

Helix-Coil Transition in Multicomponent Random Copolypeptides in Water. 3. Inclusion of Nearest-Neighbor Interactions and Application to Random Copolymers of (Hydroxybutyl)-L-glutamine, L-Alanine, L-Phenylalanine, L-Lysine, and Glycine

J. Wójcik,^{†,‡} A. Kidera,[§] A. R. Leed,[‡] A. Nakajima,^{||} and H. A. Scheraga^{*,‡}

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301, Protein Engineering Research Institute, Furuedai 6-2-3, Suita 565, Japan, and Osaka Institute of Technology, Oomiya, Asahi-ku, Osaka 535, Japan

Received December 1, 1989; Revised Manuscript Received February 15, 1990

ABSTRACT: Water-soluble random copolymers, poly[*N*⁵-(hydroxybutyl)-L-glutamine-co-L-alanine-co-L-phenylalanine-co-L-lysine-co-glycine], were synthesized, fractionated, and characterized, and the thermally induced helix-coil transitions of these five-component copolymers in water were determined by optical rotatory dispersion measurements. The melting behavior of the copolymers was analyzed with the theory reported earlier. To account for the discrepancies between the experimental and theoretical melting curves, the helix-coil transition theory for multicomponent random copolymers was extended by considering the effects of specific nearest-neighbor interactions on the melting curves, i.e., by assuming that the statistical weight of a residue depends not only on the chemical structure of the residue but also on that of its nearest-neighbor residue. By optimization of two independent parameters in the statistical weights, the extended theory accounts for the experimental results. The implications of these statistical weights for the five-component random copolymers are discussed.

I. Introduction

The host-guest technique¹ has been used to study the thermal stability of the α -helix in water. This technique has been applied² to random copolymers containing, in turn, each of the 20 naturally occurring amino acids to obtain their Zimm-Bragg parameters³ σ and s . In those studies, the nearest-neighbor Ising model was used to analyze data for the host-guest binary random copolymers with the assumption that the statistical weights (expressing the conformational preferences) of each residue depend only on its kind but, in first approximation, are independent of the kinds of its neighboring residues. This assumption is supported by the principle that short-range interactions dominate in the determination of the conformational states of the individual residues in a polymer chain.⁴ In the analysis of data from these host-guest copolymers, it is necessary that they be synthesized under conditions leading to a random distribution of their constituent amino acid residues.

In the first paper⁵ of this series, the nearest-neighbor Ising model theory for the helix-coil transition in a binary random copolymer was extended to multicomponent random copolymers consisting of any number of components. Then this model was applied to three-component copolypeptides consisting of (hydroxypropyl)-L-glutamine (HPG), L-alanine, and glycine residues⁵ and (hydroxybutyl)-L-glutamine (HBG), L-lysine, and L-phenylalanine residues,⁶ respectively. In both of the preceding papers,^{5,6} the validity of the assumption mentioned above was confirmed for each of these three-component copolypeptides.

In order to determine whether this assumption is also valid for a more complicated system, we examine here the melting behavior of water-soluble five-component random copolymers. The experimental transition curves are compared with theoretical ones for copolymers consisting of a host, *N*⁵-(4-hydroxybutyl)-L-glutamine (HBG), and four guest residues, L-alanine (Ala), L-phenylalanine (Phe), L-lysine (Lys), and glycine (Gly). Two of the guests are helix formers (Ala and Phe), and two of them are helix breakers (Lys and Gly). As a helix breaker in water, glycine shifts the transition curves to lower temperatures,⁷ but alanine⁸ and phenylalanine⁹ shift them in the opposite direction. Lysine is a rather weak or indifferent helix destabilizer.¹⁰ As previously,¹⁰ the lysine content is limited to 10% to avoid long-range electrostatic repulsions between the charged lysine side chains; it has been shown that, for copolymers of low lysine content, the charged lysine side chains are sufficiently separated from one another so that the helix-coil transition is not influenced by electrostatic repulsions. The water-soluble copolypeptides investigated here contain several different types of residues, with charged and nonpolar side chains, and can be considered as very simple models of a protein molecule.

The synthesis of water-soluble random copolymers of *N*⁵-(4-hydroxybutyl)-L-glutamine, L-alanine, L-phenylalanine, L-lysine, and glycine is described in section II. In section III, we extend the helix-coil transition theory for a multicomponent random copolymer to take into account the dependence of the statistical weights of a residue on the chemical structure of the nearest-neighbor residues. Section IV reports the experimental results and associated computations carried out with the original theory⁵ and with the extended theory described in section III. The results and the assumptions are discussed in section V.

II. Experimental Section

A water-soluble random copolymer of HBG, Ala, Phe, Lys, and Gly was prepared from the *N*-carboxyanhydrides (NCA's) of

* To whom requests for reprints should be addressed.

[†] On leave from the Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland, 1987-1990.

[‡] Cornell University.

[§] Protein Engineering Research Institute.

^{||} Osaka Institute of Technology.

Table I
Synthesis of Poly[Glu(OBzl),Ala,Phe,Lys(Boc),Gly]

time, min	completion, ^a %	composn, mol %									
		supernatant					precipitate				
		Glu	Ala	Phe	Lys	Gly	Glu	Ala	Phe	Lys	Gly
0	0	61.4	16.0	8.0	11.1	3.5	nd ^b	nd	nd	nd	nd
10		64.6	13.3	7.3	11.7	3.1	46.9	33.7	8.6	4.7	6.1
20		64.4	12.4	7.3	12.9	3.1	52.0	28.0	8.7	7.2	4.1
30		63.0	11.9	6.3	15.4	3.3	48.6	32.3	8.6	5.3	5.0
40		64.8	7.4	5.0	19.7	3.1	51.6	30.2	8.4	5.2	4.6
60	85	56.8	7.0	3.6	27.7	4.9	51.0	26.9	8.3	9.2	4.1
80		53.0	8.5	3.6	31.7	8.4	53.7	26.0	8.1	7.9	4.3
100		49.0	10.2	2.7	27.4	10.7	45.1	34.5	9.5	5.2	5.5
120	94	50.3	13.0	4.3	20.0	12.0	54.0	25.0	7.3	8.9	4.2
final polymer, after exchange							56.0	24.2	7.5	8.6	3.7

^a Percentage of NCA's reacted. ^b Not determined.

