Helix–Coil Transition in Multicomponent Random Copolypeptides in Water. 1. Theory, and Application to Random Copolymers of (Hydroxypropyl)-L-glutamine, L-Alanine, and Glycine[1]

A. Kidera,[†] M. Mochizuki,[†2a] R. Hasegawa,[†] T. Hayashi,[†] H. Sato,[†] A. Nakajima,*[,†] R. A. Fredrickson,[‡2b] S. P. Powers,[‡] S. Lee,[‡] and H. A. Scheraga*[,‡]

Department of Polymer Chemistry, Kyoto University, Kyoto 606, Japan, and Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.
Received July 28, 1982

ABSTRACT: The nearest-neighbor Ising model theory for the helix–coil transition in a binary random copolymer has been extended to multicomponent random copolymers consisting of any number of components, and the assumption that the statistical weights of a residue depend on the kind of residue but not on those of its neighbors has been examined. Exact and approximate theories of the helix–coil transition have been formulated for multicomponent random copolymers. Illustrative calculations elucidate the dependence of the melting behavior, and of the convergence of the approximate theories to the exact one, on the number of components. The theory is applied to a three-component random copolypeptide, using the host–guest technique, by studying the thermally induced helix–coil transition in water of poly[N^5-(3-hydroxypropyl)-L-glutamine-*co*-L-alanine-*co*-glycine]. Good agreement is obtained between the experimental melting curves of the copolymers and the multicomponent theoretical ones computed directly from the parameters obtained previously by the host–guest method for the corresponding binary random copolymers. Thus, the validity of the above assumption about the statistical weights has been confirmed for the system investigated here.

I. Introduction

It has been shown[3-5] that short-range interactions dominate in determining the conformational states of the individual amino acid residues in globular proteins, and various empirical schemes (summarized by Némethy and Scheraga[6]) have been developed to predict protein conformation, based on a consideration of only short-range interactions. This view also underlies the application of the nearest-neighbor Ising model to data on helix–coil transition curves for host–guest binary random copolymers to determine the helix–coil stability constants for the 20 naturally occurring amino acids in water.[7,8] In this use of the Ising model, it is assumed that the statistical weights depend on the kind of amino acid residue but are independent of the kinds of its neighboring residues.

We examine this hypothesis here by studying the helix–coil transition of water-soluble three-component random copolymers of amino acids and compare the experimental transition curves with those computed with the multicomponent nearest-neighbor Ising model, using the statistical weights obtained from experiments on binary copolymers. In section II we first extend the theory for two-component random copolymers to copolymers with any number of components and examine the dependence of the transition behavior, and of the convergence of approximate theories to the exact one, on the number of components. Then, in sections III and IV we report the experimental results and associated computations for three-component random copolymers consisting of a host [N^5-(3-hydroxypropyl)-L-glutamine] and two guest residues (L-alanine and glycine), the helix–coil stability constants for these two guest residues having been reported earlier.[9,10] Finally, in section V the results are discussed.

II. Theory

Lehman[11] and Lehman and McTague[12] have developed an exact theory for the helix–coil transition in a random-sequence copolymer of two types of units, and Poland and Scheraga[7,13,14] have proposed a series of approximations, the LAPS (Lifson–Allegra–Poland–Scheraga) hierarchy, that converge to the exact results of Lehman and McTague. In order to apply the host–guest technique[7,8] to a random-sequence copolymer consisting of any number (greater than two) of types of units, we extend the Lehman–McTague and Poland–Scheraga treatments.

As in the earlier theories, we use the one-dimensional Ising model, and assume that the statistical weights depend on the type of amino acid but not on those of its neighbors and that the nearest-neighbor 2 × 2 matrix formulation of the partition function[14,15] is applicable. The first assumption is justified by the demonstrated dominance of short-range interactions in determining protein conformation[3-5] and by the applicability of the nearest-neighbor Ising model to binary random copolymers;[7,8,16] the validity of the second assumption is demonstrated in the Appendix. The validity of the first assumption is tested further in this paper.

For a given specific-sequence copolymer, the partition function Q_N is given by

$$Q_N = \mathbf{e}\prod_{i=1}^{N}\mathbf{W}_i\mathbf{e}^{+} \tag{1}$$

where the statistical-weight matrix for the ith residue is

$$\mathbf{W}_i = \begin{bmatrix} s_i & 1 \\ \sigma_i s_i & 1 \end{bmatrix} \tag{2}$$

and

$$\mathbf{e} = [0 \quad 1] \tag{3}$$

$$\mathbf{e}^{+} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{4}$$

N is the degree of polymerization, and s_i and σ_i are the Zimm–Bragg[15] statistical weights of the ith amino acid. For a random-sequence copolymer, we must obtain the ensemble average of Q_N of eq 1 in order to compute the

[†] Kyoto University.
[‡] Cornell University.

0024-9297/83/2216-0162$01.50/0

helix content and other parameters of the helix–coil transition.

A. Exact Theory. Following Lehman and McTague[12] (designated LM), the ensemble average for noninteracting copolymer molecules is given by

$$N^{-1}\langle \ln Q_N \rangle \equiv C_N(1) \tag{5}$$

where $C_N(y)$ is defined by either eq 6 or 9.

For a finite random-sequence chain consisting of m types of amino acid residues, $C_N(1)$ is obtained as a solution of the following linear functional equation, obtained from eq 27 of LM:

$$\frac{N}{N-1}C_N(y) = \sum_{i=1}^{m} p_i\left[\ln (y + s_i) + C_{N-1}\left(\frac{y + \sigma_i s_i}{y + s_i}\right)\right] - C_{N-1}(1) + C_1(y)/(N-1) \qquad \text{for } 0 \le y < \infty \tag{6}$$

where

$$C_1(y) = \sum_{i=1}^{m} p_i \ln (y + \sigma_i s_i) \tag{7}$$

p_i is the fraction of the ith residue in the polymer, and

$$\sum_{i=1}^{m} p_i = 1 \tag{8}$$

The quantity y is a dummy variable used to solve eq 6 and has no physical significance except for $y = 1$ (the value $y = 1$ makes eq LM-19 become the partition function, given by eq LM-14). The additional factor σ_i appears in eq 7, compared to eq LM-28, because we are dealing here with polypeptides whereas LM pertains to polynucleotides, and the end vectors in the matrix formulation of the partition function differ for the two types of polymers.

For a finite chain consisting of an infinite number of components, eq 6 becomes

$$\frac{N}{N-1}C_N(y) = \int_{t_0}^{t_1} p(t)\left\{\ln [y + s(t)] + C_{N-1}\left[\frac{y + \sigma(t)s(t)}{y + s(t)}\right]\right\} dt - C_{N-1}(1) + C_1(y)/(N-1) \tag{9}$$

where

$$C_1(y) = \int_{t_0}^{t_1} p(t) \ln [y + \sigma(t)s(t)]\, dt \tag{10}$$

and the quantities s, σ, and p are continuous functions of the variable t.

$$\int_{t_0}^{t_1} p(t)\, dt = 1 \tag{11}$$

and t_0 and t_1 are finite constants that define the ranges of both σ and s. (While t_0 and t_1, in principle, could be infinite, they are taken here as finite constraints to avoid computational problems in the evaluation of the integral of eq 11; this choice is of no consequence in this problem.)

Equations 6 and 9 are recursion formulas to obtain $C_N(y)$ from $C_{N-1}(y)$, by beginning with $C_1(y)$ as explained below.

For illustrative purposes only, we take all σ_i's to be equal and let N become infinite. Then eq 6 and 9, respectively, become

$$C_\infty(y) = \sum_{i=1}^{m} p_i\left[\ln (y + s_i) + C_\infty\left(\frac{y + \sigma s_i}{y + s_i}\right)\right] - C_\infty(1) \tag{12}$$

and

$$C_\infty(y) = \int_{t_0}^{t_1} p(t)\left\{\ln [y + s(t)] + C_\infty\left[\frac{y + \sigma s(t)}{y + s(t)}\right]\right\} dt - C_\infty(1) \tag{13}$$

The solutions of eq 12 and 13 can be obtained by the functional equation approach of Lehman[11] (designated L). Equation L-83 (for infinite N) is

$$C_\infty(y) = \ln (y + W_e) + \ln [D(y)] \tag{14}$$

where

$$W_e = \frac{\hat{s} - 1}{2} + \left[\left(\frac{\hat{s} - 1}{2}\right)^2 + \hat{s}\sigma\right]^{1/2} \tag{15}$$

and

$$\hat{s} = \prod_{i=1}^{m} (s_i)^{p_i} \tag{16}$$

for a finite number of components, and

$$\hat{s} = \exp\left[\int_{t_0}^{t_1} p(t) \ln s(t)\, dt\right] \tag{17}$$

for an infinite number of components; σ_e (an "effective σ") of eq L-81 has been replaced by σ to obtain eq 15 since the value of σ_e affects only the rate of convergence of $C_\infty(y)$ and not its actual value.

