Polysarcosine

Masahiko Sisido,* Yukio Imanishi, and Toshinobu Higashimura

Department of Polymer Chemistry, Kyoto University, Kyoto 606, Japan.
Received October 10, 1975

ABSTRACT: Monte Carlo calculations were made on unperturbed *trans*-polysarcosine chain, non-self-intersecting *trans*-polysarcosine chain, and also non-self-intersecting *trans/cis*-polysarcosine chain by using a hard-sphere model. In the last case, an attempt was first made to introduce cis amide bond into the Monte Carlo calculation of polypeptide chain. Dipeptide energy maps for four different trans/cis dyad sequences were calculated. The allowed regions were consistent with the pairs of dihedral angles observed in *cyclo*-pentasarcosyl and *cyclo*-octasarcosyl. The mean-square end-to-end distance and higher even moments were obtained. The distribution function of the end-to-end distance was calculated from the even moments by using Nagai's equation and compared with the direct Monte Carlo data. The best agreement was obtained by cutting off the terms containing much higher order even moments than a critical order. The fractions of cis amide bond and of the four different trans/cis dyad sequences in polysarcosine were calculated. The results of calculation were compared with the 220-MHz NMR spectra of polysarcosine in three different solvents. Qualitative agreement was attained for longer chains.

Theoretical studies on the conformational properties of randomly coiled poly(α-amino acids) have been made for the unperturbed states and for the perturbed states owing to the excluded volume effect. For the unperturbed states, the matrix method has been found to be useful.[1] On the other hand, Monte Carlo method gives a great deal of information on the non-self-intersecting chains. Thus Monte Carlo studies have been carried out on poly(α-amino acids) by several workers.[2-11] Usually a planar amide bond of poly(α-amino acid) takes trans conformation in which a C^α atom is trans against the succeeding C^α atom. This makes the conformational calculation considerably easy since the correlation between sets of rotational angles (φ,ψ) for the neighboring units can be ignored.[1] However, if cis amide bonds are involved the correlation cannot be ignored.[12,13] In polysarcosine chain the existence of cis amide bond was first demonstrated by Bovey et al.[14] by using 220-MHz NMR spectroscopy. The present authors[15] assigned *N*-methyl resonance signals to four dyad sequences of trans/cis amide bond conformations.

Until now conformational properties of random polysarcosine chain have been calculated by three groups. Tanaka and Nakajima[16] calculated the unperturbed dimension using a matrix method. Scott and co-workers[5] made Monte Carlo calculations for the unperturbed state as well as the perturbed state. In these studies, however, the amide bond was fixed to trans form. This made the calculation very simple. Recently Burgess et al.[17] calculated the unperturbed dimension of *cis*-polysarcosine chain in which all amide bonds are fixed to cis form. However, they ignored the conformational correlation associated with the introduction of cis amide unit.

In this paper the Monte Carlo calculation on a hard-sphere polysarcosine including cis as well as trans amide bond will be described. The difficulty arising from the conformational correlation was avoided by treating rather short chains and by making several devices for the calculation (see below). Conformational properties such as even moments of end-to-end distance and the distribution function were calculated. The fraction of cis form and the fractions of four dyad sequences were also determined. The calculated values were compared with the 220-MHz NMR spectra recorded in D_2O, ethanol-d_6, and $CDCl_3$ solutions.

Experimental Section

N-Acetylpoly(sarcosine)dimethylamide-d_6 (I) was synthesized as follows. Sarcosine *N*-carboxyanhydride (NCA) was polymerized in dimethylformamide by using dimethylamine-d_6 (DMA) as an initiator. The number-average degree of polymerization is given by the molar ratio of NCA to DMA (eq 1).[18]

$$(CD_3)_2N\text{-}(\text{-}CO\text{-}CH_2\text{-}N(CH_3)\text{-})_{n-1}\text{-}COCH_3 \quad \text{I}$$

$$\bar{n}' = \bar{n} - 1 = [\text{NCA}]/[\text{DMA}] \tag{1}$$

After the polymerization was completed (1 day), the terminal secondary amino group was acetylated by adding an acetic anhydride/pyridine mixture. The polymer solution was poured into ether and the precipitated polymer was collected and washed thoroughly with ether. The number-average degree of polymerization calculated according to eq 1 was compared with that determined on the basis of the ratio of the NMR peak areas for the terminal *N*-acetyl group and the *N*-methyl group in the main chain. The results are listed in Table I. The agreement between the calculated and observed values is satisfactory. NMR spectra (220-MHz) were recorded on a Varian HR-220 instrument at ambient temperatures in 10 w/v % solutions.

Method of Calculation

Structural Parameters. Structural parameters used in this study were the same as those used by Scott and co-workers[5] (Figure 1). An amide unit was assumed to be planar and allowed to take cis as well as trans conformation. All *N*-methyl groups and the terminal acetyl methyl group were treated as a rigid sphere. The atomic radii were the same as those taken by Scott and co-workers.[5] Accordingly the radius of a methyl group was somewhat smaller than the usual value 1.75 Å. If the latter was used, almost all conformations of sarcosine dipeptide were prohibited. Furthermore, if we took a value larger than or equal to 1.525 Å for a methyl group, any conformation including cis amide unit was rejected from the dipeptide. Thus the radius of the methyl group was taken to be 1.5 Å also in the present study.