Table II
Characterization of the Fractionated Copolymers

fraction no.	composn, ^a mol %					$\bar{M}_w \times 10^{-3}$ ^c	\bar{M}_z/\bar{M}_w ^d	\overline{DP}_w
	Glu	Ala	Phe	Lys	Gly			
original I ^b	56.0	24.2	7.5	8.6	3.7			
original II ^b	56.1	20.6	7.8	10.8	4.7	175.0 ^e		
F-6	54.9	23.7	7.3	10.1	4.0	77.8	1.54	510
F-7	54.4	23.2	7.6	10.7	4.1	59.8	1.92	391
F-8	53.7	24.3	7.8	10.0	4.2	52.4	1.76	346
F-9	54.5	23.8	7.7	10.0	4.1	42.8	1.88	281

^a Determined by hydrolysis in 6 N HCl at 150 °C. ^b Original I and original II are two separate synthetic preparations. Original I was used only for the determination of \bar{D} on the unfractionated material. Original II was fractionated and yielded the remaining fractions in this table.

^c This value was obtained by conventional sedimentation equilibrium, with an extrapolation to zero concentration. ^d \bar{M}_z/\bar{M}_w is reported for the sedimentation equilibrium run at the lowest concentration. ^e By viscometry, using the relation of Fujita et al.¹² for polymers in dichloroacetic acid.

γ -benzyl L-glutamate [Glu(OBzl)], L-alanine, L-phenylalanine, glycine, and *N*- ϵ -(*tert*-butoxycarbonyl)-L-lysine [Lys(Boc)] in dioxane by using sodium methoxide as an initiator. The resulting copolymer was converted to the HBG derivative by treatment with 4-amino-1-butanol. The Boc protecting group was then removed with 4 N HCl to give the final product.

A. Materials. Glycine was purchased from Sigma (St. Louis, MO). 4-Amino-1-butanol was purchased from Fluka (Hauptpauze, NY), dried over activated 4-Å molecular sieves, distilled under reduced pressure in a nitrogen atmosphere, and stored over molecular sieves. Glu(OBzl)-NCA, Ala-NCA, Phe-NCA, and Lys-NCA were prepared previously in our laboratory^{5,6} and stored desiccated at -70 °C. Gly-NCA was synthesized as described in an earlier publication.⁵

B. Synthesis. 1. Copoly[Glu(OBzl),Ala,Phe,Lys(Boc),Gly]. The NCA's of Glu(OBzl) (3.620 g, 13.75 mmol), Ala (0.662 g, 5.75 mmol), Phe (0.335 g, 1.75 mmol), Lys(Boc) (0.681 g, 2.5 mmol), and Gly (0.133 g, 1.25 mmol) were dissolved in 200 mL of dioxane that was freshly distilled from sodium. Polymerization was initiated by adding 0.63 mL of 0.987 M sodium methoxide in benzene (0.925 mmol, A/I = 40).

The polymerization was monitored for randomness by removing 0.5-mL aliquots periodically. 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 ethanol. The supernatants and precipitates were dried in vacuo and hydrolyzed with 6 N HCl at 110 °C for 24 h. The results of amino acid analysis are shown in Table I.

After 2 h, the dioxane solution of the polymer was poured into 2 L of stirred absolute ethanol. At this point, the NCA assay¹¹ showed that 94% of the NCA molecules had reacted. The copolymer was freed of dioxane by stirring it in 750 mL of absolute ethanol (twice) and then dried in vacuo over P₂O₅ (yield 3.714 g, 86%).

The degree of polymerization (\overline{DP}) of the polymer was estimated to be approximately 1100 by the viscosity-molecular weight relationship of Fujita et al.¹²

2. Poly[N⁵-4-(hydroxybutyl)-L-glutamine-co-L-alanine-co-L-phenylalanine-co-L-lysine-co-glycine]. Copoly[Glu(O-

Bzl),Ala,Phe,Lys(Boc),Gly] (3.474 g) was dissolved in 55 mL of dioxane and stirred under nitrogen at 50 °C. 4-Amino-1-butanol (55 mL) was added in portions of about 10 mL over 3 days. The extent of the aminolysis was followed by the method of Scheule et al.,¹¹ and the reaction was terminated when 99.6% of the benzyl groups was exchanged (9 days). The solution was poured into a sufficient quantity of cold 7.5 N HCl (275 mL) to give a final concentration of 4 N HCl. The mixture was stirred with ice for 30 min to remove the Boc groups and then diluted with water to a concentration of 1 N HCl. The solution was dialyzed against water at room temperature until amines could no longer be detected by a ninhydrin test of the dialyzate.¹³ Filtration through a 0.45- μ m Millipore filter, followed by lyophilization, yielded 2.826 g (92% based on unexchanged polymer) of copoly(HBG,Ala,Phe,Lys,Gly).

C. Fractionation of Copoly(HBG,Ala,Phe,Lys,Gly). Compared to the copolymers studied previously,^{6,10} copoly(HBG,Ala,Phe,Lys,Gly) was less soluble in methanol. Therefore, the polymer ("original II" of Table II) was dissolved in a large volume of methanol (concentration 0.7% w/v) and fractionated by addition of an ether/methanol solution (90:10, v/v).¹⁴ After fractionation, the copolymers were dissolved in water, lyophilized, and dried in vacuo over P₂O₅.

D. Determination of Composition. In order to determine the amino acid composition, each of the fractions and the unfractionated copolymer were hydrolyzed in 6 N HCl at 150 °C for 1 h. An equimolar mixture of aminobutanol with glutamic acid was used as a calibration standard for HBG. Each fraction was hydrolyzed twice, and each hydrolyzate was analyzed three times. The data in Table II represent averages of two hydrolyses and three analyses each. A Waters Picotag amino acid analysis system was used for the amino acid analysis.

E. Determination of Concentration. The concentrations of the copolymer solutions were determined by UV absorbance of the Phe chromophore at 257 nm. This approach was possible since Phe is the only chromophore in this region. The apparent absorbance was corrected for light scattering.¹⁵ This was done by measuring the apparent absorbance A in the nonabsorbing region (300–400 nm) and plotting log λ vs log A. Then the resulting linear plot was extrapolated into the absorbing region,

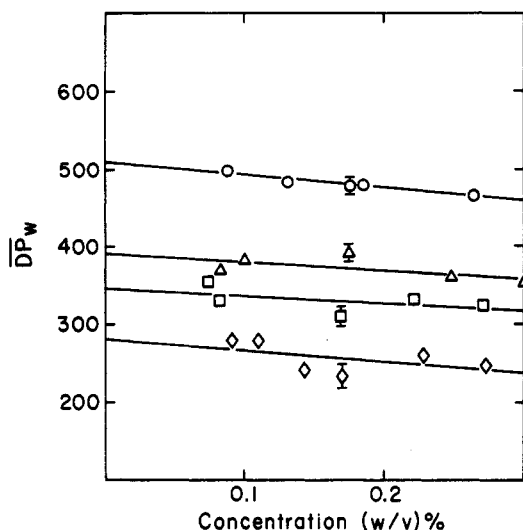


Figure 1. Determination of \overline{DP}_w for the copolymer fractions from the concentration dependence of \overline{DP}_w^{app} : (O) F-6, $\overline{DP}_w = 510$; (Δ) F-7, $\overline{DP}_w = 391$; (\square) F-8, $\overline{DP}_w = 346$; (\diamond) F-9, $\overline{DP}_w = 281$.

and the fraction of the absorbance due to light scattering could thereby be estimated. The method was calibrated by a combination of the results of micro-Kjeldahl nitrogen analysis^{16,17} and amino acid analysis.

