absence of any effect of this weak acid on the polymer conformation.

In this connection it should be recalled that $HClO_4$ also causes a progressive sharpening of the nmr peaks of the side chain groups of PBLH in TFE.[1] This fact leads to the conclusion that the side chain groups are also quite free in their motion when the polymer is in the conformation of PBLH II.

More detailed studies of the exact structure of PBLH II by X-ray diffraction methods and 220-MHz nmr spectroscopy are in progress in our laboratory and will be reported elsewhere.

Acknowledgment. The authors thanks Professor E. Scoffone and Professor M. Mammi for the suggestions and discussions during this work, and Professor V. Crescenzi and coworkers for the critical reading of the manuscript. The work has been carried out with the financial support of the Institute of Macromolecular Chemistry CNR of Milan, Italy.

Helix–Coil Stability Constants for the Naturally Occurring Amino Acids in Water. I. Properties of Copolymers and Approximate Theories[1]

P. H. von Dreele,[2] D. Poland, and H. A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14850.

Received March 11, 1971

ABSTRACT: Since it is necessary to have the Zimm-Bragg parameters σ and s of all naturally occurring amino acids (in water) for studies of protein conformation, and since these cannot be obtained from studies of the helix–coil transition in homopolymers, because of experimental difficulties, a technique has been developed to circumvent these problems. It involves the study of the thermally induced transition curves for random copolymers of "guest" amino acid residues in a water-soluble "host" poly(amino acid). The data may be interpreted with the aid of suitable theories for the helix–coil transition in random copolymers to obtain σ and s for the "guest" residues. While exact theories are available (for the one-dimensional nearest-neighbor Ising model), the computer costs involved in the computations are prohibitively high. Therefore, resort is had to approximate theories. It is shown here that, for the usual ranges of parameters found for poly(amino acids), one of the two lowest order approximations (corresponding to earlier treatments by Lifson and Allegra) is completely adequate, *i.e.*, gives results essentially identical with the exact result. In essence, the low-order approximations hold if σ and s for the two constituents of the copolymer do not differ appreciably from each other. If they do, then higher order approximations (which become exact in highest order) are required. Calculations are also reported for both regular-sequence and random copolymers in order to demonstrate how the amino acid sequence and composition of the copolymer, and also the values of σ and s of the constituents, affect the transition temperature, the breadth of the transition, and the most probable length of a helical sequence in the copolymers. It is shown that small departures from a random sequence do not affect the melting curve significantly.

Since near-neighbor interactions play a very important role in determining the conformation of an amino acid residue in a protein,[3,4] it is desirable to have a quantitative measure of the relative stabilities of the α-helix and random-coil conformations of each naturally occurring amino acid in water. Such data would then provide a basis to compute the propensity for any given sequence of amino acids in a protein to adopt the α-helical conformation.[5] In principle, these data can be obtained by measuring the fraction of α helix for each homopoly(amino acid) as a function of chain length and temperature, and interpreting the experimental results, for example, by the Zimm–Bragg[6] theory. For such an approach to be experimentally feasible, the homopoly(amino acid) must be water soluble, α helical, and capable of being melted in the temperature range between 0 and 100°C. Unfortunately, none of the homopolymers of the naturally occurring amino acids satisfies all three of these requirements. While the use of block copolymers[7–12] solves the solubility problem for any amino acid, it circumvents the other two difficulties only for a few amino acids; hence, a technique (the subject of this series of papers) has been developed which is applicable to *any* amino acid. It involves the study of random copolymers in which the desired amino acid, the "guest," is incorporated at random into a nonionic homopoly(amino acid), the "host," which does meet all three requirements. The helix–coil stability constants of the guest residue can then be determined from its influence on the melting behavior of the host homopoly(amino acid).

In this first paper of the series, we will examine the ways in which the melting behavior of a copolymer of a given composition differs from that of a homopolymer, focusing atten-

---

[1] This work was supported by research grants from the National Institute of Arthritis and Metabolic Diseases, U. S. Public Health Service (No. AM-08465), and from the National Science Foundation (No. GB-7571X).

[2] National Defense Education Act, Title IV Fellow, 1966–1969; NIH predoctoral trainee, 1969–1971.

[3] D. Kotelchuck and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **61**, 1163 (1968); **62**, 14 (1969).

[4] D. Kotelchuck, M. Dygert, and H. A. Scheraga, *ibid.*, **63**, 615 (1969).

[5] P. N. Lewis, N. Gō, M. Gō, D. Kotelchuck, and H. A. Scheraga, *ibid.*, **65**, 810 (1970).

[6] B. H. Zimm and J. K. Bragg, *J. Chem. Phys.*, **31**, 526 (1959).

[7] W. B. Gratzer and P. Doty, *J. Amer. Chem. Soc.*, **85**, 1193 (1963).

[8] H. J. Sage and G. D. Fasman, *Biochemistry*, **5**, 286 (1966).

[9] H. E. Auer and P. Doty, *ibid.*, **5**, 1708, 1716 (1966).

[10] R. T. Ingwall, H. A. Scheraga, N. Lotan, A. Berger, and E. Katchalski, *Biopolymers*, **6**, 331 (1968).

[11] R. F. Epand and H. A. Scheraga, *ibid.*, **6**, 1551 (1968).

[12] S. E. Ostroy, N. Lotan, R. T. Ingwall, and H. A. Scheraga, *ibid.*, **9**, 749 (1970).

tion on the effects of amino acid sequence, cooperativity, and enthalpy of transition on the melting temperature, the slope at the transition temperature, and the correlation length. We will also consider the theory of the melting of random copolymers, focusing particular attention on parameters which are characteristic of poly(amino acids), since previous papers on the theories[13-22] dealt with parameters which are more appropriate for polynucleotides, and the results are quite different for the two cases. We compare the results of the approximate theories to those of the exact ones, the former being easier to apply in interpreting the experimental data. In the second paper,[23] we will apply the theory to a copolymer of two amino acids, the homopolymers of which both satisfy the above three requirements, *i.e.*, either homopolymer can serve as the host, thereby allowing us to compare the guest parameters obtained from the host-guest copolymer with those obtained directly from the guest homopolymer. In the third paper,[24] the method will be applied to glycine; other amino acids will be treated in subsequent papers of the series.

I. Properties of Copolymers

(A) The Model. In order to see how the sequence of the units in a copolymer affects its properties, we shall consider both regular-sequence and random copolymers of two kinds of residues, A and B, with the overall fractional composition being f_A and f_B, respectively ($f_A + f_B = 1$). Each residue can exist in either the helix (h) or coil (c) state, with the statistical weight of a given residue being assumed to depend on the kind (A or B) of the residue (but not on the kind of any of its neighbors) and on the state (h or c) of the given residue and of its nearest neighbors.

It is also assumed that the partition functions Q_A and Q_B for the homopolymers of A and B, respectively, are given by the Zimm-Bragg[6] theory, *viz.*

$$Q_A = \mathbf{e}\mathbf{W}_A{}^N\mathbf{e}^+ \xrightarrow{N\to\infty} \lambda_A{}^N \qquad (1)$$

$$Q_B = \mathbf{e}\mathbf{W}_B{}^N\mathbf{e}^+ \xrightarrow{N\to\infty} \lambda_B{}^N \qquad (2)$$

where N is the chain length, $\mathbf{e}$ and $\mathbf{e}^+$ are appropriate end vectors, and $\mathbf{W}_A$ and $\mathbf{W}_B$ are 2 × 2 matrices

$$\mathbf{W}_A = \begin{pmatrix} s_A & 1 \\ \sigma_A s_A & 1 \end{pmatrix} \qquad (3)$$

and

$$\mathbf{W}_B = \begin{pmatrix} s_B & 1 \\ \sigma_B s_B & 1 \end{pmatrix} \qquad (4)$$

with eigenvalues λ_A and λ_B, respectively. The quantities σ and s are the well-known Zimm-Bragg parameters, with s being given by

$$s = \exp[-\Delta H/RT + \Delta S/R] \qquad (5)$$

where ΔH and ΔS are the enthalpy and entropy, respectively, for converting a coil to a helix state at the end of a long helical sequence. While there is a suggestion[25] that σ may be temperature dependent, we shall assume here that it is independent of temperature. Also, while ΔH and ΔS will be allowed to be temperature dependent in analyzing experimental data in future papers in this series, for simplicity we shall take them as independent of temperature in the illustrative calculations of the present paper. The fraction of helix states, θ_h, and the fraction of ch borders, θ_{ch}, are given by