For a finite number of components, $D(y)$ is given by eq L-84, viz.

$$D(y) = [D(1)]^{-1}\prod_{i=1}^{m}\left[G_i(y)D\left(\frac{y + \sigma s_i}{y + s_i}\right)\right]^{p_i} \tag{18}$$

where

$$G_i(y) = \left[y + s_i\frac{W_e + \sigma}{W_e + 1}\right] \Big/ (y + W_e) \tag{19}$$

For an infinite number of components

$$D(y) = \exp\left\{\int_{t_0}^{t_1} p(t) \ln \left[G(y;t)D\left(\frac{y + \sigma s(t)}{y + s(t)}\right)\right] dt - \ln [D(1)]\right\} \tag{20}$$

where

$$G(y;t) = \left[y + s(t)\frac{W_e + \sigma}{W_e + 1}\right] \Big/ (y + W_e) \tag{21}$$

We solved eq 14 by dividing the interval $0 \le y \le 1$ into 51 parts, using eq L-86, viz.

$$y_j = \sigma\left\{\exp\left[\frac{j - 1}{50} \ln (1 + \sigma^{-1})\right] - 1\right\} \tag{22}$$

for $1 \le j \le 51$ (see p 219 of L for this choice of a nonlinear subdivision of the y scale); the interval of y is taken as $0 \le y \le 1$ in order to obtain the value of $C_\infty(1)$. Since the values of $(y_j + \sigma s_i)/(y_j + s_i)$, as well as those of $[y_j + \sigma s(t)]/[y_j + s(t)]$, do not coincide with y_j, a third-order normalized B-spline equation[17,18] was used to calculate $D(y)$; i.e., it is necessary to interpolate among the 51 values of y_j to obtain the value of $D([y_j + \sigma s_i]/[y_j + s_i])$. For this purpose, we start with an initial guess of unity for $D(y)$ and improve its value by iteration [using the B-spline interpolation among the 51 values of $D(y_j)$] until $C_\infty(y)$ of eq 14 converges to the value $C_\infty(1)$. Likewise, for finite N, eq 6 is solved by computing 51 values of $C_1(y_j)$ of eq 7 and

by using the B-spline interpolation to obtain $C_1([y + \sigma_i s_i]/[y + s_i])$; then iteration gives $C_2, C_3, ..., C_N$, from which $C_N(1)$, viz., $N^{-1}\langle \ln Q_N \rangle$, is obtained.

The fraction of helical states, θ_h, and the fraction of ch (coil–helix) borders, θ_{ch}, in an infinite chain containing a finite number of components are obtained by numerical differentiation, with the aid of eq 5, by

$$\theta_h = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{m} \frac{\partial \langle \ln Q_N \rangle}{\partial \ln s_i} \tag{23}$$

$$= \sum_{i=1}^{m} s_i [C_\infty(1;s_i + \Delta s) - C_\infty(1;s_i)]/\Delta s \tag{24}$$

$$\theta_{ch} = \lim_{N\to\infty} \frac{1}{N} \frac{\partial \langle \ln Q_N \rangle}{\partial \ln \sigma} \tag{25}$$

$$= \sigma [C_\infty(1;\sigma + \Delta\sigma) - C_\infty(1;\sigma)]/\Delta\sigma \tag{26}$$

where Δs and $\Delta\sigma$ are sufficiently small constants. Instead of using eq 13 for $C_\infty(y)$ for an infinite chain containing an infinite number of components, we express $C_\infty(y)$ in functional form as

$$C_\infty(y) = \int_{t_0}^{t_1} f[y;s(t)]\, dt \tag{27}$$

Hence, θ_h is defined as

$$\theta_h = \int_{t_0}^{t_1} \left[\frac{\partial f}{\partial \ln s(t)} \right] dt \tag{28}$$

In order to evaluate θ_h from eq 28, we use a variation technique. If δs is the variation of s,

$$\delta s(t) = s(t)\Delta \tag{29}$$

where Δ is a sufficiently small constant, then the integrand in eq 28 becomes

$$\frac{\partial f}{\partial \ln s(t)} = \left[\frac{\partial f}{\partial s(t)} \right] \frac{\delta s(t)}{\Delta} \tag{30}$$

Thus, eq 28 may be written, for computational purposes, as

$$\theta_h = \frac{1}{\Delta}\int_{t_0}^{t_1} \left[\frac{\partial f}{\partial s(t)} \right] \delta s(t)\, dt = \frac{1}{\Delta}\int_{t_0}^{t_1} \delta f\, dt = \frac{\delta C_\infty}{\Delta} \tag{31}$$

where δC_∞ and δf are the variations of C_∞ and f, respectively. The expression for θ_{ch} is the same as eq 25, for uniform σ.

B. Approximate Theory. The LAPS hierarchy of approximations for a two-component system,[7,13,14] which converges in the higher approximations to the exact theory of Lehman and McTague,[12] may be generalized for a chain of N residues containing m components, where $m \geq 2$.

The ensemble average of the partition function may be written as

$$\langle \ln Q_N \rangle \cong \tilde{\mathbf{e}}\tilde{\mathbf{W}}^N\tilde{\mathbf{e}}^+ \tag{32}$$

where the matrixlike quantity[13] $\tilde{\mathbf{W}}$ is given by

$$\tilde{\mathbf{W}} = \begin{bmatrix} \sum_{i=1}^{m} \frac{\alpha_i s_i}{\mathbf{W}_i} & \sum_{i=1}^{m} \frac{\alpha_i}{\mathbf{W}_i} \\ \sum_{i=1}^{m} \frac{\alpha_i \sigma_i s_i}{\mathbf{W}_i} & \sum_{i=1}^{m} \frac{\alpha_i}{\mathbf{W}_i} \end{bmatrix} \tag{33}$$

where $\mathbf{W}_i$ is given by eq 2 and α_i is determined from

$$p_i = \frac{1}{N} \frac{\partial \langle \ln Q_N \rangle}{\partial \ln \alpha_i} \tag{34}$$

Whereas α_i was taken as p_i in the original treatment of Lifson,[19] it is taken here as an adjustable parameter to obtain the experimental value of p_i by the operation of eq 34 (see below). Since the p_i's satisfy eq 8, the α_i's are not all independent; hence, we set $\alpha_m = 1$, as was done[7] for the case of $m = 2$. The end vectorlike quantities[13] $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{e}}^+$ are given by

$$\tilde{\mathbf{e}} = \begin{bmatrix} \frac{0}{=} & \frac{1}{=} \\ \mathbf{e} & \mathbf{e} \end{bmatrix} \tag{35}$$

and

$$\tilde{\mathbf{e}}^+ = \begin{bmatrix} \frac{1}{=} \\ \mathbf{e}^+ \\ \\ \frac{1}{=} \\ \mathbf{e}^+ \end{bmatrix} \tag{36}$$

where $\mathbf{e}$ and $\mathbf{e}^+$ are given by eq 3 and 4, respectively. The procedure for treating $\tilde{\mathbf{e}}\tilde{\mathbf{W}}^N\tilde{\mathbf{e}}^+$ so that matrix multiplication is performed properly is described in ref 13. In essence, $\mathbf{W}_i$ in the denominators of eq 33 is approximated by its respective largest eigenvalue according to the order of the LAPS approximation, i.e.

$$\begin{aligned} \mathbf{W}_i &\to \lambda_i \text{ (LAPS 1, } m_a = 1) \\ \mathbf{W}_i\mathbf{W}_j &\to \lambda_{ij} \text{ (LAPS 2, } m_a = 2) \\ \mathbf{W}_i\mathbf{W}_j\mathbf{W}_k &\to \lambda_{ijk} \text{ (LAPS 3, } m_a = 3) \\ &\vdots \end{aligned} \tag{37}$$

etc., and $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{e}}^+$ are then replaced by $\mathbf{e}$ and $\mathbf{e}^+$, respectively. The Lifson approximation[19] ($\alpha_i = p_i$, $\mathbf{W}_i \to 1$) and the Lifson–Allegra approximation[20] ($\alpha_i = p_i$, $\mathbf{W}_i \to \lambda_i$) are special cases of eq 32.

To determine α_i, we regard eq 34 as a system of m simultaneous nonlinear equations, i.e.

$$\frac{1}{N} \frac{\partial \langle \ln Q_N \rangle}{\partial \ln \alpha} - \mathbf{p} = \mathbf{0} \tag{38}$$

where $\alpha = \{\alpha_i\}$, $\mathbf{p} = \{p_i\}$, and $\mathbf{0} = \{0\}$, and solve them by use of Brent's method,[21,22] which is an improved Newton method for solving such a system of equations.

For infinite N and a finite number of components, θ_h is given by

$$\theta_h = \frac{1}{m_a}\sum_{i=1}^{m} \frac{\partial \ln \lambda_{m_a}}{\partial \ln s_i} \tag{39}$$

where m_a is the order of approximation in the LAPS hierarchy and λ_{m_a} is the largest eigenvalue of $\prod_{i=1}^{m_a} \mathbf{W}_i$ in eq 37. θ_{ch} is given by

$$\theta_{ch} = \frac{1}{m_a}\sum_{i=1}^{m} \frac{\partial \ln \lambda_{m_a}}{\partial \ln \sigma} \tag{40}$$

for uniform σ.