F. Viscometry. Measurements of the viscosity of the copolymer in dichloroacetic acid were carried out with a Cannon-Ubbelohde semimicrodilution viscometer at $25.0 \pm 0.1^\circ\text{C}$. When the viscosity-molecular weight relationship of Fujita et al.¹² was used, the approximate molecular weight of the copolymer was determined.

G. Determination of Molecular Weight. The molecular weights of all fractions (shown in Table II) were determined by conventional sedimentation equilibrium using a Beckman Model E analytical ultracentrifuge. The sedimentation equilibrium data were analyzed by the procedure described by Chervenka.¹⁸ All measurements were made in water at pH 11.8 in order to eliminate polyelectrolytic effects arising from the ionization of the ϵ -amino group of the lysine residues. Measurements were carried out at a reduced temperature (5°C) to avoid any degradation of the peptide. At this pH and temperature, no degradation was observed over a period of 72 h. The solutions were prepared by dialyzing a stock solution against a large volume of solvent at pH 11.8 and 5°C under a nitrogen atmosphere. The dialyze was used as the solvent in the molecular weight measurements. The apparent weight-average molecular weight, \overline{M}_w^{app} , was determined for each fraction at several concentrations (from 0.1 to 0.3%), and \overline{M}_w was found by extrapolation of these values to zero concentration (Figure 1). \overline{M}_z was computed from the data corresponding to the lowest concentration for each fraction. The partial specific volume ($\bar{v} = 0.77$; Figure 2), determined for the polymer original I of Table II by a Shibayama digital densitometer SS-D-200 at 25°C , was used for calculating the molecular weights for all the fractions. The accuracy of the molecular weight is $\pm 4\%$.

H. ORD and CD Measurements. Optical rotatory dispersion (ORD) measurements were made on the copolymers in water with a Cary Model 60 spectropolarimeter using a 0.1-dm path length quartz cell, which was calibrated with sucrose. Temperature control was maintained to within $\pm 0.2^\circ\text{C}$ throughout the range 0 – 60°C with a water-jacketed quartz cell. Measurements were made at a concentration of 0.2% (w/v) over the wavelength range of 420–265 nm. The solutions were filtered through 0.45-mm Millipore filters before use. All transition curves were obtained by both heating and cooling in order to check that the transitions were reversible. A Lorenz-type correction was applied to the data. The parameter b_0 of the Moffitt–Yang^{19,20} equation, obtained from the ORD data, was converted to the average fraction of helix states, θ_h , by means of the equation used previously:^{5,6}

$$\theta_h = -b_0/750 \quad (1)$$

The choice of the values of 0 and -750 for b_0 for the complete

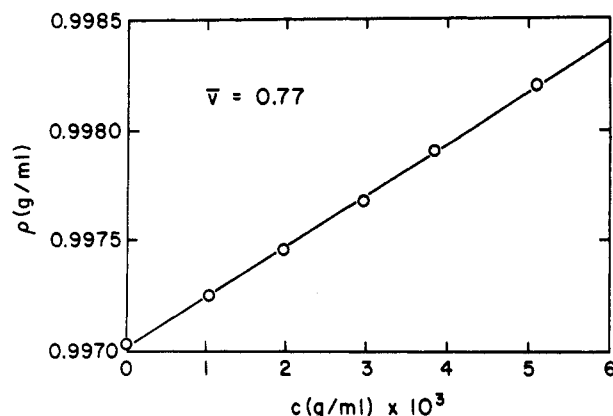


Figure 2. Plot of density vs concentration of the polymer original I of Table II. The slope gives the partial specific volume of the copolymer in water as $\bar{v} = 0.77 \text{ mL/g}$ at 25°C .

coil and α -helix, respectively, was based on the results in previous papers.^{7–10,14}

CD measurements were made with a homemade recording spectropolarimeter (built by modifying a Cary Model 14 spectrophotometer) using a 0.1-dm path length quartz cell at 23°C at a polymer concentration of 0.002% (w/v).

III. Theory

In this section, the Lehman–McTague theory²¹ is extended to the situation in which the statistical weights depend on the chemical structure of the nearest-neighbor residue.

We identify the statistical weights by two indices to designate the chemical nature of the m th and of the $(m-1)$ th amino acid residues; viz.

s_m^{m-1} : the Zimm–Bragg parameter s of the m th residue when the nature of the $(m-1)$ th residue is specified

σ_m^{m-1} : the Zimm–Bragg parameter σ of the m th residue when the nature of the $(m-1)$ th residue is specified

Thus, there are 420 kinds each of s_m^{m-1} and σ_m^{m-1} parameters for the 20 naturally occurring amino acids (including s_1^0 and σ_1^0 for the N-terminal residue, which is preceded by a zeroth or dummy residue). When these statistical weights are used, the partition function of a specific amino acid sequence may be written in the form of the Zimm–Bragg 2×2 matrix formulation, i.e.

$$Q_N = (0 \ 1) \prod_{m=1}^N \begin{pmatrix} s_m^{m-1} & 1 \\ s_m^{m-1} \sigma_m^{m-1} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (2)$$

where N is the degree of polymerization.