$$\theta_h = (1/N)(\partial \ln Q/\partial \ln s) \qquad (6)$$

and

$$\theta_{ch} = (1/N)(\partial \ln Q/\partial \ln \sigma) \qquad (7)$$

respectively. Hence, the average length of a helical sequence, $\bar{L}$, at any temperature is

$$\bar{L} = \theta_h/\theta_{ch} \qquad (8)$$

The melting or transition temperature, T_A or T_B, is that at which $\theta_h = {}^1/_2$ for an infinite chain. In the Zimm-Bragg theory, the correlation length (the value of $\bar{L}$ at T_A (or T_B) for the infinite chain) is $\bar{L} \cong \sigma^{-1/2}$, and s varies from approximately $1 + \sigma^{1/2}$ to $1 - \sigma^{1/2}$ in the range of θ_h from 0.75 to 0.25; the slope, S, of the θ_h *vs.* T curve for the homopolymer of infinite chain length at T_A (or T_B) is

$$S \equiv (\mathrm{d}\theta_h/\mathrm{d}T)_{\theta_h = 1/2} = \Delta H_A/4RT_A{}^2\sigma_A{}^{1/2} \qquad (9)$$

For finite chain length, we shall designate the temperature at which $\theta_h = {}^1/_2$ by $T_A(N)$ [and $T_B(N)$]; $T_A \neq T_A(N)$. In this paper, the B residues are taken to be the stronger helix formers, *i.e.*, $T_B > T_A$.

For copolymers, the transition or melting temperatures (where $\theta_h = {}^1/_2$) for infinite and finite chains will be designated as T_{TR} and $T_{TR}(N)$, respectively, as with homopolymers, $T_{TR} \neq T_{TR}(N)$. The definitions of the quantities θ_h, θ_{ch}, and $\bar{L}$ for copolymers are the same as in eq 6-8, with $\partial/\partial \ln s$ being $\partial/\partial \ln s_A + \partial/\partial \ln s_B$ (and similarly for $\partial/\partial \ln \sigma$); however, eq 9 does not hold for a copolymer.

The regular-sequence copolymers considered here are all of infinite chain length, but have various sizes, κ, of the repeating unit for the same f_A; *e.g.*, for a copolymer with $f_A = 0.70$, the sequence could be $\ldots[A_7B_3]_n\ldots$ or $\ldots[A_{14}B_6]_n\ldots$, where n is very large, for κ values of 10 and 20, respectively. For a particular copolymer, all of the molecules in the system have identical sequences; hence, their properties may be described by the same Zimm-Bragg partition function used for homopolymers.[26,27]

The random copolymers treated here are of both infinite and finite chain length. Since all of the molecules in a solution of a random-sequence copolymer do not have the same sequence, it is necessary to use an ensemble partition function to describe all possible sequences. We have used two methods to obtain an "exact" solution of the partition function, to

---

(13) S. Lifson, *Biopolymers*, **1**, 25 (1963).

(14) S. Lifson and G. Allegra, *ibid.*, **2**, 65 (1964).

(15) E. W. Montroll and N. S. Goel, *ibid.*, **4**, 855 (1966).

(16) H. Reiss, D. A. McQuarrie, J. P. McTague, and E. R. Cohen, *J. Chem. Phys.*, **44**, 4567 (1966).

(17) G. Allegra, *J. Polym. Sci., Part C*, **No. 16**, 2815 (1967).

(18) M. Fixman and D. Zeroka, *J. Chem. Phys.*, **48**, 5223 (1968).

(19) G. W. Lehman and J. P. McTague, *ibid.*, **49**, 3170 (1968).

(20) T. R. Fink and D. M. Crothers, *Biopolymers*, **6**, 863 (1968).

(21) D. Poland and H. A. Scheraga, *ibid.*, **7**, 887 (1969).

(22) S. S. Cohen and O. Penrose, *J. Chem. Phys.*, **52**, 5018 (1970).

(23) P. H. von Dreele, N. Lotan, V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, and H. A. Scheraga, *Macromolecules*, **4**, 408 (1971).

(24) V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, and H. A. Scheraga, *ibid.*, **4**, 417 (1971).

(25) M. Gō, N. Gō, and H. A. Scheraga, *J. Chem. Phys.*, **52**, 2060 (1970); **54**, 4489 (1971).

(26) The computer program for calculating θ_h, θ_{ch}, $\bar{L}$, etc., for regular-sequence copolymers of infinite chain length with different values of f_A and κ and various residue parameters is available.[27] All other computer programs used in this and the succeeding papers[23,24] are also available.[27]

(27) This program has been deposited as Document No. NAPS-01460 with the ASIS National Auxiliary Publication Service, c/o CCM Information Corp., 909 3rd Ave., New York, N. Y. 10022. A copy may be secured by citing the document number and by remitting $2.00 for microfiche or $7.40 for photocopies. Advance payment is required. Make checks or money orders payable to: CCMIC-NAPS.

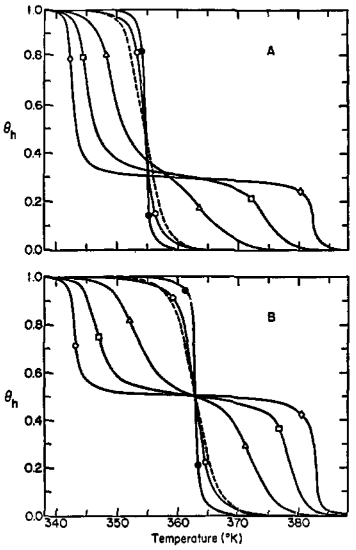

Figure 1. Theoretical melting curves for infinite chains of regular-sequence copolymers of varying κ (—) and for a random copolymer of the same composition (---) for (A) $f_A = 0.7$ and (B) $f_A = 0.5$. These curves were computed with the following parameters: $T_A = 342.5$°K, $T_B = 383.5$°K, $\Delta H_B = -8600$ cal/mol, $\Delta H_A = \Delta H_B T_A / T_B$, and $\sigma = 9 \times 10^{-5}$. The following symbols denote the length, κ, of the repeating unit in the regular-sequence copolymers: (●) 10, (○) 20, (△) 40, (□) 80, and (◇) 320.

which various approximate methods may be compared. The results of the two "exact" approaches for calculating θ_h agreed to within three significant figures. In the first of these, the ensemble partition function was evaluated by the procedure of Cohen and Penrose.[22] In this method,[26,27] the distribution function of the ratio of successive partition functions along the copolymer chain is evaluated numerically to obtain the free energy, which is then differentiated numerically to obtain θ_h. In the second,[26,27] a small ensemble of molecules[20] having sequences produced by a random-number generator was constructed and the properties of such an ensemble were evaluated (see description of finite chain below). Lehman and McTague[19] have shown that, for molecules of long chain length, the results from the ensemble average of all possible molecules of random sequences agree within 1% with those obtained using a single molecule of the same chain length, whose sequence is generated by a random-number generator. In treating infinite chains in this paper, 100,000 units were generated at random[28] and were considered to be in a single molecule of $N = 10^5$. In treating finite chains, the 100,000 units were divided into $\mathfrak{N}$ molecules, each of chainlength $10^5/\mathfrak{N}$; then θ_h was computed for each (specific sequence) copolymer, and the average value of θ_h for the ensemble of $\mathfrak{N}$ molecules was taken.[29] The values of f_A in the generated sequence agreed with the desired input values of f_A in the random-number generator to within three significant figures. The randomness of the generated sequence was checked by repeating the process several times (the computed θ_h *vs.* T curves for each generated sequence agreeing with each other to within 1%), and also by performing three statistical tests as follows: (1) it was shown that the proportion of generated numbers in each 0.1 unit interval between 0 and 1 was the same within 1.5%; (2) for the case of $f_A = 1/2$, it was shown that the numbers of AA, AB, BA, and BB sequences were equal to within 1%; (3) the distribution of A and B sequence lengths in the 100,000 generated units (with $f_A = 0.5$) agreed to within 1% of the theoretical value given by[30]

$$R = (N - L - 1)(1/2)^{L+1} \tag{10}$$

where R is the number of runs of, say, A units of length L in a chain of N units, for which the probabilities of occurrence of A and B are equal.