C. Calculation of Other Quantities. For both the exact and m_ath-order approximate theories, the sharpness of the transition, S, the average length of a helical sequence, $\bar{L}$, and the correlation length, $\bar{L}_{tr}$, for the infinite chain are obtained from

$$S = \left(\frac{\partial \theta_h}{\partial T} \right)_{\theta_h = 1/2} \tag{41}$$

$$\bar{L} = \theta_h/\theta_{ch} \tag{42}$$

$$\bar{L}_{tr} = (\bar{L})_{\theta_h = 1/2} \tag{43}$$

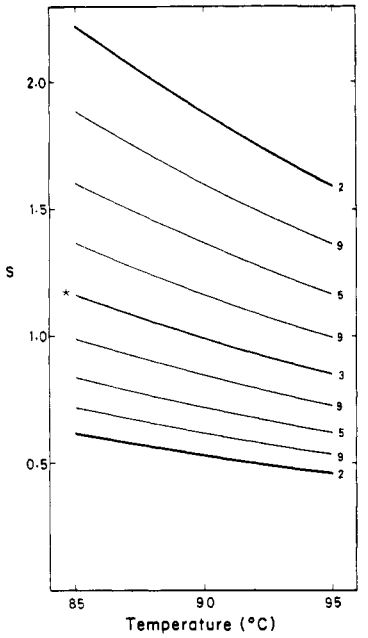

Figure 1. Zimm–Bragg parameters s for each component in systems of two, three, five, and nine components, respectively. The numbers on the curves have the following meaning: two curves (2) pertain to $m = 2$, three curves (2 and 3) to $m = 3$, five curves (2, 3, and 5) to $m = 5$, and nine curves (2, 3, 5, and 9) to $m = 9$. The curve denoted by an asterisk is used in the calculations for the one-component reference system in Figures 2–4.

D. Illustrative Calculations. In order to compare our results with those of previous authors[12,13] who treated two-component systems, we illustrate the calculations by using the same (i.e., polynucleotide) parameters. Hence, we take a temperature-independent value of σ as 9.0×10^{-5} for all components. For $s(T)$, we use the usual van't Hoff relation

$$s = \exp[-(\Delta H/R)(1/T - 1/T_{tr})] \tag{44}$$

where T_{tr} is the transition temperature [the temperature at which $s = 1$ or at which $(\theta_h)_{N\to\infty} = {}^1/_2$]. For the two-component system, previous authors[12,13] used

$$\Delta H_1 = -7600 \text{ cal/mol} \qquad (T_{tr})_1 = 342.5 \text{ K}$$

and (45)

$$\Delta H_2 = -8600 \text{ cal/mol} \qquad (T_{tr})_2 = 383.5 \text{ K}$$

For m components, we assigned these quantities to components 1 and m and divided the intervals between these quantities into m equal parts to obtain the values for components 2 to $m - 1$. The values of m used in the illustrative calculations were 2–9 and ∞. In addition, p_i was taken to be the same for all components, and N was taken as infinity, i.e.

$$p_i = 1/m \qquad 1 \le i \le m$$
$$N \to \infty \tag{46}$$

Figure 1 shows the values of the Zimm–Bragg parameter s for systems of from two to nine components that were used in the calculations. The fraction of helical states, θ_h, the fraction of ch borders, θ_{ch}, the average length of a helical sequence, $\bar{L}$, the sharpness of the transition, S, and the correlation length, $\bar{L}_{tr}$, obtained for the infinite chain by the exact theory are shown in Figures 2–5.

The transition temperature is almost independent of the number of components, within the error of the calculation. Figures 3 and 4, however, show that the conformations of the polymers at the transition temperature differ considerably as the number of components varies. The most

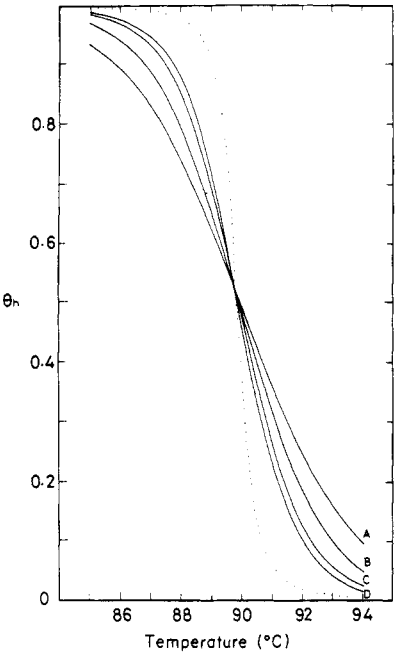

Figure 2. Dependence of the transition curve on the number of components in an infinite chain, calculated by the exact theory: (A) two components; (B) three components; (C) nine components; (D) an infinite number of components. The dotted curve is that for a one-component system having the same transition temperature and value of σ and a value of ΔH implied by the asterisked curve in Figure 1.

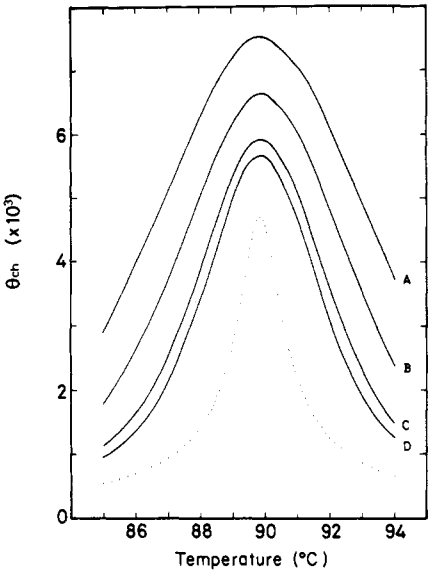

Figure 3. Dependence of θ_{ch} on temperature for the transitions of Figure 2.

characteristic feature is that the sharpness of the transition increases as the number of components increases, and both S and $\bar{L}_{tr}$ converge rapidly with $1/m$ to the corresponding values for a system with an infinite number of components (see Figure 5). In Figure 4, the dependence of $\bar{L}$ $(=\theta_h/\theta_{ch})$ on the number of components inverts at 91 °C; this can also be explained by the increase in the sharpness of the transition with the number of components.[23] Even for $m \to \infty$, however, the values of S and $\bar{L}_{tr}$ (-0.26 deg^{-1} and 87, respectively) differ appreciably from those for a homopolymer with the same transition temperature, same value of σ, and the values of ΔH implied by the asterisked curve of Figure 1 ($S = -0.82$ deg^{-1}, $\bar{L}_{tr} = 106$). Since T_{tr} and ΔH for the one-component system are averages of the two extreme values chosen, the difference between the one-component system and the system of an infinite number

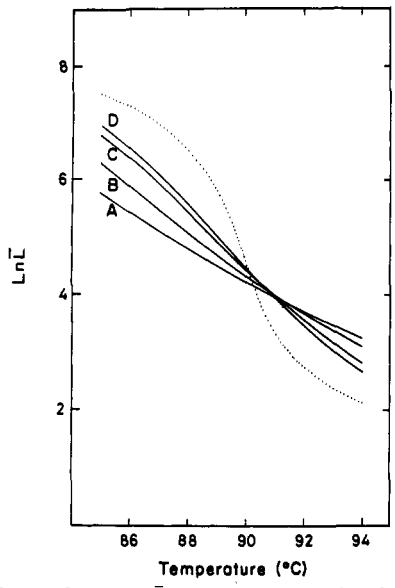

Figure 4. Dependence of $\bar{L}$ on temperature for the transitions of Figure 2.

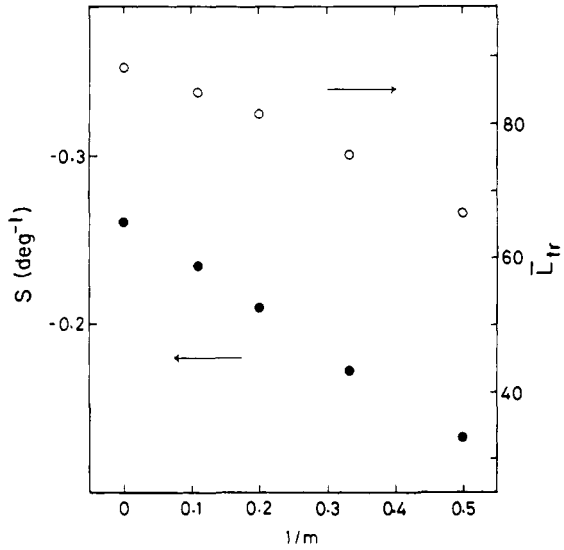

Figure 5. Dependence of S and $\bar{L}_{tr}$ on the number of components m in an infinite chain, calculated by the exact theory.

of components suggests the possibility that fluctuations in the values of s may produce a change in the transition behavior.