Our problem is now to obtain an ensemble average, $\langle \ln Q_N \rangle$, of $\ln Q_N$ for all of the random copolymers in the system. For this purpose, Lehman and McTague²¹ introduced a function

$$C_N(y) = N^{-1} \left\langle \sum_{m=1}^N \ln(y + W_m) \right\rangle \quad 0 < y < \infty \quad (3)$$

where

$$N^{-1} \langle \ln Q_N \rangle = C_N(1) \quad (4)$$

and

$$W_m = U_m/V_m \quad (5)$$

with

$$(U_m \ V_m) = (0 \ 1) \prod_{k=1}^m \begin{pmatrix} s_k^{k-1} & 1 \\ s_k^{k-1} \sigma_k^{k-1} & 1 \end{pmatrix} \quad (6)$$

If

$$P_N(\mathbf{a}_N) = P_N(a_N, \dots, a_1) \quad (7)$$

specified the probability of finding a sequence \mathbf{a}_N in our ensemble, eq 3 can be written more explicitly as

$$C_N(y) = N^{-1} \sum_{\mathbf{a}_m} \sum_{m=1}^N \ln [y + W_m(a_m, \dots, a_1)] P_m \quad (8)$$

where $\sum_{\mathbf{a}_m} = \sum_{a_m} \dots \sum_{a_1}$. Since we have included those only up to the nearest-neighbor interactions, the sequence should have only a nearest-neighbor correlation, i.e.

$$P_m(\mathbf{a}_m) = \prod_{m=2}^N p(a_m | a_{m-1}) P_1(a_1) \quad (9)$$

where $p(a_m | a_{m-1})$ is a conditional probability defined by

$$p(a|b) = P_2(a, b) / P_1(b) \quad (10)$$

where a and b refer to any two consecutive residues.

The quantity $C_N(y)$ is obtained from a recursion formula; i.e., starting from $C_1(y)$, we can calculate $C_2(y)$, $C_3(y)$ from $C_2(y)$, and finally $C_N(y)$ from $C_{N-1}(y)$.

We now introduce the following function (see Appendix) to obtain a recursion formula for $C_N(y)$:

$$C_N(y, a, b) = N^{-1} \sum_{\mathbf{a}} \sum_{m=1}^N \ln (y + W_m) p(a|b) \delta(b, a_m) P_m(\mathbf{a}_m) \quad (11)$$

where $\delta(b, a) = 1$ if $b = a$ and 0 if $b \neq a$; it is noted that

$$C_N(y) = \sum_a \sum_b C_N(y, a, b) \quad (12)$$

Equation 12 is obtained as follows. From eq 11

$$\sum_a \sum_b C_N(y, a, b) = N^{-1} \sum_a \sum_b \sum_{\mathbf{a}} \sum_{m=1}^N \ln (y + W_m) p(a|b) \delta(b, a_m) P_m(\mathbf{a}_m) \quad (13)$$

$$= N^{-1} \sum_b \sum_{\mathbf{a}} \sum_{m=1}^N \ln (y + W_m) \delta(b, a_m) P_m(\mathbf{a}_m) \quad (14)$$

$$\text{because } \sum_a p(a|b) = 1$$

$$= N^{-1} \sum_{\mathbf{a}} \sum_{m=1}^N \ln (y + W_m) P_m(\mathbf{a}_m) \quad (15)$$

$$\text{because } \sum_b \delta(b, a_m) = 1$$

and expression 15 is $C_N(y)$.

Application of eq 12 to eq 21 of Lehman and McTague²¹ yields

$$C_N(y, a, b) = N^{-1} \sum_{\mathbf{a}} \sum_{m=2}^N \{ \ln (y + s_m^{m-1}) + \ln (L_m^{m-1} y + W_{m-1}) - \ln (1 + W_{m-1}) \} p(a|b) \delta(b, a_m) P_m \quad (16)$$

where

$$L_m^{m-1} y = (y + s_m^{m-1} \sigma_m^{m-1}) / (y + s_m^{m-1}) \quad (17)$$

Then, evaluating the averages indicated in eq 16, we obtain

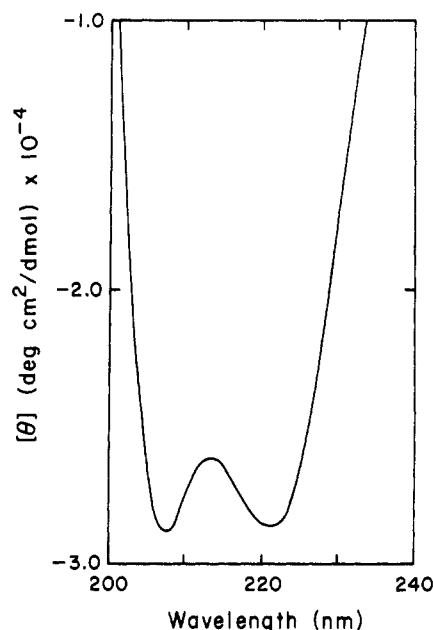


Figure 3. CD data for fraction F-8 in water at 23 °C at neutral pH.

$$[N/(N-1)] C_N(y, a, b) = \sum_c \{ P_3(a, b, c) \ln (y + s_b^0) + p(a|b) [C_{N-1}(L_b^0 y, b, c) - C_{N-1}(1, b, c)] \} + (N-1)^{-1} C_1(y, a, b) \quad (18)$$

where a , b , and c are any three consecutive residues, and

$$C_1(y, a, b) = \ln (y + s_b^0 \sigma_b^0) P_2(a, b) \quad (19)$$

(see Appendix). Equations 18 and 19 are the final forms of the recursion equation, which is the extended version of the Lehman-McTague theory.

Finally, we obtain the partition function $C_N(1)$ from eq 12, with the help of eqs 17–19. Then the average fraction of helix states, θ_h , is given by⁵

$$\theta_h = \sum_a s_a [C_N(1; s_a + \Delta s) - C_N(1; s_a)] / \Delta s \quad (20)$$

where $C_N(1; s_a + \Delta s)$ is the value of $C_N(1)$ of eq 4 obtained by setting the value of s_a equal to $s_a + \Delta s$ and keeping the values of s for the other components fixed. Δs is a sufficiently small quantity; here, a value of 0.001 was used for Δs .

IV. Results

A. Characterization of Copolymers. The fractionation procedure yielded lower molecular weight fractions with increasing amounts of added ether/methanol mixture, as seen by the average compositions and the average degrees of polymerization of the fractionated copolymers. The average compositions of the fractionated copolymers agreed with each other and with that of the unfractionated copolymer (see Table II). This result indicates that no large departures from random copolymerization occurred. The apparent degrees of polymerization, $\overline{DP}_w^{\text{app}}$, were found to be concentration-dependent, and extrapolation to infinite dilution was required to obtain the values of \overline{DP}_w (shown in Figure 1). The $\overline{M}_z/\overline{M}_w$ ratios given in Table II indicate that the fractionation procedure yielded relatively homogeneous material.