In the following subsections of section I, we will examine the properties of regular-sequence and random-sequence copolymers, specifically the dependence of the melting temperature, correlation length, and breadth of the transition on composition and sequence.

(B) Transition Temperature. We consider first the dependence of the transition temperature on composition and sequence. Figure 1 shows calculated θ_h *vs.* T curves for infinite chains of regular-sequence copolymers[26,27] with $f_A = 0.7$ and 0.5, respectively, and varying κ, and for a random copolymer of 10^5 units with the same values of f_A. It should be noted that the various curves cross at the same point for $f_A = 0.5$, but not for $f_A = 0.7$. As will be shown below, the composition $f_A = 0.5$ is a special case, and the curves do not cross at the same point for all values of $f_A \neq 0.5$.

Similar calculations were carried out for the whole range of f_A values from 0 to 1, and T_{TR} was determined. These data are plotted in Figure 2 as T_{TR} *vs.* f_A, for values of σ from 1 to 10^{-8}; the value of σ in Figure 2C is the same as that used in Figure 1. It can be seen (from the superposition of the points for low values of κ) that the T_{TR} *vs.* f_A curve approaches linearity as $\sigma \rightarrow 0$ (for *all* values of κ, even large ones, for $\sigma \ll 10^{-8}$). For all other values of σ, the curves exhibit varying amounts of sigmoidal behavior (for large values of κ), the degree of departure from linearity increasing as σ increases (up to a point), and then decreasing as $\sigma \rightarrow 1$; even for $\sigma = 1$, there is a small degree of sigmoidal character to the curve. Thus, the linearity which Marmur and Doty[31] observed for the T_{TR} *vs.* f_A curve for various DNA's probably had an undetectable amount of sigmoidal character to it, and would be best represented by the random copolymer curve in Figure 2C, which happens to be collinear with those for small values of κ; clearly a linear T_{TR} *vs.* f_A curve is not the most general one to be expected.

While, for many values of σ, T_{TR} depends on both f_A and κ, *i.e.*, on both composition and sequence, it can be seen from Figures 2A and 2D that T_{TR} becomes independent of sequence in the limiting cases of $\sigma \rightarrow 1$ and $\sigma \rightarrow 0$.

In Appendix A, the dependence of T_{TR} on f_A is computed for the limiting cases of $\sigma \rightarrow 1$ and $\sigma \rightarrow 0$; it does not appear possible to obtain an analytical expression for intermediate values of σ. For $\sigma \rightarrow 1$, it is shown (eq A-14) that

$$f_A = \frac{\tanh \int_{T_{TR}}^{T_B} \frac{\Delta H_B}{2RT^2} \mathrm{d}T}{\tanh \int_{T_{TR}}^{T_B} \frac{\Delta H_B}{2RT^2} \mathrm{d}T + \tanh \int_{T_A}^{T_{TR}} \frac{\Delta H_A}{2RT^2} \mathrm{d}T} \tag{11}$$

---

(28) The RANDU generator of the IBM Scientific Subroutines Package Version 3 for the 360 System was used here to produce a random sequence of 100,000 numbers from 0 to 1. If each generated number were less than or equal to the desired value of f_A, it was designated as an A unit; if not, it was called a B unit.

(29) The partition function was evaluated and θ_h computed using eq 18 of ref 19 for the melting of a single specific-sequence molecule. The quantities θ_{ch}, X_A, and X_B (where the latter are defined in Appendix A) were obtained by evaluating the partition function in a similar manner. Computations were carried out for $N = 64$, 128, 256, 512, and 1024, for a total number of residues of both 10^5 and 3×10^5. The θ_h *vs.* T curves were identical (to within 0.04%) for a total of both 10^5 and 3×10^5 residues, even when N was as small as 64. The computer program is available.[27]

(30) R. von Mises, "Mathematical Theory of Probability and Statistics," Academic Press, New York, N. Y., 1964, p 187.

(31) J. Marmur and P. Doty, *Nature (London)*, **183**, 1427 (1959).

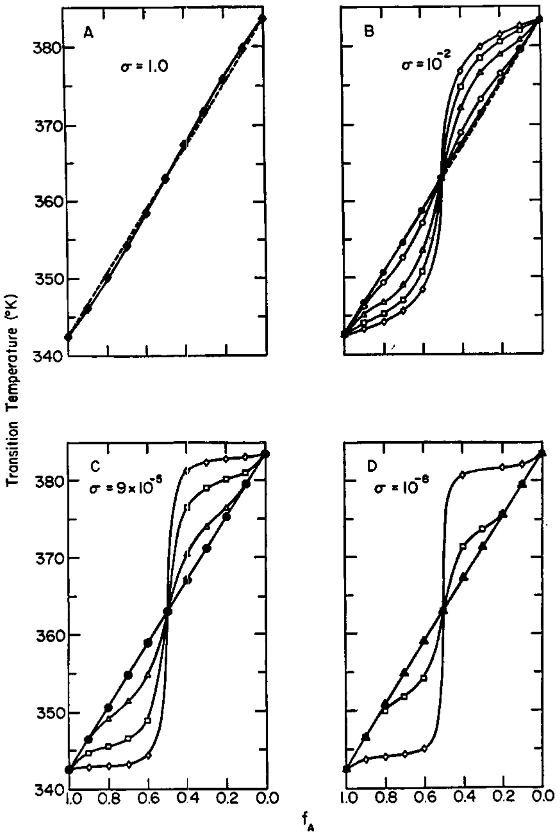

Figure 2. Dependence of melting temperature on composition for infinite chains of regular-sequence copolymers of varying κ, and for random copolymers of the same composition, using the same values of T_A, T_B, ΔH_A, and ΔH_B and symbols for κ as in Figure 1, for the σ values shown. The line for the random-sequence copolymer is indistinguishably close to that of the regular-sequence copolymer of $\kappa = 10$. The dashed line connects the points 342.5 and 383.5°K, and is intended only as a guide to the eye.

where the integrals can be evaluated for either temperature-dependent or -independent enthalpies. If ΔH_A and ΔH_B are assumed to be independent of temperature and, in addition, $\Delta H_A = \Delta H_B T_A/T_B$ (a condition used in computing the curves of Figures 1 and 2), then it can be shown that the values of T_{TR} for f_A = 0.0, 0.5, and 1.0 lie on a straight line for all values of ΔH_B. However, the values of T_{TR} for other values of f_A fall on a sigmoidal curve, and in fact eq 11 reproduces the curve of Figure 2A exactly. For $\sigma \rightarrow 0$, it is shown in Appendix A (eq A-21) that

$$f_A = \frac{\int_{T_{TR}}^{T_B} \frac{\Delta H_B}{T^2}\mathrm{d}T}{\int_{T_{TR}}^{T_B} \frac{\Delta H_B}{T^2}\mathrm{d}T + \int_{T_A}^{T_{TR}} \frac{\Delta H_A}{T^2}\mathrm{d}T} \qquad (12)$$

If, again, it is assumed that ΔH_A and ΔH_B are independent of temperature, and $\Delta H_A = \Delta H_B T_A/T_B$, eq 12 reduces to the simple linear relation

$$T_{TR} = T_B + f_A(T_A - T_B) \qquad (13)$$

for all sequences; it can be seen that the curves of Figures 2C and 2D are approaching such a linear relationship as $\sigma \rightarrow 0$.

For these one-dimensional systems, it is shown in Appendix A that the entropy of mixing of h and c states (for the case where $\sigma \rightarrow 1$) leads to a sigmoidal T_{TR} *vs.* f_A curve; the absence of an entropy of mixing, in the case where $\sigma \rightarrow 0$, leads to a linear curve. For *intermediate* values of σ, between 0 and 1, there is not only the competition between the entropy of mixing of h and c states (tending to produce short helical sections) and the unfavorable boundary free energy tending to produce long helical sequences, which would make the T_{TR} *vs.* f_A curve of Figure 2A decay monotonically to a straight line as σ passes from unity to zero. There is, in addition, the difference in free energies of A and B residues (since B is a better helix former than A), which produces the severe deflection of the T_{TR} *vs.* f_A curve from either limiting value, shown in Figures 2B, 2C, and 2D; *i.e.*, there is the additional competition arising from regional melting (preferential melting of regions which are rich in A) and the loss of conformational entropy caused by such localizations of ordered and disordered regions.