Dipeptide Maps. Ranges allowed for a pair of rotational angles (φ, ψ) were calculated for *trans,trans-*, *trans,cis-*, *cis,trans-*, and *cis,cis*-sarcosine dipeptides. Rotational angles were varied from 5 to 355° at an interval of 10°. Results are shown in Figure 2. For the present set of structural parameters, there was no difference between trans,trans and trans,cis sequences and also between cis,trans and cis,-cis sequences. The map for trans,trans sequence differed from that of Scott and co-workers[5] although the same parameters were used. In the latter the regions at φ = 300–355°, ψ = 110° and φ = 5–60°, ψ = 250° were allowed. The same regions are forbidden in the present case due to the atomic contacts, for example, between N_i–CH_3 and C_i (see

Table I
Number-Average Degree of Polymerization $\bar{n}'$ of Polysarcosine

$\bar{n}'_{calcd}$[a]	$\bar{n}'_{obsd}$[b]	$\bar{n}'_{calcd}$[a]	$\bar{n}'_{obsd}$[b]
3.0	3.9	15	16
5.0	6.3	20	25
10	10	30	29

[a] Calculated from eq 1. [b] Ratio of NMR peak areas for the terminal *N*-acetyl group and for the *N*-methyl group.

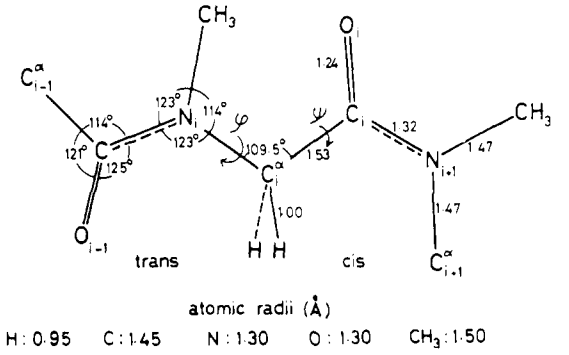
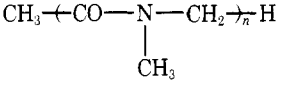


Figure 1. Structural parameters for *trans/cis*-polysarcosine chain.

Figure 1). The same regions were allotted to the considerably high-energy regions in the calculations made by Tanaka and Nakajima,[16] Conti and DeSantis,[19] and Mattice[20] who used the semiempirical potentials. These considerations suggest our dipeptide map to be reasonable. In Figure 2 are also plotted the pairs of dihedral angles observed with *cyclo*-pentasarcosyl[21] and *cyclo*-octasarcosyl[22] in the crystalline state. All the observed points lie in the neighborhood of the allowed regions calculated. This shows the reliability of the present dipeptide maps. Having varied the rotational angles from 5 to 355° at an interval of 10°, maps for trans,trans, (trans,cis) and cis,trans (cis,cis) yielded 38 and 16 allowed states, respectively.

Generation of Monte Carlo Chains. Monte Carlo calculations were performed for structure II, which corresponds to *N*-acetylpoly(sarcosine) dimethylamide.



II

Unperturbed *trans*-Polysarcosine. Monte Carlo chains of polysarcosine with only trans amide bonds were generated by using 38 allowed pairs of rotational angles in the dipeptide map shown in Figure 2. For the generation of unperturbed polysarcosine chains, only the coordinates of α carbons were calculated. The first α carbon was placed at the origin and the end-to-end distance was measured from the origin to the last α carbon. For each chain length, 5000 to 95 000 chains were generated, and the averaged quantities and the distribution of the end-to-end distance were obtained.

Non-Self-Intersecting *trans*-Polysarcosine. To generate non-self-intersecting *trans*-polysarcosine, it is necessary to calculate the coordinates of all atoms in the chain and to reject the chains having any atomic overlap. The atomic overlap was checked at every stage to add a new sarcosine unit to the chain. It was found that there was no possibility for atomic overlaps between an nth unit and any unit closer than the $(n - 4)$th unit. Therefore the check

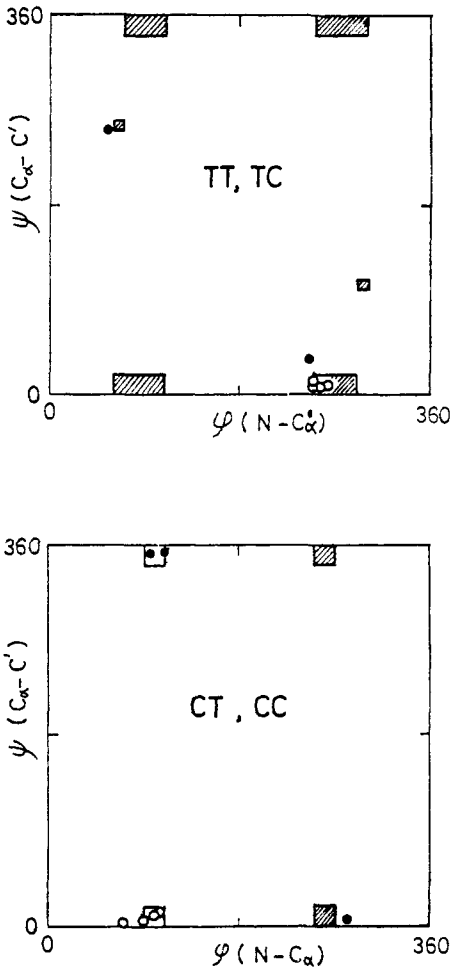


Figure 2. Dipeptide maps for the four dyad sequences of polysarcosine. The shaded areas represent the sterically allowed states. Rotational states observed with *cyclo*-pentasarcosyl[18] (●) and *cyclo*-octasarcosyl[19] (○) are also shown.

was omitted until the fourth unit was introduced and for the newly added nth unit, it is necessary to check with 1st, 2nd, . . . , $(n - 4)$th units. If an atomic overlap was detected, the chain was discarded and the calculation was restarted. For each chain length 5000 non-self-intersecting chains were generated.