B. CD and ORD Data. The CD spectrum for fraction F-8 at 23 °C at neutral pH is shown in Figure 3. Two

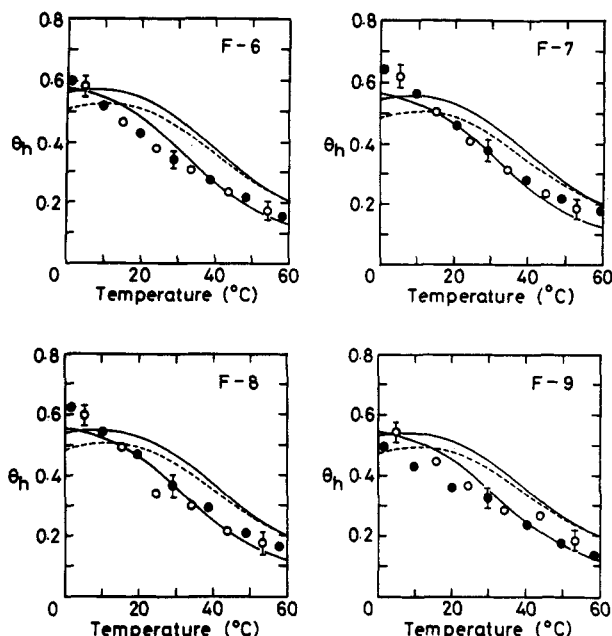


Figure 4. Comparison of the experimental data for θ_h in water at neutral pH with the calculated melting curves obtained with the original theory⁶ using previously determined values of s and σ for HBG, Ala, Phe, Lys, and Gly^{7-10,14} (Table III) (broken curves) and the extended theory described in section III using $\Delta H_{\text{Phe}}^{\text{Lys}}$ and $\Delta S_{\text{Phe}}^{\text{Lys}}$ determined with the three-component copolymers (EFK) (upper solid curves) and $\Delta H_{\text{Phe}}^{\text{Lys}}$, $\Delta S_{\text{Phe}}^{\text{Lys}}$, and $\Delta H_{\text{Ala}}^{\text{Phe}}$, $\Delta S_{\text{Ala}}^{\text{Phe}}$ ($=\Delta H_{\text{Lys}}^{\text{Ala}}$, $\Delta S_{\text{Lys}}^{\text{Ala}} = \Delta H_{\text{Phe}}^{\text{Ala}}$, $\Delta S_{\text{Phe}}^{\text{Ala}} = \Delta H_{\text{Gly}}^{\text{Phe}}$, $\Delta S_{\text{Gly}}^{\text{Phe}} = \Delta H_{\text{Lys}}^{\text{Gly}}$, $\Delta S_{\text{Lys}}^{\text{Gly}}$, respectively, i.e., model 1) (lower solid curves). Open circles: heating. Closed circles: cooling.

minima at 208 and 222 nm correspond closely to the values ascribed by Greenfield and Fasman²² to a mixture of α -helix and random coil. There is no evidence of any other structure.

The thermally induced helix-coil transition was followed spectropolarimetrically in terms of the temperature dependence of b_0 . The experimental values of θ_h are shown in Figure 4. The curves were obtained at 10° intervals. It can be seen that the helix content in water is increased by lowering the temperature. The transition curves are reversible. This reversibility suggests that no aggregation occurs at any stage of the transition. No significant influence of the charged lysine side chain is observed, which is consistent with the data obtained for the corresponding two- and three-component copolymers.^{6,10}

The errors in the experimental points, indicated by the vertical symbols in Figure 4, arise from the errors in the ORD measurements, e.g., in the absolute value of the optical rotation ($\pm 3\%$) and in the wavelength (± 0.5 nm), and from errors in the determination of b_0 . The errors in each of the theoretical curves in Figure 4 arise from the standard deviations in the values of s as determined from the ORD data ($\pm 0.075 \theta_h$ unit and $\pm 0.04 \theta_h$ unit at 0 and 70 °C, respectively) and from the experimental errors in the amino acid analyses and in the M_w measurements ($\pm 0.028 \theta_h$ unit and $\pm 0.01 \theta_h$ unit at 0 and 70 °C, respectively).

C. Computation of Melting Curves. The melting curves were computed both with the original theory⁶ and with the extended theory given in section III. The Zimm-Bragg parameters s and σ for the constituent amino acids are those obtained from the host-guest technique for binary random copolymers,^{7-10,14} which are listed in Table III. The original theory uses these values of s and σ and the degree of polymerization in Table II and gives the melting curves shown by the broken curves in Figure 4.

On the other hand, the extended theory requires s_m^{m-1} and σ_m^{m-1} for the 5² pairs of amino acids. For simplicity, we introduced the following assumptions about the statistical weights

$$(1) \quad s_a^b = s_a \alpha_a^b \quad (21)$$

$$\alpha_a^b = \exp \left[\frac{1}{R} \left(-\frac{\Delta H_a^b}{T} + \Delta S_a^b \right) \right] \quad (22)$$

where s_a is the Zimm-Bragg parameter s for amino acid "a" listed in Table III, and ΔH_a^b and ΔS_a^b are the enthalpy and the entropy of the specific interaction between amino acids a and b . The specific interaction between two consecutive amino acids is given by eq 22 and we assume:

$$(1-1) \quad \alpha_a^b = \alpha_b^a \quad (23)$$

(1-2) No specific interaction between host and guest amino acids.

(1-3) No specific interaction between the same kind of amino acids.

Assumptions (1-2) and (1-3) are supported by the series of experiments on binary random copolymers.²

(1-4) No specific interaction between alanine and glycine. In the three-component random copolymers of HBG, Ala, and Gly (EAG),⁵ the best fitted value of $\alpha_{\text{Gly}}^{\text{Ala}}$ to the experimental data was found to be close to unity.

(1-5) The value of $\alpha_{\text{Lys}}^{\text{Phe}}$ was obtained from the experimental results on the three-component copolymers of HBG, Phe, and Lys (EFK)⁶ by minimizing F

$$F = \sum_i \sum_j [\theta_{h,j}^{\text{exptl}}(T_i) - \theta_{h,j}^{\text{calcd}}(T_i)]^2 \quad (24)$$

in terms of $\Delta H_{\text{Lys}}^{\text{Phe}}$ and $\Delta S_{\text{Lys}}^{\text{Phe}}$, where $\theta_{h,j}^{\text{exptl}}(T_i)$ is the experimental value of the helical fraction at temperature T_i for the j th copolymer and $\theta_{h,j}^{\text{calcd}}(T_i)$ is that of the computed value. The optimized values of $\Delta H_{\text{Lys}}^{\text{Phe}}$ and $\Delta S_{\text{Lys}}^{\text{Phe}}$ are -1.83 kcal/mol and -5.60 cal/deg-mol, respectively, for the transition from coil to helix. The value of the target function F was reduced from 0.039 to 0.018 by including $\alpha_{\text{Lys}}^{\text{Phe}}$ in eq 21.

$$(2) \quad \sigma_a^b = \sigma_a \beta_a^b \quad \text{with } \beta_a^b = 1 \quad (25)$$

where σ_a is the value of σ for amino acid a listed in Table III. Even if $\beta = 10^3$, no significant effect was detected in the five-component copolymer.