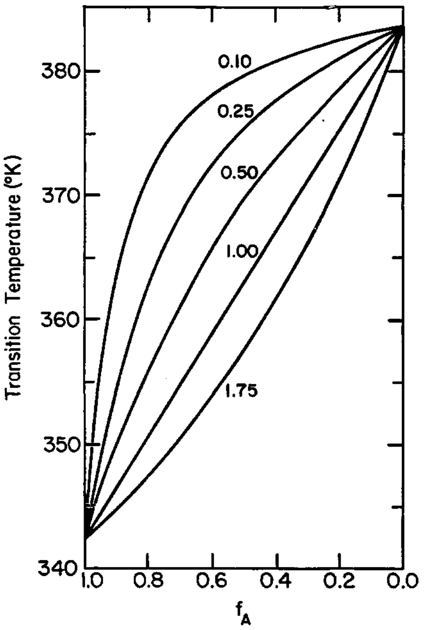

Figure 3. Effect of variation of the ratio $\Delta S_A/\Delta S_B$ on the shape of the T_{TR} *vs.* f_A curve. Identical curves were produced in two ways. First, eq 12 (which holds for all κ as $\sigma \rightarrow 0$, and for low values of κ for $0 < \sigma < 1$) was solved for various values of the input parameters. Second, an infinite-chain-length regular-sequence copolymer of $\kappa = 10$ and $\sigma = 9 \times 10^{-5}$ was used to compute θ_h *vs.* T. In both cases, T_A = 342.5°K, T_B = 383.5°K, and ΔH_B = −8600 cal/mol, and ΔS_A was adjusted to give the ratios $\Delta S_A/\Delta S_B$ shown. Additionally, identical curves were also produced using ΔH_B = −200 cal/mol, with ΔS_A adjusted to give the same ratios $\Delta S_A/\Delta S_B$. Numbers on curves are for $\Delta H_A T_B/\Delta H_B T_A$, *i.e.*, $\Delta S_A/\Delta S_B$.

It should be noted that, in order to obtain a linear T_{TR} *vs.* f_A curve, *i.e.*, to obtain eq 13 from eq 12, in the limit of $\sigma \rightarrow 0$, or for small values of κ when $\sigma \neq 0$, it was necessary to assume that $\Delta H_A/T_A = \Delta H_B/T_B$ (or $\Delta S_A = \Delta S_B$). When this equality does not hold, *i.e.*, when $\Delta S_A \neq \Delta S_B$, the T_{TR} *vs.* f_A curve shows large positive or negative deviations from the curve for which this equality holds (see Figure 3). Similar departures would be observed in temperature–composition diagrams for the distillation of binary mixtures (see Appendix A) if $\Delta S_A \neq \Delta S_B$.

The results of Figures 2 and 3 show the various possible behaviors to be expected for the dependence of T_{TR} on composition, sequence, and entropy for noninteracting A and B units. We thus see how special is the case of the almost linear T_{TR} *vs.* f_A curve obtained for DNA's by Marmur and Doty[31] (besides the fact that the nearest-neighbor Ising model is a

Table I
Effect of Sequence on the Correlation Length in Regular-Sequence Copolymers[a]

Repeating unit	For nucleic acid parameters[b] Correlation length	X_A[d]	For amino acid parameters[c] Correlation length	X_A[d]
A_1B_1	101	0.498	46	0.500
A_2B_2	96	0.496	46	0.500
A_4B_4	79	0.488	46	0.500
A_8B_8	41	0.438	46	0.498
$A_{16}B_{16}$	17	0.120	46	0.492
$A_{32}B_{32}$	32	0.046	45	0.476
$A_{64}B_{64}$	63	0.022	46	0.438
$A_{128}B_{128}$	125	0.010	46	0.394
$A_{256}B_{256}$	244	0.004	47	0.368
$A_{512}B_{512}$	464	0.002	48	0.356

[a] The calculations were carried out for polymers of $f_A = 0.5$ at $T = 363.0$°K, where $\theta_h = {}^1/_2$ for all the copolymers. [b] $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma_A = \sigma_B = 9 \times 10^{-5}$, $\Delta H_B = -8600$ cal/mol, $\Delta H_A = \Delta H_B T_A/T_B$, $\sigma^{-1/2} = 105$. [c] $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma_A = \sigma_B = 5 \times 10^{-4}$, $\Delta H_B = -200$ cal/mol, $\Delta H_A = \Delta H_B T_A/T_B$, $\sigma^{-1/2} = 45$. [d] As defined in Appendix A, this is the fraction of helical residues which are A units.

Table II
Effect of Enthalpy and Cooperativity on the Correlation Length in Random-Sequence Copolymers[a]

ΔH_B,[b] cal/mol	$\sigma = 9 \times 10^{-5}$ Correlation length	X_A[c]	$\sigma = 5 \times 10^{-4}$ Correlation length	X_A[c]
−200	106	0.496	46	0.496
−800	106	0.484	46	0.486
−2000	104	0.466	45	0.464
−3200	98	0.454	44	0.448
−4500	90	0.444	43	0.434
−6500	77	0.430	39	0.416
−8600	65	0.416	36	0.400

[a] The calculations were carried out for polymers of $f_A = 0.5$ at $T = 363.0$°K, where $\theta_h = {}^1/_2$ for all copolymers. [b] $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma_A = \sigma_B$, and $\Delta H_A = \Delta H_B T_A/T_B$. See Table I for values of $\sigma^{-1/2}$. [c] Defined in Appendix A.

poor one for DNA, unless a correction for internal loops is included).

(C) Correlation Length. The correlation length is the value of $\bar{L}$ (defined by eq 8) at the transition temperature for an infinite chain and, for a homopolymer, is essentially $\sigma^{-1/2}$. However, because of the competitive effects mentioned in section IB, $\bar{L}$ is no longer such a simple function of σ for a copolymer. The effect of this competition on $\bar{L}$ can be seen by the following two examples.

First, we compute $\bar{L}$ at the transition temperature for a series of regular-sequence copolymers, for which f_A, ΔH_A, ΔH_B, σ, T_A, and T_B are kept constant, but the sequences are altered (by varying κ) to produce increasingly long blocks of A residues.[26,27] The results are shown in Table I for two sets of parameters. We see that, as the length of the A region increases, the fraction of the helical phase composed of A residues decreases. For nucleic acid parameters, $\bar{L}$ is close to $\sigma^{-1/2}$ (the value for a homopolymer) when A and B alternate, decreases to a minimum value at a value of X_A (defined in Appendix A) which implies that the coil regions exist primarily among the A units, and then increases as the length of the B region increases (approaching the length of the B region at this temperature where very little of the B region had melted). These data indicate how the A and B units differ in the melting properties which they exhibit as homopolymers, *i.e.*, for low values of κ, B units near A units tend to keep the latter from melting, and A units induce the melting of B units, whereas the A and B regions become independent of each other as κ becomes very large. For the set of amino acid parameters considered here (*i.e.*, low values of ΔH and ΔT), the effects are not as dramatic because of a lower tendency toward preferential melting in the A-rich regions, as indicated by the fairly large values of X_A. The change in $\bar{L}$ as κ increases is barely perceptible, but of the same pattern as observed for nucleic acid parameters.

Second, $\bar{L}$ was computed[26,27] for a random-sequence copolymer of 10^5 units with given values of f_A, T_A, T_B, and σ, but variable ΔH_A and ΔH_B. The results in Table II show that $\bar{L}$ decreases monotonically as ΔH_B increases (in absolute value). The helical regions become shorter, tending to localize more closely in those regions of the random chain in which the B content is slightly higher than the average value (*i.e.*, f_B).

(D) Breadth of Transition. For a homopolymer, the slope, S, of the melting curve for an infinite chain at T_A (or T_B) is given by eq 9. However, this equation is not applicable to a copolymer, *i.e.*, S is no longer a simple function of T_{TR}, σ, and ΔH—again because of the competitive effects mentioned in section IB. Instead, the preferential melting of the A-rich regions of a copolymer affects S, just as it did T_{TR} and $\bar{L}$.