Figures 6 and 7 show the transition curves for infinite chains calculated by the approximate theories for the two- and nine-component systems, respectively. Finally, the ratio of the sharpness of the transition of the infinite chain, S_{m_a}/S_∞, where S_{m_a} was obtained by the approximate theory of order m_a and S_∞ was obtained by the exact theory, is plotted against the reciprocal of the order of approximation, m_a, in Figure 8.

From these illustrative calculations, we conclude that the LAPS hierarchy is applicable to a multicomponent system of any number of components and that the convergence of the LAPS hierarchy exhibits little dependence on the number of components, for the values of the parameters chosen here. These values, which correspond to those for polynucleotides, exaggerate the deviation of the lower orders of the LAPS hierarchy from the exact values. As shown in section V-A, the values of the parameters for poly(amino acids) are such that even low order (e.g., m_a = 1 or 2) of the LAPS hierarchy suffice for simulation of the exact theory.

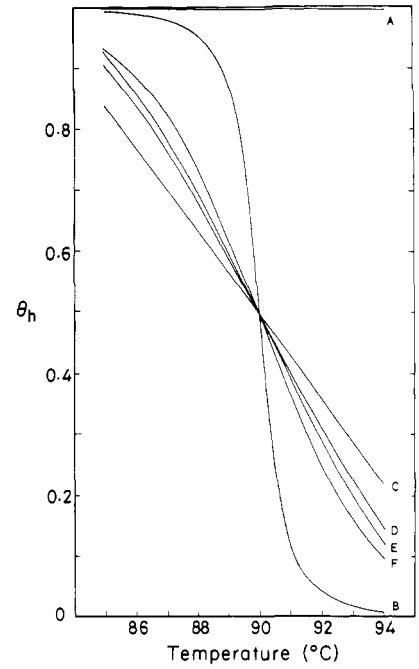

Figure 6. Dependence of the transition curve for an infinite chain of two components, calculated by various theories: (A) Lifson;[19] (B) Lifson–Allegra;[20] (C) LAPS 1; (D) LAPS 2; (E) LAPS 4; (F) exact.

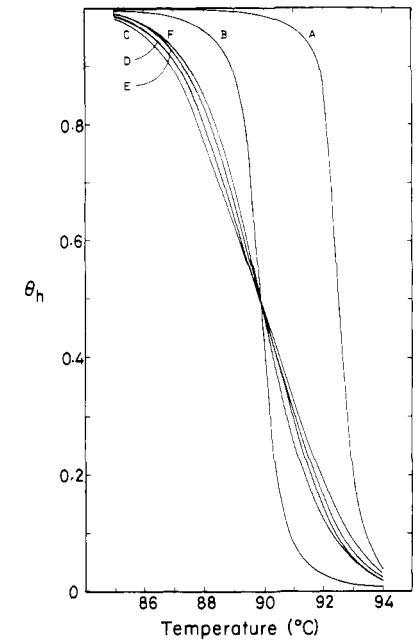

Figure 7. Dependence of the transition curve for an infinite chain of nine components, calculated by the same theories (with the same designations) as in Figure 6.

The main advantage of the LAPS hierarchy over the exact theory is that it requires considerably less computer time to determine the parameters σ and s from experimental data. If, however, as in this paper, we are interested in computing only θ_h from given values of σ and s, then the use of the exact theory is preferable since such a calculation avoids approximations and does not require much computer time. Having illustrated the theory for infinite chains in the foregoing section, we will consider its application to finite chains in sections IV and V.

III. Experimental Section

Water-soluble random copolymers of N^5-(3-hydroxypropyl)-L-glutamine, L-alanine, and glycine, P(HPG:Ala:Gly), were pre-

Table I
Synthesis of Poly[Glu(OBzl),Ala,Gly], Copolymers 1, 2, and 3

copolymer	A/I[a]	time,[b] min	composition, mol % supernatant Glu	supernatant Ala	supernatant Gly	precipitate Glu	precipitate Ala	precipitate Gly	DP[c]
1	40	0	42	48	9.6	nd[d]	nd	nd	3070
		10	46	44	9.5	nd	nd	nd	
		20	52	36	12	45	50	4.1	
		30	50	37	12	46	49	4.2	
		40	38	44	17[e]	47	49	4.1	
		50	38	47	15[e]	49	47	3.8	
		60				47	48	4.1	
		80				48	48	4.0	
		100				46	50	3.6	
		120				48	48	4.0	
2	25	0	nd	nd	nd	nd	nd	nd	1700
		5	44	51	5.0	47	49	4.3	
		10	47	47	5.7	40	56	4.4	
		20	49	41	8.5	42	53	4.4	
		30	46	44	10.4	44	52	4.0	
		40	43	44	12.7	45	51	4.4	
		50	43	44	13.2	46	50	4.2	
		100				47	50	3.8	
		180				43	52	4.3	
3	7.5	0	nd	nd	nd	nd	nd	nd	860
		5	nd	nd	nd	42	52	5.3	
		10	55	36	8.8	45	51	4.4	
		20	54	38	7.6	46	51	4.3	
		30	56	35	8.9	45	50	4.2	
		40	52	37	10.2	46	50	4.2	
		50				45	51	4.4	
		60				44	51	4.1	
		100				45	51	4.2	

[a] Ratio of anhydride to initiator. [b] Reaction was monitored by quenching 0.5 mL of reaction mixture with 5 mL of 0.1 M HCl-ethanol. [c] By viscometry in dichloroacetic acid, using the relationship of Fujita et al.[31] [d] Not determined. [e] The data for the supernatant become unreliable toward the end of the polymerization reaction because of the relatively large contribution from nonreactive soluble byproducts such as diketopiperazines and other small-ring compounds. Hence, inconsistencies between the data for the supernatants and precipitates are not significant toward the end of the reaction. For example, for copolymer 1 at 40 and 60 min, only 3.4% and 1.2%, respectively, of the original NCA's remained.

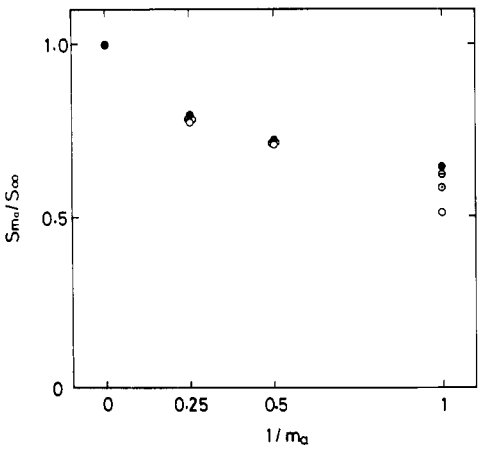

Figure 8. Dependence of the ratio of the sharpness of the transition, S_{m_a}/S_∞, for an infinite chain on the order of approximation, m_a, of the LAPS hierarchy: ○, two components; ⊙, three components; ⊖, five components; ●, nine components.

pared from the *N*-carboxyanhydrides (NCA's) of γ-benzyl L-glutamate, L-alanine, and glycine in dioxane, with sodium methoxide as an initiator. The γ-benzyl blocking groups were subsequently exchanged by reaction with (3-hydroxypropyl)amine. Preliminary calculations, based on earlier experimental data[9,10] that showed that L-alanine is a helix maker and glycine is a helix breaker, indicated that an appropriate ratio of HPG:Ala:Gly would be approximately 45:50:5, and the NCA's were therefore mixed in this ratio. Three independent polymerizations (A–C) were carried out, with use of different anhydride/initiator ratios to obtain different degrees of polymerization. The experimental details are presented below.

A. Materials. L-Glutamic acid was purchased from Mann Research Laboratories (New York), L-alanine and 3-aminopropanol from Aldrich (Milwaukee, WI), and glycine from Sigma (St. Louis, MO). 3-Aminopropanol was dried over Linde molecular sieves (4 Å), distilled under reduced pressure in a nitrogen atmosphere, and stored over molecular sieves. All other reagents and solvents were identical in quality and preparation with those used earlier.[24]

B. Synthesis. 1. *N*-Carboxyanhydrides. L-Alanine *N*-carboxyanhydride was prepared by the action of phosgene on a suspension of the finely ground alanine in dioxane for several hours at 45–55 °C, as described by Platzer et al.[9] Recrystallization from ethyl acetate–hexane gave a product with a melting point of 92–93 °C (lit.[9,24] mp 92 °C).

Glycine *N*-carboxyanhydride was prepared as described by Oya et al.[25] and by Fuller et al.[26] To a suspension of glycine (4.0 g) in 200 mL of dioxane was added 3.9 N $COCl_2$/benzene (50 mL) at 55 °C with stirring. Stirring was continued until the suspension became clear, after which it was purged with nitrogen and evaporated to yield an oil. The oil was crystallized from hot ethyl acetate–*n*-hexane and recrystallized from hot ethyl acetate; yield 1.72 g (32%); mp >100 °C dec [lit.[27] mp >100 °C dec].