Non-Self-Intersecting *trans/cis*-Polysarcosine. In this case an amide bond was allowed to take cis as well as trans conformation with an equal probability. Two major difficulties arose by the introduction of the cis amide bond. The first is the correlation of conformation between the neighboring units. As seen from Figure 2, the allowed regions for the trans unit succeeding to trans unit (trans,-trans) differs from those succeeding to cis unit (cis,trans). This conformational correlation can be avoided formally by adopting the "extended allowed regions" as the allowed regions for a sarcosine unit irrespective of the dyad sequences. The extended allowed regions include all regions allowed for the four dyad sequences and actually are the same as the allowed regions for trans,trans (trans,cis) sequence. In the generation of a polysarcosine chain a newly added sarcosine unit chooses the conformation of amide bond (trans or cis) and the rotational angles (φ, ψ) among 38 extended allowed states at once.

Using this procedure the conformational correlation was avoided, but severe chain attrition occurred. In fact an atomic overlap may occur even between the neighborng sarcosine units. This is the second major problem. This difficulty was reduced partly by using the Wall-Erpenbeck

Table II
Mean-Square End-to-End Distance $\langle r^2\rangle_0$ and Average Fourth Power $\langle r^4\rangle_0$ of Unperturbed *trans*-Polysarcosine

	$\langle r^2\rangle_0 \times 10^2$, Å^2			$\langle r^4\rangle_0 \times 10^4$, Å^4	
n	Monte Carlo	Exact[b]	Wormlike[c]	Monte Carlo	Wormlike[d]
5	1.86 ± 0.01[a]	1.86	1.67	3.74 ± 0.05[a]	3.14
10	4.45 ± 0.06	4.49	4.24	25.3 ± 0.5	23.0
15	7.27 ± 0.10	7.15	6.90	72.7 ± 1.8	65.4
20	9.78 ± 0.15	9.81	9.55	140 ± 4	131
30	15.3 ± 0.3	15.1	14.9	353 ± 11	334
40	20.6 ± 0.4	20.5	20.2	660 ± 23	631

[a] 90% confidence limit. [b] Calculated from eq 3. [c] Calculated from eq 6. [d] Calculated from eq 7.

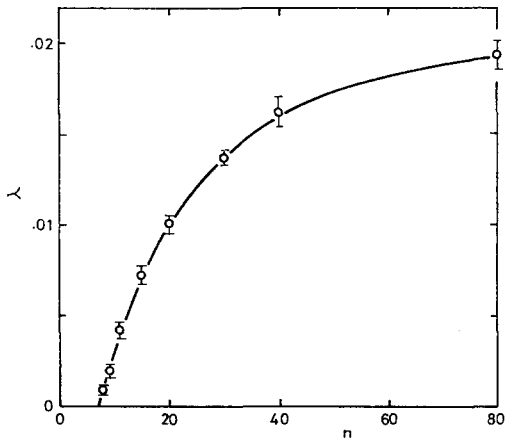


Figure 3. A plot of attrition parameter λ vs. chain length n for non-self-intersecting *trans*-polysarcosine chains. The vertical bars represent the 90% confidence limits.

chain enrichment technique.[23] It was found that the best parameters were $s = 2$ and $p = 3$. The computation time was further reduced by listing the allowed conformations of sarcosine dimers and trimers. Using these lists we can decide whether an nth newly added sarcosine unit overlaps with any atoms in the $(n-1)$th and $(n-2)$th units, without calculating the atomic coordinates. If no overlapping was detected the atomic coordinates of the newly added unit were calculated and checked for overlaps with any atoms in the rest of the chain. The number of completely independent samples was 5000 for short chains ($n \leq 15$), 2500 ($n = 29$), and 250 ($n = 59$). All computations were carried out at the Data Processing Center of Kyoto University by using a FACOM 230/75 system.

Results and Discussion

Chain Attrition. For non-self-intersecting chains, the attrition parameter λ is defined[23] as

$$W_n = W_0 \exp(-\lambda n) \tag{2}$$

where W_n is the number of non-self-intersecting chains having chain length n, and W_0 is the number of trials required to obtain W_n chains. In this section the attrition parameter of non-self-intersecting *trans*-polysarcosine chains will be discussed for the comparison with usual poly(α-amino acids). In Figure 3 is shown a plot of attrition parameter vs. chain length of non-self-intersecting *trans*-polysarcosine consisting of up to 80 sarcosine units. Although λ is still increasing at $n = 80$, it appears that the value at infinite chain length (λ_∞) lies between 0.02 and 0.025. If λ is plotted against $1/n$, λ_∞ should be 0.023 ± 0.001, which seems more reliable. This value of λ_∞ is very small as compared with those for usual poly(α-amino acids)[2-9] and even for polysarcosine reported by Scott and co-workers.[5] This may be explained in terms that few rotational states are allowed for the coiled conformation of *trans*-polysarcosine chain (Figure 2).

The attrition parameter of non-self-intersecting *trans/cis*-polysarcosine chain was about 0.5 for the chain length of 5 to 59. This extraordinary large attrition parameter comes partly from the adoption of the extended allowed regions and partly from the high tendency of cis amide bond to cause intrachain atomic overlaps.