The melting curves for regular-sequence copolymers of the type listed in Table I were computed for nucleic acid and one set of amino acid parameters, and are shown in Figures 1B and 4 for $f_A = 0.5$. The curve for a homopolymer whose units have the average statistical weight $\langle s \rangle$ defined as

$$\langle s \rangle = s_A{}^{f_A} s_B{}^{f_B} \tag{14}$$

is collinear with the curve for $\kappa = 10$ in these figures. As seen in these figures, the largest slope at T_{TR} is obtained for the homopolymer of average statistical weight $\langle s \rangle$. For nucleic acid parameters (especially Figure 1B), there is a large change in slope from a maximum (for $\kappa = 10$) to a smaller value (for $\kappa = 320$), the curve for the latter copolymer being nearly biphasic because of the regional melting. For the set of amino acid parameters being considered, the slopes are very similar for most values of κ. This insensitivity of the melting curve for copolymers of *some* amino acids to severe changes in sequence has a very useful practical advantage in the experimental studies of random copolymers, *viz.*, small deviations from randomness will not change the melting curve significantly. This result does not apply to all amino acid copolymers—only to those whose values of s_A and s_B for the individual residues are both close to unity at T_{TR} and whose value of σ is not too small.

The slope of a random-sequence copolymer at T_{TR} is influenced by two opposing effects. In general, S increases as ΔH_A and ΔH_B increase, not only for a homopolymer, as indicated in eq 9, but also for a copolymer such as $(\ldots A_8B_8 \ldots)$, which is shown in Figures 1B and 4. However, if the copolymer contains some A-rich regions, then, as ΔH_A and ΔH_B increase, the localized melting and shorter correlation length lead to a decrease in slope. The net effect, shown in Figure 5, is that the slope passes through a maximum whose location depends on ΔH, σ, and $T_B - T_A$.

II. Theories of Random-Sequence Copolymers

Having seen how the properties of the melting curve of a copolymer depend on composition and sequence, we turn

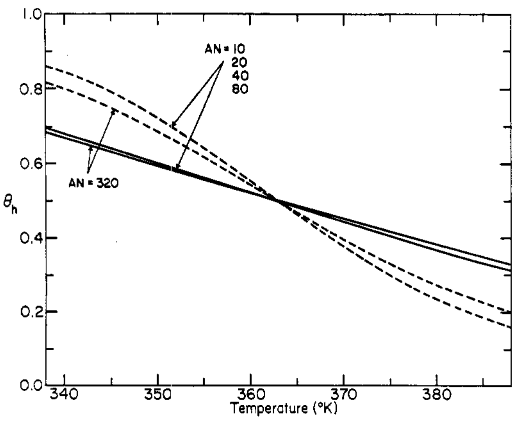

Figure 4. Theoretical melting curves for infinite chains of regular-sequence copolymers of varying κ, for $f_A = 0.5$, $\Delta H_B = -200$ cal/mol, and $\sigma = 9 \times 10^{-5}$ (---) or $\sigma = 5 \times 10^{-4}$(——). The parameters used to compute these curves are $T_A = 342.5$°K, $T_B = 383.5$°K, and $\Delta H_A = \Delta H_B T_A/T_B$.

now to the analysis of experimental data on copolymers. Further computational details are given in paper II.[23] In this series of papers, we analyze the melting curves of random copolymers of a "guest" amino acid and a "host" amino acid to obtain the values of σ and s which characterize the thermally induced helix–coil transition in aqueous solution for the "guest" amino acid. The values of σ and s for the "host" amino acid are known from studies of homopolymers of the "host" amino acid which are water soluble and undergo the desired conformational change in the interval between 0 and 100°C. Using these values of σ and s for the "host," the values of σ and s for the "guest" are determined by adjustment to give the best fit to the experimental θ_h *vs.* T curves for copolymers of given N and f_A.

In order to obtain σ and s for the "guest" amino acid residues by this procedure, it is necessary to have a theory for the helix–coil transition in a random copolymer. While exact theories[19,22] have been developed, they require extensive iterative numerical analysis and thus become computationally prohibitive for use in analyzing experimental data. On the other hand, it is very easy to apply the approximate theories[13,14,17,21] in order to deduce parameters from experimental data on the melting of copolymers. It is thus of primary importance to establish, by comparison with the exact theories, the validity of the numerical results obtained by use of the approximate theories. Such comparisons have already been made,[20,21] but with parameters characteristic of polynucleotides of infinite chain length. It is thus necessary to establish here the validity of the approximate theories, when parameters characteristic of poly(amino acids) are used. The theory for infinite chains will be discussed in section A, and that for finite chains in section B.

(A) Infinite Chains. For a specific-sequence copolymer, *e.g.*, ABABBAAA, the partition function would be

$$Q\{A,B\} = \mathbf{e}\mathbf{W}_A\mathbf{W}_B\mathbf{W}_A\mathbf{W}_B\mathbf{W}_B\mathbf{W}_A\mathbf{W}_A\mathbf{W}_A\mathbf{e}^+ \tag{15}$$

where {A,B} designates a particular sequence of A's and B's. However, in a solution of a random copolymer, prepared from a mixture of A and B monomers, all molecules do not have the same random sequence. Instead, the solute is an ensemble of $\mathfrak{N}$ molecules of varying sequence but of uniform chain length N, whose conformational partition function Z is given by

$$Z = \prod_{\{A,B\}} Q\{A,B\}^{\mathfrak{N}\{A,B\}} = \prod_{\{A,B\}} Q\{A,B\}^{\mathfrak{N}P\{A,B\}} \tag{16}$$

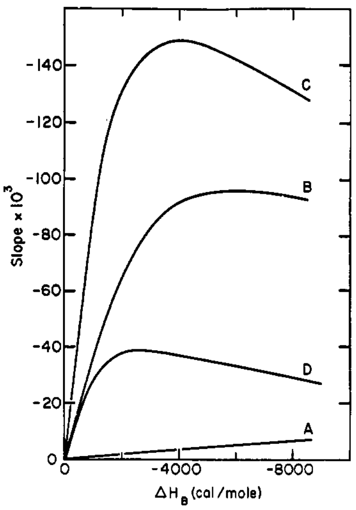

Figure 5. Dependence of the slope of a random-sequence copolymer of $N = 10^5$ at T_{TR} on ΔH for (A) $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma = 1.0$; (B) $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma = 5 \times 10^{-4}$; (C) $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma = 9 \times 10^{-5}$; (D) $T_A = 313.0$°K, $T_B = 413.0$°K, $\sigma = 5 \times 10^{-4}$.

where $\mathfrak{N}\{A,B\}$ is the number of molecules having a particular sequence {A,B} (with partition function $Q\{A,B\}$) and $P\{A,B\}$ is its probability of occurrence. In this paper, $P\{A,B\}$ is regarded as that for a random-sequence copolymer (with no nearest-neighbor correlations). The helix content is then

$$\theta_h = \frac{1}{\mathfrak{N}}\frac{1}{N}\frac{\partial \ln Z}{\partial \ln s} = \frac{1}{N}\sum_{\{A,B\}} P\{A,B\}\frac{\partial \ln Q\{A,B\}}{\partial \ln s} \tag{17}$$

where

$$\frac{\partial}{\partial \ln s} = \frac{\partial}{\partial \ln s_A} + \frac{\partial}{\partial \ln s_B} \tag{18}$$

Poland and Scheraga[21,32] have developed a treatment for obtaining θ_h which involves a hierarchy of approximations; the lowest order approximations reduce to earlier theories of Lifson,[13] Lifson and Allegra,[14] and Allegra,[17] while the result is exact in the limit of infinite order. This treatment, which we will call the Lifson–Allegra–Poland–Scheraga (LAPS) hierarchy, is particularly useful, since one can examine the results of various order approximations of the hierarchy for sample parameters characteristic of polynucleotides or poly(amino acids), and then use the simplest approximation that gives adequate results. Fortunately, we will find that the lowest order approximations are completely adequate for some poly(amino acids) treated in this series of papers. For the other poly(amino acids) it can be used to locate the approximate values of σ and s, and an exact theory then used for the final analysis of the data. The LAPS hierarchy is based on the observation that θ_h of eq 17 can be equally well computed from another partition function, $Z^\dagger$, which is not equal to Z of eq 16, but defined as

$$Z^\dagger = \sum_{\{A,B\}} \frac{P\{A,B\}}{C\{A,B\}} Q\{A,B\} = 1 \tag{19}$$

where $Q\{A,B\}$ is a differentiable function of s_A and s_B, and $C\{A,B\}$ is a number which is the numerical value of $Q\{A,B\}$;

---

(32) Further details (including the extension to a Markoff distribution $P\{A,B\}$) may be found in D. Poland and H. A. Scheraga, "Theory of Helix-Coil Transitions in Biopolymers," Academic Press, New York, N. Y., 1970, Chapter 8.