The target function for optimizing ΔH and ΔS of eq 22 is the same as F in eq 24; in this case, F is a function of ΔH and ΔS for four pairs of amino acids; i.e.

$$F = F(\Delta H_{\text{Ala}}^{\text{Phe}}, \Delta S_{\text{Ala}}^{\text{Phe}}, \Delta H_{\text{Ala}}^{\text{Lys}}, \Delta S_{\text{Ala}}^{\text{Lys}}, \Delta H_{\text{Gly}}^{\text{Phe}}, \Delta S_{\text{Gly}}^{\text{Phe}}, \Delta H_{\text{Gly}}^{\text{Lys}}, \Delta S_{\text{Gly}}^{\text{Lys}}) \quad (26)$$

Since there are too few experimental data for the five-component copolymers to lead to optimum values of these eight variables, we used five different models (listed in Table IV) for the specific interactions. Examples of the melting curves based on these parameters (model 1 of Table IV) are illustrated by the solid curves in Figure 4 for the five-component copolymers. The melting curves for the other models in Table IV are basically identical with those of model 1 shown in Figure 4. Figure 5 shows

Table III
Values of the Parameters σ and s Used for Calculating the Curves of Figure 4

temp, °C	values of s				
	HBG ^a ($\sigma = 0.000\ 67$)	Ala ^b ($\sigma = 0.000\ 80$)	Phe ^c ($\sigma = 0.001\ 80$)	Lys ^d ($\sigma = 0.000\ 10$)	Gly ^e ($\sigma = 0.000\ 01$)
0	1.045	1.081	1.061	0.857	0.510
10	1.032	1.079	1.084	0.909	0.550
20	1.019	1.071	1.086	0.939	0.591
30	1.008	1.058	1.069	0.947	0.615
40	0.998	1.042	1.047	0.939	0.625
50	0.988	1.025	1.016	0.926	0.625
60	0.979	1.008	1.003	0.911	0.631

^a Computed from the values of ΔH and ΔS given in Table II of ref 14. ^b From ref 8. ^c From ref 9. ^d From ref 10. ^e From ref 7.

Table IV
Optimized Values of ΔH and ΔS in Equation 22 for the Transition from Coil to Helix

model ^a	amino acid pairs ^b	min (F) ^c	ΔH , kcal/mol	ΔS , cal/deg-mol
Three-Component Copolymers (EFK)				
	Phe-Lys	0.018 ^d	-1.83 ^d	-5.60 ^d
Five-Component Copolymers				
1	Ala-Phe	0.082	-0.87	-3.15
	Ala-Lys			
	Gly-Phe			
	Gly-Lys			
2	Ala-Phe	0.083	-0.99	-3.58
	Ala-Lys			
3	Ala-Phe	0.079	-2.23	-8.08
4	Ala-Lys	0.075	-2.01	-7.22
5	Gly-Phe	0.185	-1.07	-4.61
	Gly-Lys			

^a These models optimize only two values of ΔH and ΔS , assuming that all the specific interactions listed in the second column are identical. ΔH_{Lys}^{Phe} and ΔS_{Lys}^{Phe} for the five-component copolymers were taken as those determined for the three-component copolymers.⁶ The amino acid pairs whose specific interactions are considered in the optimization. ΔH and ΔS for the amino acid pairs not listed here are taken as zero during the optimization, except for ΔH_{Lys}^{Phe} and ΔS_{Lys}^{Phe} (see footnote a). ^c min(F) is the value of the target function (eq 24) after optimization. The values of F before optimization were 0.039 and 0.321 for the three- and five-component copolymers, respectively. ^d These values were determined for the three-component copolymers (EFK).⁶

the melting curves for the three-component copolymers (EFK)²³ based on the optimized values of ΔH_{Lys}^{Phe} and ΔS_{Lys}^{Phe} .

V. Discussion

The five-component copolymers investigated here, which are much more complicated than the three-component systems EGA⁵ and EFK,⁶ were designed to serve as simple models of protein molecules containing both helix-forming (Ala and Phe) and helix-breaking residues (Gly and Lys) and both charged (Lys) and hydrophobic side chains (Phe). This study was intended to determine whether the short-range interaction model of the original theory⁵ is valid for a system as complicated as the five-component copolymers. The result is shown clearly in Figure 4. The experimental melting curves of the five-component copolymers depart significantly from the theoretical curves computed by the original theory. This implies that the five-component copolymers are too complicated to be treated by the original helix-coil transition theory using statistical weights that do not depend on the kind of neighboring residues. When we recast the experimental results of the three-component copolymers (EFK)⁶ in Figure 5 in terms of the extended theory, we notice that the theoretical melting curves (based on the original theory) underestimate the experimental data at low temperature. This also indicates that even the

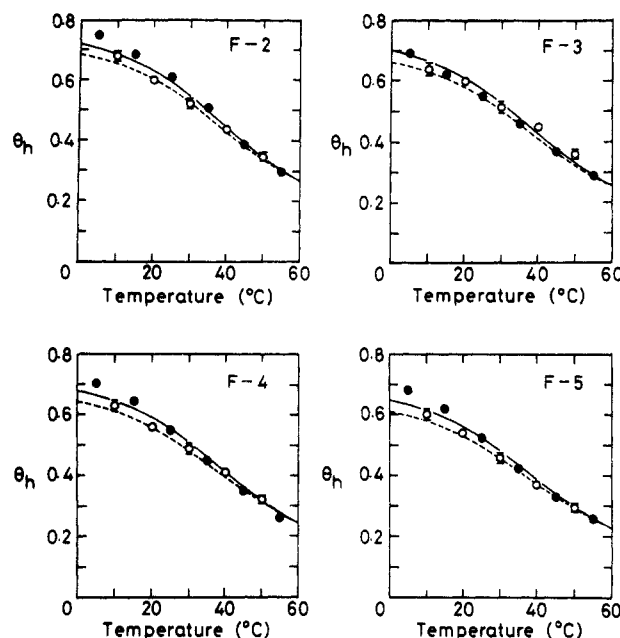


Figure 5. Same as Figure 4 but for the three-component copolymers EFK. The experimental data are those of Figure 3 of ref 6.