γ-Benzyl L-glutamate *N*-carboxyanhydride was prepared by treatment of γ-benzyl L-glutamate (synthesized according to the procedure of Prestige et al.[28]) with phosgene, with use of dioxane as a solvent as described by Blout and Karlson;[29] mp 92–93 °C (lit.[29] mp 93–94 °C).

2. Copoly[Glu(OBzl),Ala,Gly]. Copolymers 1–3. Random ternary copolymers of 45% γ-benzyl L-glutamate, 50% L-alanine, and 5% glycine were synthesized by polymerization of the NCA's in dioxane with sodium methoxide as the initiator.[24]

The NCA's of Glu(OBzl) (1.18 g, 4.5 mmol), Ala (0.576 g, 5 mmol), and Gly (0.05 g, 0.5 mmol) were dissolved in 80 mL of dioxane that had been freshly distilled from sodium. The polymerization was initiated with three different concentrations of

Table II
Characterization of the Fractionated Copolymers

fraction no.	composition, mol % Glu	Ala	Gly	$\bar{M}_w$[a] ($\times 10^{-3}$)	$\bar{M}_z/\bar{M}_w$[a]	$\overline{DP}_w$[a]
original I	47.1	49.0	3.9	45.1	2.05	360
1	45.9	48.4	5.7	84.6	1.08	680
3	47.4	49.0	3.6	75.9	1.10	610
4	47.2	49.1	3.7	67.9	1.13	540
6	47.3	49.1	3.6	63.1	1.12	500
8	45.2	50.2	4.6	40.3	1.16	320
original II	44.0	51.1	4.9	71.1	1.72	590
11	44.3	51.5	4.2	100.8		830
12	44.5	51.8	3.7	80.5	1.54	660
13	44.6	51.5	3.9	69.1		560
14	44.4	51.7	3.9	57.7	1.45	480
15	44.1	52.1	3.8	45.8		390
original III	45.9	49.9	4.2	28.0	1.47	230
21	48.5	47.9	3.6	34.7		280
22	46.5	50.3	3.2	31.8	1.36	260
23	46.3	50.3	3.4	30.4		250
24	46.2	50.5	3.3	29.1	1.35	240
25	45.4	50.9	3.7	23.5		190

[a] By conventional sedimentation equilibrium with an extrapolation to zero concentration. $\bar{v} = 0.737$ was obtained from experiments on fractions 6, 11, and 25 and is essentially independent of molecular weight.

sodium methoxide in benzene as shown in Table I (copolymers 1–3).

The polymerization was monitored for randomness by taking 0.5-mL aliquots every 10 min for the first hour and every 20 min thereafter. The aliquots were quenched in ethanol (5 mL) that was 0.1 M in HCl. Any polymer that precipitated was separated from the supernatant by centrifugation and was washed three times with EtOH. Appropriate volumes of the supernatant were dried and hydrolyzed with 6 N HCl at 110 °C for 20 h. Samples of the precipitated polymer were hydrolyzed in the same manner. The results of amino acid analysis (Table I) indicate that there was no significant difference in the rates of reaction of the different NCA's; i.e., the copolymers are random.

After 2–3 h, the dioxane solution of the polymer was poured into 800 mL of stirred absolute ethanol. At this point, the NCA assay[30] showed that more than 98% of the NCA molecules had reacted. The copolymer was freed of dioxane by stirring in 300 mL of absolute ethanol (twice) and then was dried in vacuo over KOH (average yield of polymers 70–80%).

The chain lengths of the copolymers were determined roughly with the viscosity–molecular weight relationship of Fujita et al.,[31] using dichloroacetic acid (see Table I).

3. Poly[N^5-(3-hydroxypropyl)-L-glutamine-*co*-L-alanine-*co*-glycine]. Copolymers I–III. The exchanged copolymers were obtained by dissolving 1.00 g of copoly[Glu(OBzl),Ala,Gly] in 20 mL of dioxane at 50 °C and adding 20 mL of 3-aminopropanol over a period of 2 days. The reaction was monitored as in Scheule et al.[30] and was terminated when no fewer than 99.5% of the benzyl groups had been exchanged. The reaction mixture was then poured into an excess of 1 N acetic acid in water and dialyzed against water at room temperature until amines could no longer be detected by a ninhydrin test on 0.1 mL of the dialysate.[32] The polymer solution was passed through a Millipore filter, and the crude product was obtained by lyophilization. Average yields of copolymers I–III were about 70% based on the number of moles of poly[Glu(OBzl),Ala,Gly].

C. Fractionation of Copoly(HPG,Ala,Gly). The water-soluble copolymers I–III were fractionated by addition of dioxane to aqueous solutions of the copolymers. After fractionation, the copolymers were dissolved in water, lyophilized, dried in vacuo, and stored in a refrigerator.

D. Determination of Composition. The amino acid compositions of all copolymer fractions were determined as described previously.[33] The data are shown in Table II.

E. Viscosity, ORD, and CD Measurements. Viscosity measurements were carried out as described previously.[33] Optical rotatory dispersion (ORD) and circular dichroism (CD) measurements were made with a JASCO J-20 recording spectropolarimeter using a 0.1-dm path length quartz cell, which was calibrated with sucrose for ORD, with *d*-10-camphorsulfonic acid for CD,[34] and with testosterone for a wavelength check.[35] Solutions were prepared by dissolving the lyophilized material in twice-distilled water and filtering through a Millipore filter HA (0.45-μm pore size). Most transition curves were obtained by both heating and cooling in order to check that the transitions were reversible. Temperature was kept constant to within ±0.2 °C throughout the range of 5–70 °C. A Lorentz-type correction was applied to the data. The parameter b_0 of the Moffitt–Yang equation, obtained from the ORD data in the range of 280–450 nm, was converted to the fraction of helix states, θ_h, by means of the same equation used previously,[9,10,16] viz.

$$\theta_h = -b_0/750 \tag{47}$$

The choice of the values of 0 and −750 for b_0 for the complete coil and α-helix, respectively, was based on measurements of several fractions in dichloroacetic acid and trifluoroethanol, respectively.

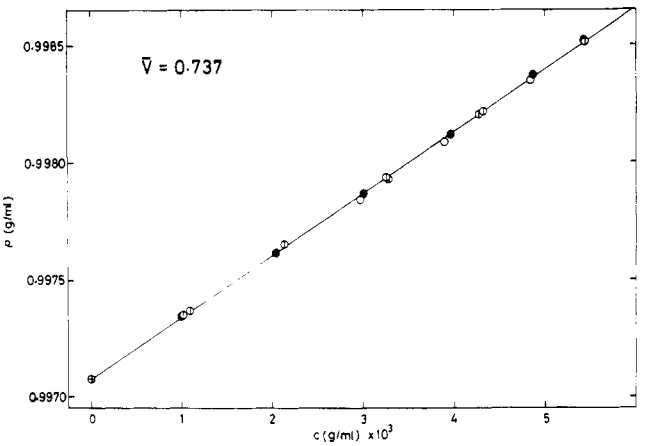

Figure 9. A plot of density (g/mL) vs. concentration (g/mL). ○, ⊕, and ●, fractions 6, 11, and 25, respectively. The slope gives the partial specific volume of the copolymer in water as $\bar{v} = 0.737$ mL/g at 25.0 °C.

F. Determination of Molecular Weight. The molecular weights of fractions from polymers I–III, shown in Table II, were determined in aqueous solution by conventional sedimentation equilibrium, as reported earlier,[16] using interference optics with an MOM ultracentrifuge 3170-b (Hungarian Optical Works, Hungary). The initial concentration was determined from a calibration curve of fringe shift vs. polymer concentration, as in ref 36. The concentration dependence of the weight-average molecular weight, $\bar{M}_w$, was determined for each sample, and the *z*-average molecular weight, $\bar{M}_z$, was computed from the run at the lowest concentration for each fraction. The estimated precision in the values of $\bar{M}_w$ was ±5%.

The partial specific volume, $\bar{v}$, of the fractions, required for the calculation of molecular weights, was determined with a Shibayama digital densitometer SS-D-200 (Shibayama Scientific Co. Ltd., Japan) at 25.0 ± 0.0001 °C. As shown in Figure 9, a value of $\bar{v} = 0.737$ mL/g was obtained for fractions 6, 11, and 25. It can be seen that $\bar{v}$ is essentially independent of molecular weight.

G. Gel Permeation Chromatography. As a check on the molecular weight heterogeneity of the fractions listed in Table II, fraction 6 was examined by gel permeation chromatography (GPC) with a Toyo-Soda high-speed liquid chromatography apparatus (HLC-803D). Two columns (each of 7 mm i.d. and 500 mm length) of TSK-Gel Type G4000SW, C-No. SW-46A-0015, were used. The carrier solvent was 0.05 M aqueous phosphate buffer (pH 7.5). The flow rate was 0.5 mL/min at 25 °C. The GPC unit was calibrated to determine the relation between V_e, the peak elution volume, and M_e, the molecular weight of elution standards, i.e., the effect of molecular weight heterogeneity on bandwidth, by using several fractionated samples of poly[N^5-(3-hydroxypropyl)-L-glutamine] (PHPG). Six fractions of PHPG, with a range of $\bar{M}_w$ of 12 000–270 000 and a range of $\bar{M}_z/\bar{M}_w$ of 1.08–1.15 (determined by sedimentation equilibrium) were employed in the calibration.