Table III

The LAPS Hierarchy

Value of m	Treatment of α	Original author(s)	Nature of treatment
0^a	$\alpha = p_A/(1 - p_A)$	Lifson[b]	Approximate
0^a	α adjusted	Allegra[c]	Approximate
1	$\alpha = p_A/(1 - p_A)$	Lifson–Allegra[d]	Approximate
1	α adjusted	Poland–Scheraga[e]	Approximate
2	α adjusted		
4	α adjusted		
.			
.			
.			
N	$\alpha = p_A/(1 - p_A)$	Lehman–McTague[f] and Cohen–Penrose[g]	Exact

[a] The case of $m = 0$ is defined as that for which $C\{A,B\} = 1$. If α is adjusted (Allegra[17] approximation), then this case of $m = 0$ is equivalent to that of $m = 1$ in the Poland–Scheraga[21] treatment. [b] Reference 13. [c] Reference 17. [d] Reference 14. [e] Reference 21. [f] Reference 19. [g] Reference 22.

the last equality of eq 19 follows from the facts that $C\{A,B\} = Q\{A,B\}$ numerically and $P\{A,B\}$ is a normalized probability. Then

$$\theta_h = \frac{1}{N}\frac{\partial \ln Z^\dagger}{\partial \ln s} = \frac{1}{N}\frac{\partial Z^\dagger}{\partial \ln s} = \frac{1}{N}\sum_{\{A,B\}}\frac{P\{A,B\}}{C\{A,B\}}\frac{\partial Q\{A,B\}}{\partial \ln s} \tag{20}$$

Since the last terms on the right-hand sides of eq 17 and 20 are identical, the values of θ_h computed from both equations are the same.

The advantage of introducing $Z^\dagger$ is that, unlike Z, it can be generated with the aid of a matrix-like quantity, *viz.*

$$Z^\dagger = \left(\frac{0}{\mathbf{e}}, \frac{1}{\mathbf{e}}, \frac{0}{\mathbf{e}}, \frac{1}{\mathbf{e}}\right)\mathbf{W}^N\begin{pmatrix}\frac{1}{\mathbf{e}^+}\\ \frac{1}{\mathbf{e}^+}\\ \frac{1}{\mathbf{e}^+}\\ \frac{1}{\mathbf{e}^+}\end{pmatrix} = \tilde{\mathbf{e}}\mathbf{W}^N\tilde{\mathbf{e}}^+ \tag{21}$$

where

$$\mathbf{W} = \begin{pmatrix}\frac{\alpha s_A}{\mathbf{W}_A} & \frac{\alpha}{\mathbf{W}_A} & \frac{s_B}{\mathbf{W}_B} & \frac{1}{\mathbf{W}_B}\\ \frac{\alpha\sigma_A s_A}{\mathbf{W}_A} & \frac{\alpha}{\mathbf{W}_A} & \frac{\sigma_B s_B}{\mathbf{W}_B} & \frac{1}{\mathbf{W}_B}\\ \frac{\alpha s_A}{\mathbf{W}_A} & \frac{\alpha}{\mathbf{W}_A} & \frac{s_B}{\mathbf{W}_B} & \frac{1}{\mathbf{W}_B}\\ \frac{\alpha\sigma_A s_A}{\mathbf{W}_A} & \frac{\alpha}{\mathbf{W}_A} & \frac{\sigma_B s_B}{\mathbf{W}_B} & \frac{1}{\mathbf{W}_B}\end{pmatrix} = \begin{pmatrix}\frac{\alpha s_A}{\mathbf{W}_A} + \frac{s_B}{\mathbf{W}_B} & \frac{\alpha}{\mathbf{W}_A} + \frac{1}{\mathbf{W}_B}\\ \frac{\alpha\sigma_A s_A}{\mathbf{W}_A} + \frac{\sigma_B s_B}{\mathbf{W}_B} & \frac{\alpha}{\mathbf{W}_A} + \frac{1}{\mathbf{W}_B}\end{pmatrix} \tag{22}$$

and $\mathbf{W}_A$ and $\mathbf{W}_B$ are given by eq 3 and 4, respectively. The second equality in eq 22 holds if, as assumed here, the statistical weights are independent of neighbor type. In eq 21, it is to be understood that the indicated operations are carried out so that the multiplication of factors in the "numerators" and "denominators" are performed independently and in parallel until finally the matrix products accumulating in the denominators are transformed into scalar quantities by operation with the end vectors. Upon matrix multiplication in this manner, the matrices $\mathbf{W}_A$ and $\mathbf{W}_B$ lead to the quantity $C\{A,B\}$ and the factors α (for the case of a random distribution considered here) generate the quantities $P\{A,B\}$, where

$$P\{A,B\} = p_A^{N_A}p_B^{N_B} = (1 - p_A)^N\left(\frac{p_A}{1 - p_A}\right)^{N_A} \tag{23}$$

and p_A and p_B are the *a priori* probabilities of occurrence of A and B ($p_A + p_B = 1$) and N_A and N_B are the average numbers of A and B units per chain ($N_A + N_B = N$), respectively. Since the factor $(1 - p_A)^N$ is a constant for all $\{A,B\}$, which cancels in all calculations, we drop it and assign each A unit the factor

$$\alpha = p_A/(1 - p_A) \tag{24}$$

In $Z^\dagger$ of eq 21, with α defined by eq 24, the *a priori* probability of occurrence of A, p_A, is taken to be the experimentally determined fraction of A in the copolymers, f_A. On evaluation of $Z^\dagger$, the theoretically calculated fraction of A in the copolymer, F_A, will be equal to f_A and p_A. The solution to eq 21 in this manner would be an exact one for the model assumed. However, for large N, it is much too difficult to carry out the indicated operations explicitly.

This difficulty is circumvented by the series of approximations for $Z^\dagger$ represented by the LAPS hierarchy. In these approximations, the matrix products in the denominators of eq 22, obtained after m explicit matrix multiplications have been performed, are replaced by scalar quantities, *i.e.*

$$Z^\dagger = \tilde{\mathbf{e}}\mathbf{W}^N\tilde{\mathbf{e}}^+ = \tilde{\mathbf{e}}(\mathbf{W}^m)^{N/m}\tilde{\mathbf{e}}^+ \cong \mathbf{e}(\mathcal{W}_m)^{N/m}\mathbf{e}^+ \equiv Z_m^\dagger \tag{25}$$

where $\mathcal{W}_m$ is now a proper matrix. For infinite chains, the scalars in the denominators of $\mathcal{W}_m$ are the largest eigenvalues (λ_A, λ_{AB}, and λ_{ABAA}, respectively) of the matrix products in the denominators of $\mathbf{W}^m$ for orders of approximation $m = 1$, 2, and 4, respectively. In the succession of approximations we compute, in the lowest ($C\{A,B\} = 1$) and highest (exact, *i.e.*, $C\{A,B\}$ equal to $Q\{A,B\}$ numerically) orders, values of θ_h which are the same as those calculated from the partition functions corresponding respectively to the arithmetic mean ($\Sigma\mathfrak{N}\{A,B\}Q\{A,B\}$) and geometric mean ($\Pi Q\{A,B\}^{\mathfrak{N}\{A,B\}}$) of the molecular partition function $Q\{A,B\}$. The LAPS hierarchy is then a progression from the incorrect arithmetic to the correct geometric mean of the molecular partition function. The rate of convergence is governed by the condition that the arithmetic mean be a good approximation to the geometric mean, which will be true when the values of $Q\{A,B\}$ are very nearly equal to each other, which in turn is true when σ_A and s_A are nearly equal to σ_B and s_B, respectively.