Mean-Square End-to-end Distance and Higher Even Moments. The mean-square end-to-end distances for unperturbed *trans*-polysarcosine obtained by the present Monte Carlo calculation are listed in Table II, together with the exact values calculated by using the analytical equation of Flory.[1]

$$\langle r^2\rangle_0 = nl^2\left\{(\mathbf{E} + \langle \mathbf{T}\rangle)(\mathbf{E} - \langle \mathbf{T}\rangle)^{-1} - \frac{2}{n}\langle \mathbf{T}\rangle(\mathbf{E} - \langle \mathbf{T}\rangle^n)(\mathbf{E} - \langle \mathbf{T}\rangle)^{-2}\right\}_{11} \tag{3}$$

$$l = 3.80 \text{ Å} \tag{4}$$

In eq 3 $\mathbf{E}$ denotes a unit matrix of the order three and l is a distance between the neighboring α-carbon atoms.[1] The average transformation matrix, $\langle \mathbf{T}\rangle$, was calculated in accordance with the dipeptide map shown in Figure 2.

$$\langle \mathbf{T}\rangle = \begin{bmatrix} 0.628 & 0.641 & 0.0 \\ -0.104 & -0.215 & 0.0 \\ 0.0 & 0.0 & -0.048 \end{bmatrix} \tag{5}$$

The agreement between Monte Carlo values and the analytical solutions was satisfactory. The characteristic ratio at infinite chain length $C_\infty = \langle r^2\rangle_0/nl^2$, $n \to \infty$, was calculated from eq 3 to be 3.68. This value is greater than the reported values, which have been obtained using the semiempirical potentials[14] and the hard-sphere model.[5] This is again a consequence of the absence of the coiled states in the dipeptide map.

Premilat and Hermans[11] suggested the resemblance of the conformational properties of short unperturbed poly(α-amino acid) chains to those of Porod–Kratky wormlike chains.[24] In the present study the mean-square end-to-end distance and the fourth moment[25] were calculated according to the wormlike chain model as well,

$$\langle r^2\rangle_0 = 2aL\{1 - (a/L)[1 - \exp(-L/a)]\} \tag{6}$$

$$\langle r^4\rangle_0 = 5(2aL)^2/3 - [26/9 + \exp(-L/a)]L(2a)^3 + \{[\exp(-3L/a) - 1]/54 + 2[1 - \exp(-L/a)]\}(2a)^4 \tag{7}$$

The parameters for characterizing the wormlike chain (L,

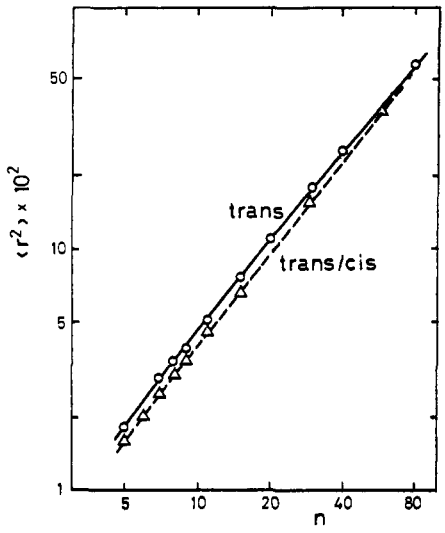


Figure 4. Double logarithmic plot of the mean-square end-to-end distance $\langle r^2 \rangle$ vs. chain length n for non-self-intersecting *trans*- (○) and *trans/cis*- (△) polysarcosine chains. The 90% confidence limits are included within open circles and open triangles.

a) were determined from the chain length and the characteristic ratio,[1]

$$L = 0.95nl = 3.61n \text{ Å}$$
$$a = C_\infty nl^2/2L = 7.37 \text{ Å} \tag{8}$$

The results are compared with the Monte Carlo data and the analytical solutions in Table II. The agreement is fairly good. This indicates that the wormlike chain model is adequate to describe short polysarcosine chains. Adequacy of the wormlike chain model has also been observed in the calculation of the ring-closure probability.[26]

The logarithmic plot of the mean-square end-to-end distance against the chain length is shown in Figure 4 for non-self-intersecting *trans*- and *trans/cis*-polysarcosine chains. Usually $\langle r^2 \rangle$ is related to n according to the following type of equation,[2-8]

$$\langle r^2 \rangle = an^b \tag{9}$$

For the lack of values for longer chains, the least-square fit was not carried out. The approximate values of a and b were $33._0$ and 1.1_8 for *trans*-polysarcosine and $22._5$ and 1.2_5 for *trans/cis*-polysarcosine, respectively.

The higher even moments, $\langle r^{2p} \rangle$, were also obtained and used to estimate the distribution of end-to-end distances as described in the next section.

Distribution of End-to-End Distances. In Figures 5–7 are shown the distribution of end-to-end distances for unperturbd *trans*-, non-self-intersecting *trans*-, and non-self-intersecting *trans/cis*-polysarcosine chains having 7, 11, and 20 units. Circles in the figures represent the distribution determined by counting the Monte Carlo chains having a given value of end-to-end distance. The count was made with 1 Å increments irrespective of the chain length, so the plot seems more scattered for longer chains. The distributions for unperturbed *trans*-polysarcosine chains are compared with the corresponding Gaussian distribution functions in Figures 5–7. The Monte Carlo results show more pronounced maxima than the Gaussian functions, and the deviation from the Gaussian curve is more distinct for shorter chains.

In Figures 5–7 are also shown the moment-based distribution function which was reconstructed by using the even