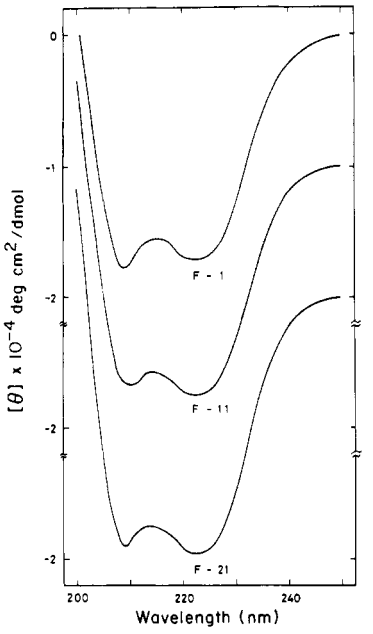

Figure 10. CD data for three fractions in water at 24.0 °C.

The chromatogram from fraction 6 exhibited a single fairly sharp peak, similar to those of PHPG, with a very small tailing on the larger elution-volume side. A molecular weight distribution curve was computed from the experimental GPC curve by assuming a linear relation[37] between V_e and log M_e. The value of M_e at the peak of the chromatogram was 61 000 and the value of $\bar{M}_w/\bar{M}_n$ was 1.28. These may be compared with the values of $\bar{M}_w$ = 63 100 and 1.12, respectively (Table II). These data indicate that the fractions are sufficiently homogeneous for the purposes of this investigation.

H. Concentration Determination. Concentrations were determined by dry weight, and experience indicates that these polymers contain from 0% to 5% moisture after "drying" in vacuo. All calculations reported here were carried out by assuming a moisture content of 5%. They were repeated (values not reported here) by assuming a moisture content of 0%. The only changes were a shift of $\bar{v}$ from 0.737 to 0.753 mL/g (with corresponding changes in the calculated molecular weights) and a shift in the average deviation in Figure 12 from 0.027 to 0.0006. However, the agreement between the experimental and calculated melting curves is essentially the same for both 0% and 5% moisture contents.

IV. Results

The CD spectra at 24.0 °C for representative fractions of the copolymers are shown in Figure 10. The curves are ascribed to a mixture of right-handed α-helix and random coil and show no evidence of β or any other structure.

The experimental values for θ_h for all of the fractions used in the ORD measurements are shown in Figure 11. The transition curves are reversible and exhibited no concentration dependence in the range of 0.06–0.33% w/v. Figure 11 also shows the associated melting curves computed directly from the theory for *finite* chains given in section II; the values of σ and s for HPG, L-alanine, and glycine were obtained from experiments on the corresponding binary copolymers,[9,10,36] and the compositions and the degrees of polymerization are given in Table II. All of the computations in Figure 11 were carried out with the exact theory of Lehman and McTague[12] to avoid introducing approximations,[7] although the LAPS hierarchy yielded results in good agreement with those from the exact theory, as discussed in the next section.

The errors in the experimental points, indicated by the vertical symbols in Figure 11, arise from errors in the determination of b_0 and from those inherent in the ORD measurements: errors in the absolute value of the optical rotation (±3%) and in the wavelength (±0.5 nm). There are also errors in each of the theoretical curves in Figure 11; they arise from the standard deviations in the values

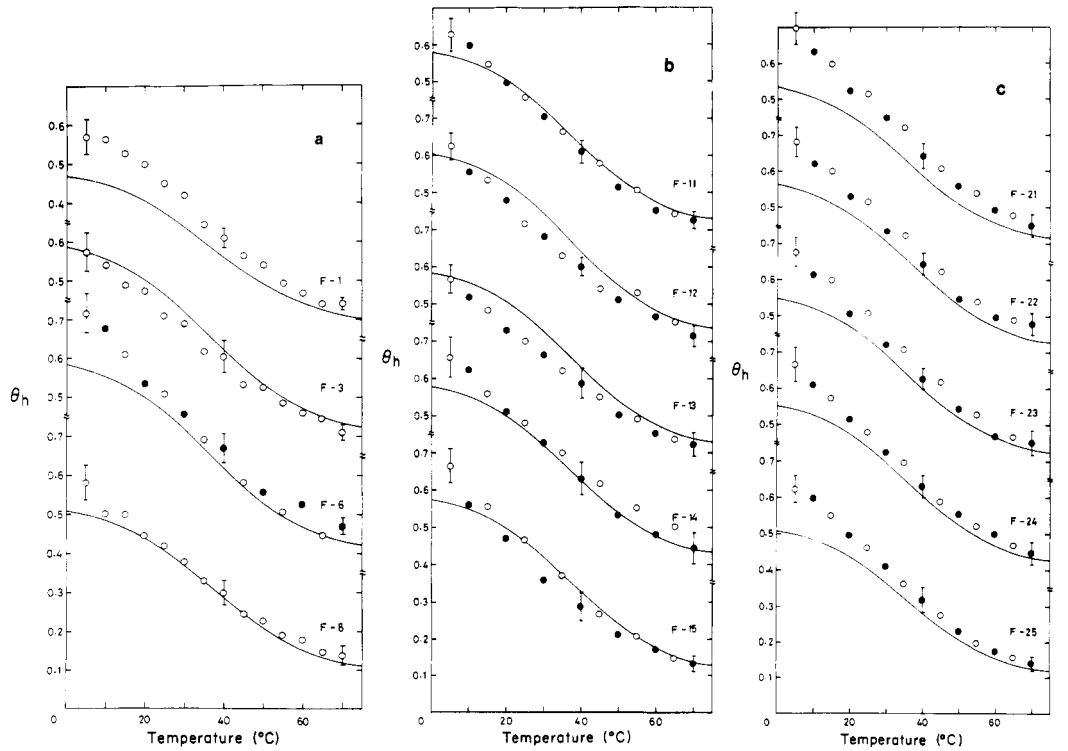

Figure 11. Comparison of the experimental data for θ_h with the calculated melting curves obtained with the exact theory, using previously determined values of σ and s for HPG, L-alanine, and glycine. ○ and ● are experimental points obtained by cooling and heating, respectively. The solid curves are theoretical. The error symbols indicate the errors in the experimental points caused by errors in the determination of b_0 and in measurement of the optical rotation. It should be noted that errors are also associated with the theoretical curves (see text).

Table III
Comparison of the Values of θ_h Calculated with Approximate and Exact Theories[a]

t, °C	0	10	20	30	40	50	60	70
θ_h(Lifson)[b]	0.5792	0.5435	0.4805	0.3872	0.2861	0.2029	0.1468	0.1160
θ_h(L-A)[c]	0.5517	0.5146	0.4534	0.3656	0.2722	0.1953	0.1430	0.1135
θ_h(LAPS 1)[d]	0.4740	0.4559	0.4145	0.3431	0.2613	0.1903	0.1407	0.1121
θ_h(LAPS 2)[d]	0.4739	0.4558	0.4144	0.3430	0.2611	0.1903	0.1407	0.1120
θ_h(LAPS 4)[d]	0.4733	0.4553	0.4141	0.3427	0.2609	0.1901	0.1406	0.1120
θ_h(exact)[e]	0.4699	0.4519	0.4097	0.3404	0.2595	0.1898	0.1400	0.1119

[a] The data for fraction 1 were used in the calculations, viz., N = 680, p(Glu) = 0.459, p(Ala) = 0.484, and p(Gly) = 0.057. [b] Lifson approximation. [c] Lifson-Allegra approximation. [d] LAPS m_a means Lifson-Allegra-Poland-Scheraga hierarchy of of m_ath order. [e] The exact theory of Lehman and McTague.

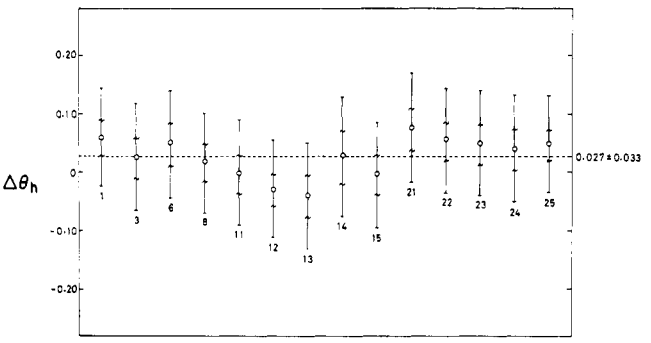

Figure 12. $\Delta\theta_h$ of eq 48 for each fraction. The symbols ⌶ and ⌷ represent the maximum errors in the experimental values only and in both the experimental and theoretical values, respectively. The dashed line indicates the average value of $\Delta\theta_h$ for all the fractions (calculated by assuming 5% moisture in the "dried" samples; this value is reduced to 0.0006 if the moisture content is assumed to be 0%).

of s as determined from the ORD data [±0.08 θ_h unit (0 °C) and ±0.02 θ_h unit (70 °C) at the most], and from the experimental errors in the amino acid analyses and in the determination of $\overline{DP}_w$ [±0.01 θ_h unit (0 °C) and ±0.003 θ_h unit (70 °C) at the most].