EFK copolymers have interactions that cannot be accounted for by the original theory.

The values of s and σ , determined by the host-guest technique involving binary random copolymers,² are intrinsic values that reflect the interaction only *within* a given residue, i.e., between a side chain and its own backbone. When these parameters are applied to the treatment of specific-sequence copolymers, it is necessary to modify these intrinsic values by inclusion of specific medium and long-range interactions.^{23,24} To interpret the experimental melting curves of the random copolymers, we use the extended helix-coil transition theory outlined in section III. As seen in Figures 4 and 5, the agreement with the experimental data was improved significantly by the extended theory with the optimized values of ΔH and ΔS for the specific interactions (eq 22) listed in Table IV. In the case of the five-component copolymers, even after adopting the assumptions described above, there could still be eight variables involved in the optimization [ΔH and ΔS for four different specific interactions (Ala-Phe, Ala-Lys, Gly-Phe, Gly-Lys)], which is too large a number to be determined from the four melting curves in Figure 4. Taking this into account, we checked the five model combinations listed in Table IV for the specific interactions; all of these contain only two variables, ΔH and ΔS , which were assumed to be common to all the specific interactions considered in the optimization.²⁵

The values of the target function for models 1-4 are the same within 0.01, which is much less than the range of the

experimental error (~ 0.03). In these copolymers, the percentages of Phe, Lys, and Gly are much smaller than those of Glu and Ala. Consequently, the probabilities of occurrence of Gly-Phe and Gly-Lys pairs are too small to enable model 5 to account for the experimental data. Further, in random copolymers having the composition of Table II, it is possible to calculate the frequencies of occurrence of different pairs of guest residues. If these frequencies are multiplied by the optimized values of ΔH and ΔS , respectively, the resulting products are essentially the same for models 1-4. These quantities can be regarded as a rough estimate of the total enthalpy and entropy of the specific interactions in a polymer chain. The constancy of these quantities among models 1-4 indicates that each value of ΔH and ΔS is determined mainly by the fraction of amino acid and is almost independent of the kind of amino acid. Hence, it is concluded that the optimization cannot determine which model is the best one to account for the experimental data.

Although the details of the specific interactions are not obvious, the data in Table IV may reflect the nature of these interactions. If the five-component copolymers involved specific interaction between only Phe and Lys, which was found in the three-component copolymers (EFK),⁶ their melting curves would be largely overestimated at high temperature, as shown in Figure 4, because of a strong helix-forming interaction between Phe and Lys as given in the parameters of Table IV. These melting curves indicate that the other specific interactions have a helix-breaking nature at high temperature. From the parameters in Table IV, it is noted that there is a common trend in ΔH and ΔS , i.e., a rather small negative ΔH and a large negative ΔS . As demonstrated earlier by Poland and Scheraga,²⁶ interaction within the coil form can also influence the melting curves. These results reflect possible interactions in both the helix and coil forms. Presumably, these are mainly hydrophobic interactions in the coil state between hydrocarbon groups in the side chains of the guest residues, which stabilize the coil state at high temperature (a small enthalpic contribution $\Delta H \leq 0$ and a large entropy gain in the coil state $\Delta S < 0$).

VI. Conclusion

The melting curves of the five-component random copolymers depart significantly from the theoretical curves computed by the original helix-coil transition theory using statistical weights that do not depend on the kind of neighboring residues. The departures of the computed melting curves from the experimental ones for the five-component random copolymers can be accounted for by modifying the intrinsic values of s and σ (determined from binary random copolymers) by inclusion of nearest-neighbor interactions in the helical and/or coil forms.²⁶ Thus, the statistical weight of a residue depends to some extent on the chemical nature of its neighboring residues.

Acknowledgment. This work was supported by research grants from the National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases, National Institutes of Health, U.S. Public Health Service (DK-08465), from the National Science Foundation (R-MPC-0043), and from the Japan Society for the Promotion of Science (MPCR-035). We thank T. W. Thannhauser for calibrating the UV method for determining polymer concentration, R. W. Sherwood for the amino acid analyses, and V. G. Davenport for help with the sedimentation equilibrium measurements.

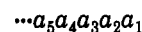
Appendix. Derivation of Equations 11, 18, and 19

Equation 11 is obtained by analogy with the original Lehman and McTague theory for nearest-neighbor correlated sequences.²¹ They defined the recursive equation

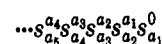
$$C_N(y, a) = N^{-1} \sum_{m=1}^N \langle \ln [y + W_m(a_m, a_{m-1}, \dots, a_1)] p(a|a_m) \rangle \quad (A1)$$

to obtain the partition function. Here "a" stands for the component amino acid and $\langle \rangle$ means the ensemble average.

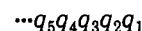
In our system, because the Zimm-Brugg parameter for the m th residue also depends on the nature of the $(m-1)$ th residue, eq A1 cannot be used without any modification. Let us consider an n -component copolymer of the sequence



The corresponding Zimm-Brugg parameters s are



It is possible to reassign this sequence as



where q_m represents two consecutive amino acid residues as

$$q_1 = a_1 \quad (\text{an amino acid at the N-terminus})$$

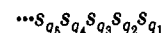
$$q_2 = a_2 a_1$$

$$q_3 = a_3 a_2$$

:

:

Now the Zimm-Brugg parameters s are



where

$$s_{a_1}^0 = s_{q_1}, s_{a_2}^0 = s_{q_2}, \dots$$

This representation is simply the original model for an n^2 -component copolymer in which the Zimm-Brugg parameters do not depend on the neighboring residue, but here a constraint is introduced, in which account must be taken of the sequence $q_1 q_2 q_3 \dots$. This constraint can be expressed explicitly by the probability of the sequence distribution

$$\left[\prod_{m=2}^N p(a_m | a_{m-1}) \right] P_1(a_1) = \left[\prod_{m=2}^N p(q_m | q_{m-1}) \delta(t_m, s_{m-1}) \right] P_1(q_1) \quad (A2)$$

where $q_m = s_m t_m$ (two consecutive residues)

$$\delta(t, s) = 1 \quad \text{when } t = s \\ = 0 \quad \text{when } t \neq s$$

It should be noted in eq A2 that we can apply eq A1 to our system by replacing the conditional probability p by $p\delta$. Therefore, we define eq 11

$$C_N(y, a, b) = N^{-1} \sum_a \sum_{m=1}^N \ln (y + W_m) p(a|b) \delta(b, a_m) P_m(a_m) \quad (11)$$