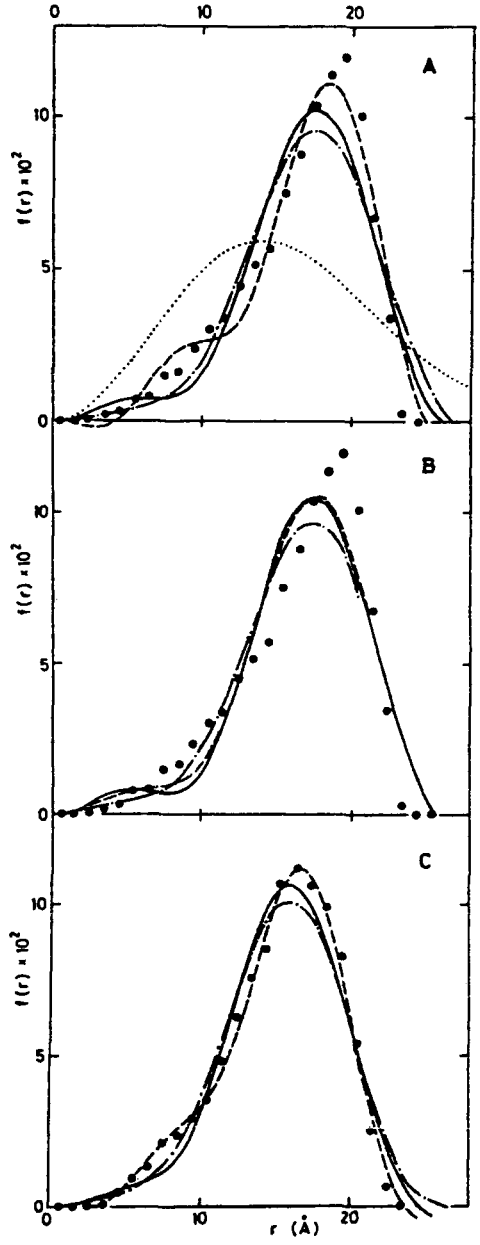


Figure 5. Distribution of the end-to-end distance of Monte Carlo chains of (A) unperturbed *trans*-polysarcosine, (B) non-self-intersecting *trans*-polysarcosine, and (C) non-self-intersecting *trans/cis*-polysarcosine, $n = 7$. Moment-based distribution functions calculated by using up to 16th (- · -), 24th (—), and 32nd (- - -) even moments are also shown. The dotted line in A shows the Gaussian distribution function for chains having the same average end-to-end distance as the analytical solution of eq 3.

moments obtained in the present Monte Carlo calculation. According to Nagai,[27] the distribution function may be expanded in terms of the even moments as follows,

$$W(\rho) = \pi^{-3/2} \exp(-\rho^2)[(1 + 15g_4 + 105g_6 + 945g_8 + \ldots) - (20g_4 + 210g_6 + 2520g_8 + 34650g_{10} + \ldots)\rho^2 + (4g_4 + 84g_6 + 1512g_8 + \ldots)\rho^4 - (8g_6 + 288g_8 + \ldots)\rho^6 + \ldots] \tag{10}$$

$$W(\rho) = (2\langle r^2 \rangle/3)^{3/2} W(r) \tag{11}$$

$$g_{2p} = 2^{-p}\left[-\frac{\Delta_4}{2!(p-2)!} + \frac{\Delta_6}{3!(p-3)!} - \ldots \frac{(-1)^{p-1}}{p!}\Delta_{2p}\right] \tag{12}$$

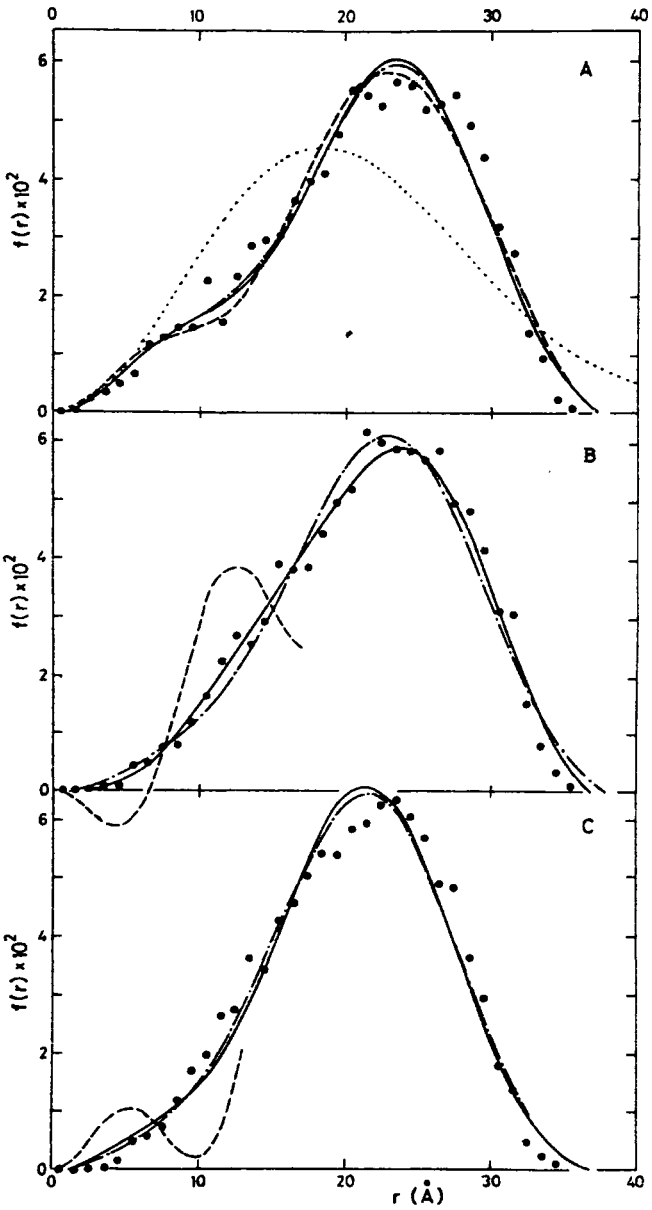


Figure 6. Distribution of the end-to-end distance of Monte Carlo chains, n = 11. For other notations, see Figure 5.

$$\Delta_{2p} = 1 - \left[\frac{3^p}{3 \times 5 \times 7 \times \ldots (2p+1)}\right] \frac{\langle r^{2p} \rangle}{\langle r^2 \rangle^p} \qquad (13)$$

$$p > 0$$

In the above equations Flory's notation was followed.[1] Three types of distribution functions were drawn using up to 16th, 24th, and 32nd even moments. It is seen that the convergence is not necessarily achieved by taking higher even moments. This is expected if we consider the slow convergency of the moment-based distribution function and the inaccuracy of higher even moments due to the sensitiveness to the highly extended conformations which are relatively few.[28] Therefore, to obtain the best fit to the circles, we had better cut off the terms containing higher order even moments than a critical order.