V. Discussion

A. Convergence of the Approximate Theory to the Exact One. Here we examine how the approximate theories converge to the exact one for a poly(amino acid) with a finite chain length, using the data for fraction 1, viz., N = 680, p(Glu) = 0.459, p(Ala) = 0.484, and p(Gly) = 0.057. Table III shows the values of θ_h calculated by both the approximate and the exact theories. It can be seen that the deviations of the approximate values from the exact ones are essentially of the same magnitude as those for the binary random copolymers of (hydroxybutyl)-L-glutamine and -glycine[10] and that even low orders of the LAPS hierarchy suffice to simulate the exact theory. Hence, the conclusion in section II, that the convergence of the LAPS hierarchy exhibits little dependence on the number of components, is also valid for poly(amino acids) with finite chain lengths.

B. Validity of the Assumption about Statistical Weights. In Figure 12, the results of Figure 11 were recast by using the average deviation $\Delta\theta_h$ of the experimental value θ_h^{exptl} from the calculated value θ_h^{calcd} defined for each fraction by the following equation:

$$\Delta\theta_h = \sum_{i=1}^{14} [\theta_h^{exptl}(T_i) - \theta_h^{calcd}(T_i)]/14 \qquad (48)$$

where the average is taken over the 14 experimental points for each fraction. The error symbols ⌶ and ⌷ represent the maximum errors in θ_h^{exptl} and those in the sum of θ_h^{exptl} and θ_h^{calcd}, respectively. The dashed line indicates the average value of $\Delta\theta_h$ for all of the fractions. It can be seen that $\Delta\theta_h$ is small compared with the magnitude of the error in this quantity, with no evidence of any trend in the data. These results support the assumption that the statistical weight of the residue is independent of the kind of its neighboring residues. This means that side chain–side chain interactions play a minor role in determining the conformational stabilities of these copolymers.

Platzer et al.[9] drew the same conclusion from their study of binary copolymers of HPG and L-alanine, in which the transition parameters for L-alanine were found to be independent of the L-alanine content up to 50% (their fractions XA-XF) and, further, to be in substantial agreement with those from block copolymers.[38] Since the samples used in this study can be regarded as copolymers that are the same as their fractions XA-XF but with about 4% glycine incorporated, the results of this study, viz., that the effect of the incorporation of the glycine residue can be explained by the theory based on the assumption that the statistical weights are independent of the neighboring residues, enable us to extend the conclusion of Platzer et al. to these three-component systems.

Itou et al.[39] reported that alternating-sequence copoly(amino acids) of γ-benzyl L-glutamate and ε-*N*-[(benzyloxy)carbonyl]-L-lysine in a binary organic solvent exhibited large deviations in their melting behavior from that calculated by a theory based on the same assumption. The transitions investigated were of the *inverse* type, in which the helix content increases with increasing temperature, unlike the behavior in water in which the transition is of the *normal* type. In the inverse transition, interactions with the solvent may play a more important role in determining the conformational stability than in those of the normal type; hence, one might expect cooperative interactive solvent effects of neighboring residues (e.g., involving preferential adsorption[40]) to exist when a copoly(amino acid) is in a *binary* solvent and to be less important when the polymer is in a one-component solvent. These comments may account for the observations of Itou et al.[39] for a copolymer undergoing an inverse transition in a binary solvent and the apparent absence of such effects in the system investigated here, in which the copolymers undergo a normal transition in the one-component solvent, water.

VI. Conclusions

A theory for the helix–coil transition has been formulated for random copolymer systems consisting of any number of components. Illustrative calculations with the exact and approximate theories for the set of parameters used in this study led to the following conclusions: (1) the sharpness of the transition and the correlation length are almost linear functions of the reciprocal of the number of components; (2) the convergence of the LAPS hierarchy exhibits little dependence on the number of components.

The thermally induced helix–coil transition of the three-component random copolymers consisting of HPG, L-alanine, and glycine in water were studied with the aid of ORD measurements. The fact that good agreement was

Table IV
Definition and Statistical Weights of Eight States of a Residue[41]

state	dihedral angles[a]	hydrogen bond on[b] CO	hydrogen bond on[b] NH	statistical weight real	statistical weight dummy
1	c	no	no	1	p_1
2	c	no	yes	$\sigma^{-1/4}s^{1/2}$	p_2
3	c	yes	no	$\sigma^{-1/4}s^{1/2}$	p_3
4	c	yes	yes	$\sigma^{-1/2}s$	p_4
5	h	no	no	$\sigma^{1/2}$	p_5
6	h	no	yes	$\sigma^{1/4}s^{1/2}$	p_6
7	h	yes	no	$\sigma^{1/4}s^{1/2}$	p_7
8	h	yes	yes	s	p_8

[a] The symbol h corresponds to the states of the dihedral angles ϕ and ψ of a residue in which they are confined to the small range in the (ϕ,ψ) space, characteristic of the right-handed α-helix; if the values of ϕ and ψ have no constraint at all, then the residue is in a c state. [b] The existence or absence of a hydrogen bond on the CO and NH groups is designated by yes or no, respectively.

Table V
Three Different Helix Contents for Two- and Nine-Component Random Copolymers with $N = 500$

t, °C		components: two[a]	components: nine[a]	components: nine[b]
87	θ_h	0.830	0.912	0.817
	θ_h^H	0.835	0.908	0.817
	θ_h^A	0.845	0.915	0.837
90	θ_h	0.514	0.511	0.511
	θ_h^H	0.509	0.505	0.500
	θ_h^A	0.523	0.517	0.528
93	θ_h	0.169	0.062	0.191
	θ_h^H	0.152	0.055	0.175
	θ_h^A	0.161	0.060	0.197

[a] $\sigma = 9.0 \times 10^{-5}$. [b] $\sigma = 9.0 \times 10^{-4}$.

obtained between the experimental melting curves and the theoretical ones computed directly from the parameters obtained by the host–guest technique for binary copolymers supports the assumption in the theory that the statistical weights of residues are independent of the kinds of their neighbors.

Acknowledgment. We are indebted to T. W. Thannhauser for carrying out the amino acid analyses.

Appendix

Applicability of Nearest-Neighbor Ising Model to Random Copoly(amino acids) of Many Components. We examine here the validity of our assumption that the 2×2 matrix formulation of the partition function for the nearest-neighbor Ising model (NNIM) is applicable to random copoly(amino acids) of many components. Its validity has already been established for homopoly(amino acids)[15] and for specific-sequence binary copolypeptides.[41] We consider this question by calculating the fraction of helical states for two- and nine-component random copolypeptides both by the NNIM and by the 11×11 matrix model of Gō et al.,[41] which is adopted here as an exact reference one.

A. NNIM. For a multicomponent system, the fraction of helical states is given, according to the Appendix of ref 13, by

$$\theta_h = \frac{1}{N}\sum_{i=1}^{m}\frac{\partial\langle \ln Q_N\rangle}{\partial \ln s_i} = \frac{1}{N}\mathbf{e}[\mathbf{I} \quad \mathbf{0}]\langle\prod_{i=1}^{N}\hat{\mathbf{W}}_i/Q_N\rangle\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}\mathbf{e}^+ \quad \text{(A-1)}$$

where $\mathbf{I}$ and $\mathbf{0}$ are the identity and null matrices, respectively, and Q_N, $\mathbf{e}$, and $\mathbf{e}^+$ are given by eq 1, 3, and 4, respectively. The matrix $\hat{\mathbf{W}}_i$ is given by

$$\hat{\mathbf{W}}_i = \begin{bmatrix}\mathbf{W}_i & \mathbf{W}_i'\\ \mathbf{0} & \mathbf{W}_i\end{bmatrix} \quad \text{(A-2)}$$

where $\mathbf{W}_i$ is given by eq 2, and

$$\mathbf{W}_i' = \sum_{i=1}^{m}\frac{\partial \mathbf{W}_i}{\partial \ln s_i} \quad \text{(A-3)}$$

B. Model of Gō et al.[41] This model requires the 11×11 statistical-weight matrix $\mathbf{U}$, where

$$\mathbf{U} = \begin{bmatrix}
p_1 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & p_5 & 0 & 0 & 0 & 0 & 0 & p_7 & 0 & 0\\
0 & 0 & 0 & p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & p_5 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & p_5 & 0 & 0 & 0 & 0\\
p_2 & p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & p_6 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & p_6 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & p_5 & 0 & 0 & 0 & 0 & p_7 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p_5 & 0 & 0 & p_7\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p_6 & 0 & 0 & p_8
\end{bmatrix} \quad \text{(A-4)}$$

where the statistical weights in eq A-4 are defined in Table IV.