Using eq 11, it is easy to obtain the recursion equation for $C_N(y, a, b)$, i.e., to derive eq 18 from eq 17. The first term on the right-hand side of eq 17 is

$$\sum_{m=2}^N \sum_{a_m} p(a|b) \delta(b, a_m) \ln(\gamma + s_m^{m-1}) P_m =$$

$$\sum_{m=2}^N \sum_{a_{m-1}} p(a|b) p(b|a_{m-1}) \ln(\gamma + s_b^{m-1}) P_1(a_{m-1}) =$$

$$(N-1) \sum_c p(a|b) p(b|c) \ln(\gamma + s_b^c) P_1(c)$$

The second term is

$$\sum_{m=2}^N \sum_{a_m} p(a|b) \delta(b, a_m) \ln(L_m^{m-1} \gamma + W_m) P_m =$$

$$p(a|b) \sum_{m=2}^N \sum_c \sum_{a_{m-1}} p(b|c) \delta(c, a_{m-1}) \ln(L_b^c \gamma + W_{m-1}) P_{m-1} =$$

$$p(a|b) \sum_{m=1}^{N-1} \sum_c \sum_{a_m} p(b|c) \delta(c, a_m) \ln(L_b^c \gamma + W_m) P_m =$$

$$(N-1) p(a|b) \sum_c C_{N-1}(\gamma, b, c)$$

The third term can be written in the same way as the second term. Finally, from the definition, the last term is simply

$$\sum_{a_1} p(a|b) \delta(b, a_1) \ln(\gamma + W_1) P_1(a_1) = C_1(\gamma, a, b)$$

Equation 19 is obtained by letting $N = 1$ in eq 11, viz.

$$C_1(\gamma, a, b) = \sum_{a_1} \ln(\gamma + W_1) p(a|b) \delta(b, a) P_1(b)$$

$$= \ln(\gamma + W_1) p(a|b) P_1(b)$$

Since $W_1 = s_1^0 s_1^0$ and $p(a|b) P_1(b) = P_2(a, b)$, eq 19 results.

References and Notes

- (1) von Dreele, P. H.; Poland, D.; Scheraga, H. A. *Macromolecules* 1971, 4, 396.
- (2) Wójcik, J.; Altmann, K. H.; Scheraga, H. A. *Biopolymers*, in press.
- (3) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* 1959, 31, 526.
- (4) Scheraga, H. A. *Pure Appl. Chem.* 1973, 36, 1; 1978, 50, 315.
- (5) Kidera, A.; Mochizuki, M.; Hasegawa, R.; Hayashi, T.; Sato, H.; Nakajima, A.; Fredrickson, R. A.; Powers, S. P.; Lee, S.; Scheraga, H. A. *Macromolecules* 1983, 16, 162.
- (6) Miki, T.; Kidera, A.; Oka, M.; Hayashi, T.; Nakajima, A.; Meinwald, Y. C.; Thannhauser, T. W.; Scheraga, H. A. *Macromolecules* 1985, 18, 1069.
- (7) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* 1971, 4, 417.
- (8) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* 1972, 5, 177.
- (9) Van Wart, H. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* 1973, 6, 266.
- (10) Dygert, M. K.; Taylor, G. T.; Cardinaux, F.; Scheraga, H. A. *Macromolecules* 1976, 9, 794.
- (11) Scheule, R. K.; Cardinaux, F.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* 1976, 9, 23.
- (12) Fujita, H.; Teramoto, A.; Yamashita, T.; Okita, K.; Ikeda, S. *Biopolymers* 1966, 4, 781.
- (13) Ferger, M. F.; Jones, W. C.; Dyckes, D. F.; Du Vigneaud, V. J. *Am. Chem. Soc.* 1977, 94, 982.
- (14) von Dreele, P. H.; Lotan, N.; Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* 1971, 4, 408.
- (15) Leach, S. J.; Scheraga, H. A. *J. Am. Chem. Soc.* 1960, 82, 4790.
- (16) Lang, C. A. *Anal. Chem.* 1958, 30, 1692.
- (17) Noel, R. J.; Hambleton, L. G. *J. Assoc. Off. Anal. Chem.* 1976, 59, 134; *Chem. Abstr.* 1976, 84, 149347.
- (18) Chervenka, C. H. *A Manual of Methods for the Analytical Ultracentrifuge*; Beckman Instruments: Palo Alto, CA, 1969; pp 47-49.
- (19) Moffitt, W.; Yang, J. T. *Proc. Natl. Acad. Sci. U.S.A.* 1956, 42, 596.
- (20) Urnes, P. J.; Doty, P. *Adv. Protein Chem.* 1961, 16, 401.
- (21) Lehman, G. W.; McTague, J. P. *J. Chem. Phys.* 1968, 49, 3170.
- (22) Greenfield, N.; Fasman, G. D. *Biochemistry* 1969, 8, 4108.
- (23) Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* 1985, 82, 5585.
- (24) Vásquez, M.; Scheraga, H. A. *Biopolymers* 1988, 27, 41.
- (25) Only a small improvement was obtained in the optimization when eight different values of ΔH and ΔS were used for the four specific interactions; the value of the target function was 0.070.
- (26) Poland, D.; Scheraga, H. A. *Biopolymers* 1965, 3, 283.

A Reptation Model for Polymer Dissolution

Michael F. Herman* and S. F. Edwards

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, U.K.

Received May 15, 1989; Revised Manuscript Received January 30, 1990

ABSTRACT: A model for the dissolution of a simple linear polymer in a solvent is presented. When the pure polymer and solvent are brought into contact, the solvent penetrates the polymer, causing it to swell. This swelling induces a nonrandom distribution of orientations for the polymer chains. The contribution to the free energy and chemical potentials of polymer and solvent due to this nonrandom distribution of orientations is evaluated in closed form from reptation theory. When this contribution is sufficiently large compared with the ordinary mixing terms, the system undergoes a phase separation into a gel-like concentrated solution phase and a dilute solution phase. The presence of this gel-like phase can significantly slow the dissolution process. Simple model calculations are present to illustrate the expected behavior of solvent/polymer systems for which the magnitude of the orientational contribution to the free energy is sufficiently large to support a phase boundary between concentrated and dilute solution regions.

I. Introduction

The dissolution of a polymer in a solvent is a process of considerable technological importance. If a specific

polymer has a desirable property in solution but the kinetics of dissolution are prohibitively slow, then the slow rate of dissolution may limit its usefulness. By understanding the dissolution process and the underlying physical parameters that affect the nature and rate of this process, it may be possible to alter the polymer/solvent

* Permanent address: Department of Chemistry and the Quantum Theory Group, Tulane University, New Orleans, LA 70118.