If we compare the distribution for unperturbed trans chains with those for non-self-intersecting trans chains, it is clear that the chain attrition is more significant for the chains having shorter end-to-end distances. This indicates that the excluded volume effect plays an important role to determine the conformational properties of the chains having short end-to-end distances. For example, the ring-closure probabilities for non-self-intersecting *trans*-polysarcosine chains were substantially smaller than the values for unperturbed chains.[26] The distribution for non-self-intersecting *trans/cis*-polysarcosine is little different from those for trans chain, except a small shift of the peak position for the former to a shorter end-to-end distance.

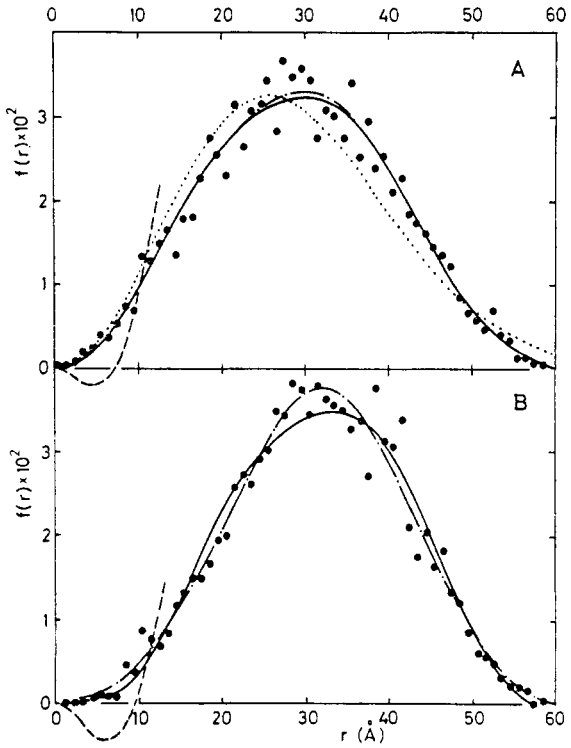


Figure 7. Distribution of the end-to-end distance of Monte Carlo chains, n = 20. For other notations, see Figure 5.

Features in Figures 5–7 are much the same as those observed for unperturbed polyglycine and poly-L-alanine.[11]

Fractions of Cis Amide Unit and Cis/Trans Dyad Sequence in Non-Self-Intersecting *trans/cis*-Polysarcosine. As described in the previous section, the occurrence of cis and trans amide bond was set being equally probable at the beginning of Monte Carlo calculation of non-self-intersecting *trans/cis*-polysarcosine chain. As a result of chain attrition, however, the fractions of cis form and trans/cis dyad sequences took other values than the original values, 0.5 and 0.25, respectively. In the present model of polymer structure (II) the terminal *N,N*-dimethylamide bond is symmetric, and trans and cis forms are distinguishable only for the first $(n - 1)$ amide bonds in an n-meric chain. Thus the total cis fraction was calculated by averaging the cis fractions for the first $(n - 1)$ sarcosine units. Similarly the fraction of each dyad sequence was obtained by averaging the fractions for the first $(n - 2)$ dyad sequences. The cis fraction f_{cis} and the fractions of the four dyad sequences obtained in the present Monte Carlo calculation are shown in Figures 8 and 9, respectively. As seen from Figure 8 the cis fraction was about 0.19 and increased slightly with increasing the chain length. In the same figure the cis fraction for cyclic polymer is also shown. The cyclic polymer is defined as the polymer chain with its end-to-end distance being less than 10 Å. The cis fraction was considerably increased in the cyclic conformation of short polysarcosine chains. This trend was more pronounced if the end-to-end distance was confined less than 6 Å or less. In Figure 9 were observed a slight but definite increase in cis,-

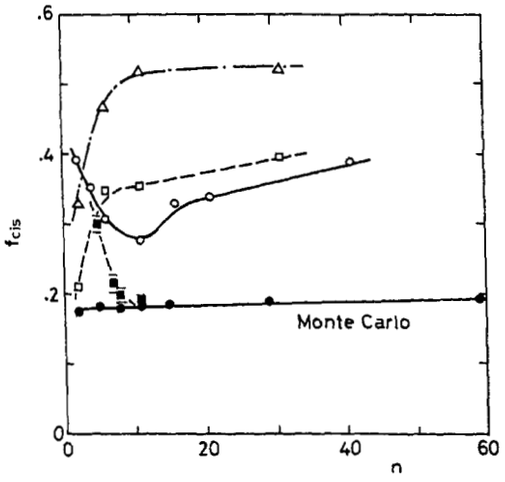


Figure 8. Fraction of cis amide bond in polysarcosine chain as a function of chain length n. (●) Monte Carlo results for non-self-intersecting *trans/cis*-polysarcosine; 90% confidence limits are included within closed circles. (■) Monte Carlo results for cyclic polysarcosine (see text). The horizontal bars represent the 90% confidence limits. (○) Observed values for NMR spectra recorded in D_2O solution. (-·-△-) in ethanol-d_6 solution. (-□-) in $CDCl_3$ solution.