In the paper of Poland and Scheraga,[21] in which infinite chains were treated, the rate of convergence of the approximate to the exact result was accelerated by using a device introduced by Allegra,[17] *i.e.*, the value of α was taken as an adjustable parameter to satisfy the condition

$$f_A = (1/N)(\partial \ln Z_m^\dagger/\partial \ln \alpha) \tag{26}$$

where f_A is the experimental value of the fraction of A units. In the limit, as $m \rightarrow N \rightarrow \infty$, α will approach the value given by eq 24. The LAPS hierarchy can be viewed as a composite of special cases, *i.e.*, of particular values of m and α, as shown in Table III. Details of calculations for $m = 0$ and 1 are

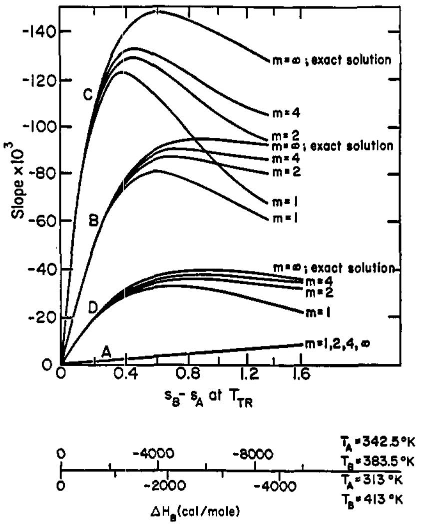

Figure 6. Effect of ΔH, σ, $(T_B - T_A)$, and $s_B - s_A$ on the rate of convergence of the LAPS hierarchy (expressed in terms of S_m at T_{TR}) for *infinite* chain lengths, with α adjustable to force F_A to be 0.5. In all cases $\sigma_A = \sigma_B$ and $\Delta H_A = \Delta H_B T_A / T_B$: (A) $\sigma = 1$, $T_A = 342.5$°K, and $T_B = 383.5$°K ($\Delta T = 41$°); (B) $\sigma = 5 \times 10^{-4}$, $T_A = 342.5$°K, and $T_B = 383.5$°K ($\Delta T = 41$°); (C) $\sigma = 9 \times 10^{-5}$, $T_A = 342.5$°K, and $T_B = 383.5$°K ($\Delta T = 41$°); (D) $\sigma = 5 \times 10^{-4}$, $T_A = 313.0$°K, and $T_B = 413.0$°K ($\Delta T = 100$°).

given in Appendix B; the computer programs for these and higher values of m are available.[27]

The rate of convergence of the LAPS hierarchy to the exact solution was examined for infinite chains, using a variety of nucleic acid and amino acid parameters. For this purpose, the melting curves were characterized by their melting temperatures, T_{TR}, and slopes, S, at T_{TR}. We find that the values of T_{TR} for each order of approximation coincide with those of the exact solution when α is adjusted to give the proper f_A. The various orders of approximation (m) are assessed solely in terms of the slope of each order of approximation, S_m. Figure 6 shows how S_m depends on ΔH, σ, $(T_B - T_A)$, and $s_B - s_A$ at T_{TR}, for $m = 1, 2, 4$, and ∞. It can be seen that, when $s_B - s_A$ is small, the convergence is so rapid that all orders of approximation give the correct result; for these cases, both s_A and s_B are close to unity over much of the transition region of the copolymers, and hence the arithmetic and geometric means of the partition function are about equal. The point at which the approximation fails to hold is seen to depend not only on s but also on σ.

For parameter set D of Figure 6, for example, the convergence of the approximate to the exact solution is not rapid. Further, the curves for $m = 2$ and $m = 4$ will cross the one for the exact result (not shown in Figure 6) at $\Delta H_B = -5600$ cal/mol ($s_B - s_A = 2.2$) and $\Delta H_B = -6600$ cal/mol ($s_B - s_A = 2.7$), respectively, indicating that the approach of S_m to S_∞ is not monotonic at high values of ΔT and ΔH. This is readily seen by plotting S_m/S_∞ against $1/m$ (Figure 7). For some amino acid parameters, a lower order of approximation (Allegra's method[17]) is suitable for analyzing experimental data from copolymers of infinite chain length, *i.e.*, for the range of parameters in Figure 6 where all orders of approximation agree with the exact result. This range includes some but not all amino acids. In fact, even Lifson's earlier procedure[13] (with α given by eq 24) is adequate for certain poly(amino acid) copolymers in this range of parameters; however, we find that Allegra's method is valid over a wider range of parameters than Lifson's. The breakdown of the latter approximation would be indicated by a discrepancy in the value of T_{TR} from that obtained with the exact theory.

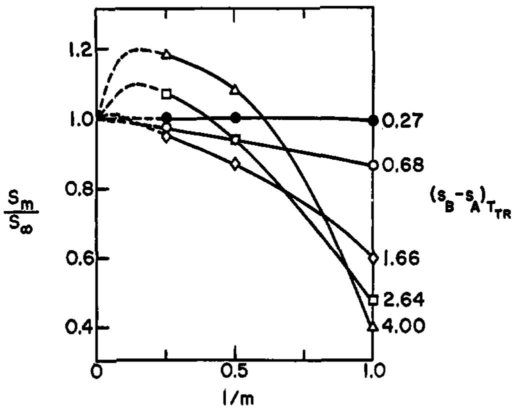

Figure 7. A plot of S_m/S_∞ against $1/m$ for *infinite* chains with $f_A = 0.5$, using the parameters $T_A = 313$°K, $T_B = 413$°K, $\sigma = 5 \times 10^{-4}$, $\Delta H_A = \Delta H_B T_A / T_B$: (●) $\Delta H_B = -800$, (○) $\Delta H_B = -2000$, (◊) $\Delta H_B = -4500$, (□) $\Delta H_B = -6500$, (Δ) $\Delta H_B = -8600$ cal/mol. The curves are intended only as guides to the eye, and not as representatives of the functional form of the relation shown by the points.

(B) Finite Chains. We now examine the situation for finite chains, since most experimental data are obtained for copolymers, for which the infinite-chain approximation may not be applicable. In principle, the LAPS hierarchy will give the correct result as $m \rightarrow N$, even when N is not large, if proper normalizing factors are inserted in the denominators of eq 21 and 22. However, in treating finite chains, the replacement of these denominators by their maximum eigenvalues would not lead to the correct value of θ_h in the limit $m \rightarrow N$. The correct value of θ_h (*i.e.*, the correct values of $C\{A,B\}$ and $P\{A,B\}$) for $m = N$, of course, would be obtained if the matrix multiplication of eq 21 were carried out explicitly. However, since this procedure consumes too much computer time for large N, we resort to the following approximate procedure for *small* values of m, in the spirit of the LAPS hierarchy, in order to analyze experimental data. The matrices in the denominators of $\mathbf{W}^m$ are replaced by their maximum eigenvalues, and α is adjusted (according to eq 26) to make $f_A = F_A$ at each value of m and N; while these are infinite-chain procedures, we retain the finite-chain aspect of $Z_m^\dagger$ of eq 25; *i.e.*, after $\mathcal{W}_m$ is evaluated, it is inserted in eq 25 together with the end vectors $\mathbf{e}$ and $\mathbf{e}^+$ (whose values are given in eq B-3 of Appendix B). It is recognized that, unless the geometric and arithmetic means of the partition function are nearly equal, this choice of normalizing factor will not allow the hierarchy of approximations to converge to the exact solution as $m \rightarrow N$; however, for *low* values of m, the approximate solution will move more rapidly to its limiting value than if no normalizing factor were introduced.

The success of this approximate method in yielding the correct melting curve can be judged by comparing the resulting approximate values of $T_{TR}(N)$ and the slope at $T_{TR}(N)$ with the exact values. It was found that the values of $T_{TR}(N)$, computed for $m = 1, 2$, and 4, were all the same and only slightly higher than those computed with the exact theory. The difference in $T_{TR}(N)$ for $m = 1, 2$, and 4 and $m = N$ is largest for small N and large values of $(s_B - s_A)$ at $T_{TR}(N)$ (being about 0.5°C for a copolymer of $N = 64$, $\sigma = 5 \times 10^{-4}$, and $(s_B - s_A) = 3.7$); of course, this difference approaches zero as $N \rightarrow \infty$. Figure 8 shows the slope of the melting curves [at $T_{TR}(N)$] for the low-order approximations

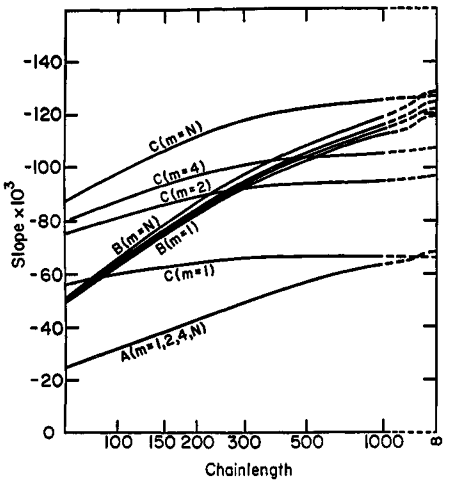

Figure 8. Semilogarithmic plot comparing the slopes at $T_{TR}(N)$ for low values of m (LAPS hierarchy) with those for the exact solution ($m = N$), for several values of ΔH_B. In all cases, $f_A = 0.5$, $T_A = 342.5$°K, $T_B = 383.5$°K, $\sigma = 9 \times 10^{-5}$, and $\Delta S_A = \Delta S_B$: (A) $\Delta H_B = -800$, (B) $\Delta H_B = -2000$, (C) $\Delta H_B = -8600$ cal/mol. In case B, the (unlabeled) curves for $m = 2$ and 4 lie between those for $m = 1$ and N.

and the exact result for a few selected values of ΔH_B. It can be seen that the approximate method leads to the correct slope for finite chains for small values of ($s_B - s_A$) (which correspond to small values of ΔT and ΔH). For the other values of ($s_B - s_A$) illustrated, the slopes for $m = 2$ and 4 lie between those for $m = 1$ and N, *i.e.*, the higher order approximations at low m do lead toward the exact result for $m = N$.