In this model, the fraction of helical states can be defined in two ways, viz., as θ_h^H or as θ_h^A. The first definition, in terms of θ_h^H, is

$$\theta_h^H = \frac{1}{N}\sum_{i=1}^{m}\frac{\partial\langle \ln Z_N\rangle}{\partial \ln s_i} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=3,4,7,8}\frac{\partial\langle \ln Z_N\rangle}{\partial \ln p_j} =$$

$$\frac{1}{N}\sum_{i=1}^{m}\sum_{j=2,4,6,8}\frac{\partial\langle \ln Z_N\rangle}{\partial \ln p_j} = \frac{1}{N}\mathbf{v}[\mathbf{I} \quad \mathbf{0}]\langle\prod_{i=1}^{N}\hat{\mathbf{U}}_i/Z_N\rangle\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}\mathbf{v}^+ \quad \text{(A-5)}$$

where

$$Z_N = \mathbf{v}\prod_{i=1}^{N}\mathbf{U}_i\mathbf{v}^+ \quad \text{(A-6)}$$

$$\mathbf{v} = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0] \quad \text{(A-7)}$$

$$\mathbf{v}^+ = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^t \quad \text{(A-8)}$$

$$\hat{\mathbf{U}}_i = \begin{bmatrix}\mathbf{U}_i & \mathbf{U}_i'\\ \mathbf{0} & \mathbf{U}_i\end{bmatrix} \quad \text{(A-9)}$$

where $\mathbf{U}_i$ is given by eq A-4, and

$$\mathbf{U}_i' = \sum_{i=1}^{m}\sum_{j=3,4,7,8}\frac{\partial \mathbf{U}_i}{\partial \ln p_j}$$

$$\text{or} = \sum_{i=1}^{m}\sum_{j=2,4,6,8}\frac{\partial \mathbf{U}_i}{\partial \ln p_j} \quad \text{(A-10)}$$

Equation A-5 for θ_h^H means that a helical state is defined as one in which either the CO or the NH group of a residue is involved in the formation of an intrachain hydrogen bond. The second definition, in terms of θ_h^A, is that in which the residue is in the h state defined in Table IV, i.e., with its dihedral angles ϕ and ψ confined to a small range

of the (ϕ,ψ) space characteristic of the right-handed α-helix.

$$\theta_h{}^A = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=5,6,7,8}\frac{\partial\langle \ln Z_N\rangle}{\partial \ln p_j}$$

$$= \frac{1}{N}v[\mathbf{I} \quad \mathbf{0}]\langle\prod_{i=1}^{N}\hat{\mathbf{U}}_i/Z_N\rangle\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}v^+ \qquad \text{(A-11)}$$

where $\hat{\mathbf{U}}_i$ is given by eq A-9 and

$$\mathbf{U}_i' = \sum_{i=1}^{m}\sum_{j=5,6,7,8}\frac{\partial \mathbf{U}_i}{\partial \ln p_j} \qquad \text{(A-12)}$$

C. Computations. The ensemble averages of the partition functions, $\langle Q_N\rangle$ and $\langle Z_N\rangle$, were calculated by generating ~2000 random chains by a Monte Carlo method with $N = 500$ and the polynucleotide parameters described in section II. For the value of σ, we used 9×10^{-4}, as well as 9×10^{-5}. The NNIM is a mathematically simple one but is physically unrealistic especially for describing the statistical weight of a short helical sequence because of the arbitrariness in its definition of helix and coil states. Thus, the validity of the NNIM should be assessed by determining how accurately it reproduces θ_h for different values of the average length of a helical sequence, $\bar{L}$, or equivalently, σ.

Table V shows the values of the fraction of helical states for the two- and nine-component systems calculated by the NNIM and by the model of Gō et al.,[41] i.e., by eq A-1 and A-5 and A-11, respectively. The differences between θ_h, $\theta_h{}^H$, and $\theta_h{}^A$ are invariably less than 2% *absolute* even for the nine-component system with $\sigma = 9 \times 10^{-4}$.

Consequently these results demonstrate that the NNIM is indeed a good approximation even when applied to random-sequence copolypeptides of many components.

Registry No. Poly[Glu(OBzl),Ala,Gly], 84303-23-1.

References and Notes

(1) This work was supported by research grants from the Japan Society for the Promotion of Science (MPCR-035), from the National Science Foundation (PCM79-20279 and R-MPC-0043), and from the National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, U.S. Public Health Service (AM-08465).
(2) (a) On leave from Unitika Co. (b) NIH Predoctoral Trainee.
(3) Kotelchuck, D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *61*, 1163; **1969**, *62*, 14.
(4) Kotelchuck, D.; Dygert, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *63*, 615.
(5) Scheraga, H. A. *Pure Appl. Chem.* **1973**, *36*, 1; **1978**, *50*, 315.
(6) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.
(7) von Dreele, P. H.; Poland D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 396.
(8) For the most recent paper in this series, see: Denton, J. B.; Powers, S. P.; Zweifel, B. O.; Scheraga, H. A. *Biopolymers* **1982**, *21*, 51.
(9) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 177.
(10) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 417.
(11) Lehman, G. W. In "Statistical Mechanics"; Bak, T., Ed.; W. A. Benjamin: New York, 1967; p 204.
(12) Lehman, G. W.; McTague, J. P. *J. Chem. Phys.* **1968**, *49*, 3170.
(13) Poland, D.; Scheraga, H. A. *Biopolymers* **1969**, *7*, 887.
(14) Poland, D.; Scheraga, H. A. "Theory of Helix–Coil Transitions in Biopolymers"; Academic Press: New York, 1970; Chapter 8.
(15) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.
(16) von Dreele, P. H.; Lotan, N.; Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 408.
(17) These calculations were carried out by using the FACOM SSL II (Scientific Subroutine Library II) subroutines E12-31-0302 DBIC3 and E11-31-0301 DBIF3. DBIC3 computes the coefficients of interpolating splines expanded by normalized B-splines.[18] DBIF3 computes the indefinite integral or the interpolated value at a given point based on the interpolating spline.
(18) de Boor, C. *J. Approx. Theory* **1972**, *6*, 50.
(19) Lifson, S. *Biopolymers* **1963**, *1*, 25.
(20) Lifson, S.; Allegra, G. *Biopolymers* **1964**, *2*, 65.
(21) Cosnard, M. Y. Cornell University Computer Science Technical Report TR75-248, 1975.
(22) These calculations were carried out by using the FACOM SSL II subroutine C24-11-0101 DNOLBR.
(23) Since we used the same temperature-independent value of σ for all components, the dependence of θ_{ch} on the number of components is less affected by temperature than that of θ_h; i.e., as shown in Figures 2 and 3, the separations of the curves for θ_h varies with temperature, whereas the separations of the curves for θ_{ch} is essentially independent of temperature. Hence, $\bar{L}$ defined as θ_h/θ_{ch}, exhibits a dependence on the number of components similar to that of θ_h rather than to that of θ_{ch}.
(24) Maxfield, F. R.; Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 479.
(25) Oya, M.; Uno, K.; Iwakura, Y. *J. Polym. Sci., Polym. Chem. Ed.* **1970**, *8*, 1851.
(26) Fuller, W. D.; Verlander, M. S.; Goodman, M. *Biopolymers* **1976**, *15*, 1869.
(27) Hamilton, R. D.; Lyman, D. J. *J. Org. Chem.* **1969**, *34*, 243.
(28) Prestige, R. L.; Harding, D. R. K.; Battersby, J. E.; Hancock, W. S. *J. Org. Chem.* **1975**, *40*, 3287.
(29) Blout, E. R.; Karlson, R. H. *J. Am. Chem. Soc.* **1956**, *78*, 941.
(30) Scheule, R. K.; Cardinaux, F.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 23.
(31) Fujita, H.; Teramoto, A.; Yamashita, T.; Okita, K.; Ikeda, S. *Biopolymers* **1966**, *4*, 781.
(32) Ferger, M. F.; Jones, W. C., Jr.; Dyckes, D. F.; du Vigneaud, V. *J. Am. Chem. Soc.* **1972**, *94*, 982.
(33) Fredrickson, R. A.; Chang, M. C.; Powers, S. P.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 625.
(34) DeTar, D. F. *Anal. Chem.* **1969**, *41*, 1406.
(35) Djerassi, C.; Riniker, R.; Riniker, B. *J. Am. Chem. Soc.* **1956**, *78*, 6377.
(36) Van Wart, H. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 266.
(37) Moore, J. C. *J. Polym. Sci., Part A* **1964**, *2*, 835.
(38) Ingwall, R. T.; Scheraga, H. A.; Lotan, N.; Berger, A.; Katchalski, E. *Biopolymers* **1968**, *6*, 331.
(39) Itou, S.; Lee, D. C.; Teramoto, A. *Macromolecules* **1977**, *10*, 1061.
(40) Strazielle, C. In "Light Scattering from Polymer Solutions"; Huglin, M. B., Ed.; Academic Press: New York, 1972; pp 640, 641.
(41) Gō, N.; Lewis, P. N.; Gō, M.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 692.