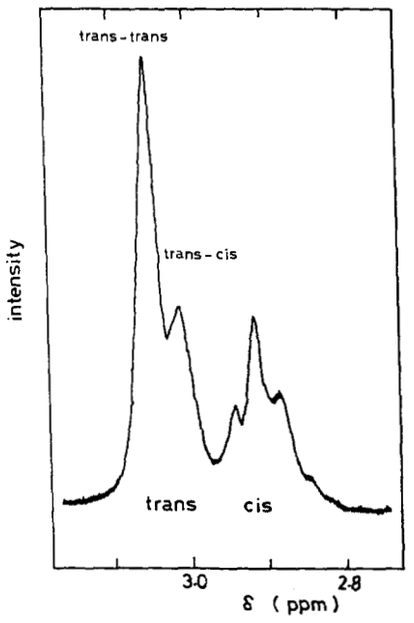


Figure 10. NMR spectrum (220 MHz) of polysarcosine in 10 w/v % D_2O solution at an ambient temperature. Only the N-methyl region of the resonance signal is shown.

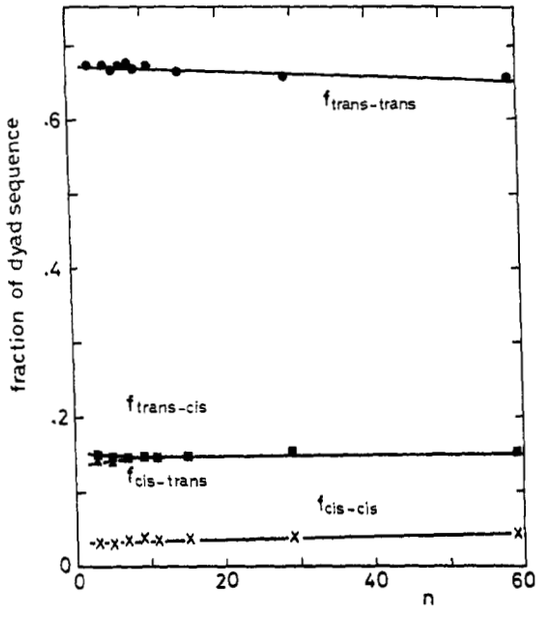


Figure 9. Fractions of four dyad sequences in the Monte Carlo chains of non-self-intersecting *trans/cis*-polysarcosine. Each fraction is plotted as a function of chain length n.

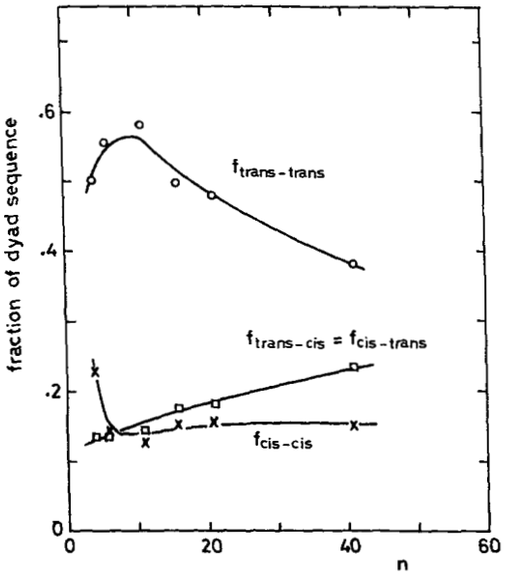


Figure 11. Observed fractions of four dyad sequences of polysarcosine chain in D_2O solution plotted as a function of chain length n.

cis sequence and a decrease in trans,trans sequence with increasing the chain length. The fractions of trans,cis and cis,-trans sequence were nearly unchanged.

In order to compare with the Monte Carlo results, polysarcosines (I) having different average degrees of polymerization were synthesized and their 220-MHz NMR spectra were recorded in D_2O, ethanol-d_6, and $CDCl_3$ solutions. The absorption from the terminal dimethylamide unit which does not show trans/cis isomerism was excluded by using dimethylamine-d_6 as an initiator to prepare the polymer sample. A typical spectrum of the N-methyl band in D_2O solution is shown in Figure 10. A peak assignment has been reported by the present authors.[15] The lowest field peak is ascribed to trans,trans and the second lowest is trans,cis. These two peaks in the lower field correspond to trans conformation. The three peaks in the higher field correspond to cis conformation. Although these three peaks cannot be assigned to the remaining two dyad sequences, the fractions were evaluated by assuming that the fraction of cis,trans sequence is identical with that of trans,cis sequence. As the resolution of spectrum was not sufficient in ethanol-d_6 and $CDCl_3$ solutions, only the fraction of cis amide bond was determined in these solvents.

The cis fraction f_{cis} and the fractions of four dyad sequences $f_{trans,trans}$, $f_{trans,cis}$ ($=f_{cis,trans}$), and $f_{cis,cis}$ are presented in Figures 8 and 11 as functions of the chain length n, respectively. In Figure 8 the cis fractions in the three solvents were substantially larger than the Monte Carlo

values. Since the calculations have been made assuming an equal energy for either cis or trans form, this result implies that the cis form is in practice more stabilized than the trans form in these solvents. In D_2O solution the observed cis fraction showed a distinct minimum at $n = 10$. For longer chains the calculated chain length dependence agreed well with the observed ones in the three different solvents. The qualitative agreement was also observed for the fractions of dyad sequences in D_2O solution (Figure 11) with the Monte Carlo results (Figure 9). As in the case of cis fraction, the caculated chain length dependences of the dyad sequences for longer chains agreed well with the observed ones. As compared with the Monte Carlo results, however, the observed cis,cis fraction was considerably larger and the trans,trans fraction was smaller. The anomalous behavior observed for shorter chains may be a result of specific end effects which depend strongly on the nature of the solvent.

References and Notes

(1) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969.
(2) K. K. Knaell and R. A. Scott III, *J. Chem. Phys.*, **54**, 566 (1971).
(3) K. K. Knaell and R. A. Scott III, *J. Chem. Phys.*, **54**, 3556 (1971).
(4) H. E. Warvari, K. K. Knaell, and R. A. Scott III, *J. Chem. Phys.*, **54**, 2020 (1971).
(5) H. E. Warvari, K. K. Knaell, and R. A. Scott III, *J. Chem. Phys.*, **56**, 2903 (1972).
(6) H. E. Warvari, K. K. Knaell, and R. A. Scott III, *J. Chem. Phys.*, **57**, 1146 (1972).
(7) H. E. Warvari and R. A. Scott III, *J. Chem. Phys.*, **57**, 1154 (1972).
(8) H. E. Warvari, K. K. Knaell, and R. A. Scott III, *J. Chem. Phys.*, **57**, 1161 (1972).
(9) S. Tanaka and A. Nakajima, *Macromolecules*, **5**, 708 (1972).
(10) S. Tanaka and A. Nakajima, *Macromolecules*, **5**, 714 (1972).
(11) S. Premilat and J. Hermans, Jr., *J. Chem. Phys.*, **59**, 2602 (1973).
(12) J. C. Howard, F. A. Momany, R. H. Andreatta, and H. A. Scheraga, *Macromolecules*, **6**, 535 (1973).
(13) S. Tanaka and A. Nakajima, *Polym. J.*, **1**, 71 (1970).
(14) F. A. Bovey, J. J. Ryan, and F. P. Hood, *Macromolecules*, **1**, 305 (1968).
(15) M. Sisido, Y. Imanishi, and T. Higashimura, *Biopolymers*, **11**, 399 (1972).
(16) S. Tanaka and A. Nakajima, *Polym. J.*, **2**, 717 (1971).
(17) A. W. Burgess, Y. Paterson, and S. J. Leach, *J. Polym. Sci., Polym. Symp.*, **49**, 75 (1975).
(18) C. H. Bamford, A. Elliott, and W. E. Hanby, "Synthetic Polypeptides", Academic Press, New York, N.Y., 1956, Chapter 3.
(19) F. Conti and P. DeSantis, *Biopolymers*, **10**, 2581 (1971).
(20) W. L. Mattice, *Macromolecules*, **6**, 855 (1973).
(21) K. Titlestad, P. Groth, and J. Dale, *J. Chem. Soc., Chem. Commun.*, 646 (1973).
(22) K. Titlestad, P. Groth, J. Dale, and M. Y. Ali, *J. Chem. Soc., Chem. Commun.*, 346 (1973).
(23) F. T. Wall and J. J. Erpenbeck, *J. Chem. Phys.*, **30**, 634 (1959).
(24) O. Kratky and G. Porod, *Recl. Trav. Chim. Pays-Bas*, **68**, 1106 (1949).
(25) J. J. Hermans and R. Ullman, *Physica*, **18**, 951 (1952).
(26) M. Sisido, Y. Imanishi, and T. Higashimura, *Macromolecules*, **9**, 320 (1976).
(27) K. Nagai, *J. Chem. Phys.*, **38**, 924 (1963).
(28) In the Nagai's series expansion (eq 10), the continuity of the end-to-end distance distribution is tacitly assumed. In real chains, however, the end-to-end distance cannot take values larger than that for maximum extension. The effect of this difference becomes more prominent when higher even moments are considered and this is also a likely reason for the divergence of the distribution function truncated at a higher even moment.

Geometrical Criteria for Formation of Coiled-Coil Structures of Polypeptide Chains[1]

Ken Nishikawa and Harold A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14853.

Received October 16, 1975

ABSTRACT: Crick's general formulas describing a coiled coil are expressed in a different form to combine the parameters of a coiled coil with the backbone dihedral angles of a polypeptide chain, assuming that the bond lengths and bond angles of the chain are fixed. While the existence of a low-energy coiled-coil conformation depends on energetic considerations, these formulas, which pertain to single-stranded structures and, by application of symmetry operations, to multistranded structures, provide the geometrical criteria for the existence of coiled coils. The concept of "the averaged structure of the minor helix", introduced here, makes it possible to relate the shape of the major helix to that of the minor helix. It is shown, in the analysis of a simple model of a single-stranded coiled-coil β structure, that strong geometrical restrictions exist for the formation of coiled-coil structures from a given minor helix conformation of a polypeptide chain; these restrictions are expressed in a general form that is applicable to any coiled-coil of any number of residues in a repeat unit. As an application, the possible existence of a two-stranded coiled-coil antiparallel β structure is considered, both geometrically and energetically, and discussed in relation to the observed twisted β structures in globular proteins. The proposed coiled-coil models of α-helical proteins are also examined briefly.

It has been widely accepted that some of the fibrous proteins are in the form of coiled coils. For example, collagen exists as a three-stranded coiled coil,[2,3] α-keratin probably as a three-stranded α-helical coiled coil,[4,5] and muscle proteins[6] (especially tropomyosin[7]) as two-stranded α-helical coiled coils. Each of these coiled-coil models has been deduced from x-ray diffraction data on fibers rather than on single crystals; however, the coiled-coil model for collagen has been supported by a recent crystal structure determination[8] of the repeating polytripeptide $(Pro\text{-}Pro\text{-}Gly)_{10}$.

Crick has derived a general formula[9] which relates the parameters of the major and minor helices, the minor helix being the small one (e.g., the α helix) and the major helix being the larger one which coils with a larger period than the smaller one. Crick's formula pertains to a coiled-coil polypeptide chain in which the geometry (bond lengths and bond angles) varies continuously and periodically from one residue to another.[10] However, most treatments of polypeptide chains (including the computation of conformational energies) consider the geometry to be rigid and the peptide groups to be in the planar trans form, and use only the dihedral angles for rotation about single bonds as variables. Therefore, in the treatment developed in this paper, we will assume rigid geometry.

We do not deal here (except for a brief example in section III) with the intra- and intermolecular energetic criteria which determine whether a coiled coil exists and, if so, how many chains constitute the coiled-coil structure. However, this question is treated elsewhere[11] for the collagenlike repeating