For the amino acid copolymers studied in papers II[23] and III,[24] it was found that the low-order approximations, $m = 0$ and 1, were sufficient to obtain good agreement with the exact theory. In such a case it is not necessary to introduce a normalizing factor in the denominator, since α is adjusted (see Appendix B), *i.e.*, we are effectively using the Allegra approximation ($m = 0$) of Table III for finite chains. It can be seen that this approximation (shown as $m = 1$ in Figure 8) approaches the exact result ($m = N$) more closely when N is finite than when $N \rightarrow \infty$. This suggests a simple test to determine whether Allegra's method will suffice for any finite chain length, *viz.*, first use Figure 6 to see if it is valid for the *infinite* chain for the particular values of $(s_B - s_A)_{T_{TR}}$ and σ of interest; if it is satisfactory for infinite chains, then it will also be satisfactory for any finite chain length. When ($s_B - s_A$) at T_{TR} is small and σ_A and σ_B are large (even as small as 10^{-4} in some cases), the zero-order approximations (the methods of Lifson[18] and Allegra,[17] the details of which are given in Appendix B) are simple and valid methods for analyzing experimental data. Otherwise, the exact procedure (*e.g.*, that of Cohen and Penrose[22]) can be used.

(C) Effect of Composition on the Rate of Convergence of the LAPS Hierarchy for Infinite Chains. So far, we have considered the behavior of infinite and finite chains for the case where $f_A = 0.5$. We now consider the case of *infinite* chains for which $f_A \neq 0.5$. For a random copolymer, the slope at T_{TR} is not a linear function of the slopes of the homopolymers. Instead, S exhibits a minimum, as a function of f_A, as a result of regional melting. As shown in Figure 9, the convergence of the LAPS hierarchy is slowest at the composition of the minimum, and better for all other values of f_A. Thus, while Figure 6 shows the first-order (Allegra[17]) approximation for the most unfavorable case ($f_A = 0.5$), it can be seen from Figure 9 that the first-order approximation is much better (for the same values of ΔH, ΔT, and σ) at other values of f_A.

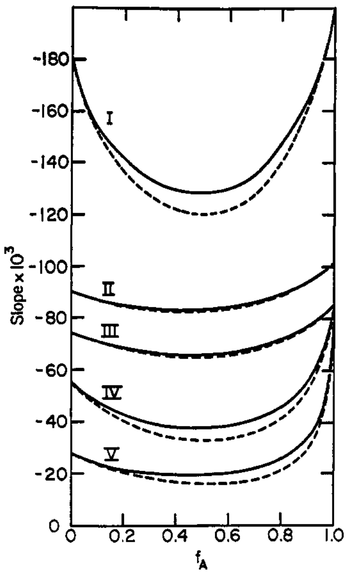

Figure 9. Effect of composition on the comparison (expressed as S at T_{TR}) of the first-order (Allegra[17]) approximation (---) with the exact result (—) for *infinite* chains: (I) $T_A = 342.5$°K, $T_B = 383.5$°K, $\Delta H_B = -2000$ cal/mol, $\sigma = 9 \times 10^{-5}$; (II) $T_A = 342.5$°K, $T_B = 383.5$°K, $\Delta H_B = -1000$ cal/mol, $\sigma = 9 \times 10^{-5}$; (III) $T_A = 342.5$°K, $T_B = 383.5$°K, $\Delta H_B = -2000$ cal/mol, $\sigma = 5 \times 10^{-4}$; (IV) $T_A = 213.0$°K, $T_B = 313.0$°K, $\Delta H_B = -1000$ cal/mol, $\sigma = 5 \times 10^{-4}$; (V) $T_A = 113.0$°K, $T_B = 313.0$°K, $\Delta H_B = -500$ cal/mol, $\sigma = 5 \times 10^{-4}$. In all cases $\Delta S_A = \Delta S_B$. Since T_{TR} varies with f_A, the quantity ($s_B - s_A$) at T_{TR} will also vary with f_A; the values of ($s_B - s_A$) at T_{TR} for $f_A = 0.5$ are (I) 0.30, (II) 0.15, (III) 0.30, (IV) 0.62, and (V) 0.77.

Many of the copolymers studied in this series are of finite chain length and contain only small amounts of the "guest" residue, *i.e.*, $f_A < 0.5$. In order to determine whether the Allegra approximation will be valid for fitting the experimental data for particular values of σ and s, the calculations are first carried out for infinite chains and $f_A = 0.5$ (see Figure 6); if Figure 6 shows that the Allegra approximation suffices for the infinite chain with $f_A = 0.5$, then it will surely hold for finite chains with $f_A < 0.5$.

Finally, it is of interest to make the following points about the convergence of the approximate to the exact solution for the various sets of parameters illustrated in Figure 9. First, the fact that convergence has occurred in case II does not mean that it will occur for all poly(amino acids) with ΔH smaller (in absolute value) than -1000 cal/mol, since the values of ΔT and σ also influence the rate of convergence (see case IV for which $\Delta H = -1000$ cal/mol but $\Delta T = 100$° and case V for which $\Delta H = -500$ cal/mol but $\Delta T = 200$°). Second, the fact that near convergence occurs in case III, for which ($s_B - s_A$) at T_{TR} is 0.30, does not mean that it will occur in all cases where ($s_B - s_A$) at T_{TR} is 0.30, since the value of σ also influences the rate of convergence (see case I).

III. Discussion

We have shown that the LAPS hierarchy converges very rapidly for certain ranges of parameters applicable to poly(amino acids), and that the first-order approximation is essentially exact in certain ranges, for both infinite and finite chains; this makes it very easy to carry out calculations on the helix–coil transition in random copolymers. In essence, as σ_A and s_A approach σ_B and s_B, the copolymer becomes essentially a homopolymer, and the need to take explicit account of the copolymer character (*e.g.*, the required use of the $C\{A,B\}$'s) becomes less important. Only when σ_A and s_A differ appreciably from σ_B and s_B do we see the large effect of the copolymer character, which then requires a high-order approximation for adequate treatment.

In paper II,[23] we present evidence which indicates that there is good agreement between theory and experiment, and further that the host–guest technique yields good values of the (known) homopolymer parameters, thus providing some degree of confidence that the host–guest technique can be used to obtain "guest" helix–coil parameters for the naturally occurring amino acids in water, even though the guest homopolymers do not satisfy the three requirements set forth in the introductory section. We now turn to a brief discussion of the validity of the model on which the theory is based.

The mathematics of sections I and II is given in terms of a model involving nearest-neighbor interactions. Yet, in the α helix, the amide groups and side chains four residues apart are those in closest contact; further, the dipole–dipole interactions of the amide groups in the α helix are long range, becoming negligible only after about five to ten residues. We may then ask how justified is a formulation in terms of nearest-neighbor interactions. The answer (discussed in detail elsewhere[33,34]) is that, for long sequences, the free energy of formation of a helical sequence of j residues from the coil state is

$$\Delta G_{\text{helix}}(j) = j\Delta G_{c\rightarrow h} - RT \ln \sigma \text{ for } j \geq j_{tr} \qquad (27)$$

where $-RT \ln \sigma$ represents the (positive) contribution to $\Delta G_{\text{helix}}(j)$ from end effects, and j_{tr} is the truncation length beyond which interresidue interactions become negligible. It should be noted that eq 27 holds for $j \geq j_{tr}$, and thus all the interactions mentioned at the beginning of this paragraph are included in eq 27, *i.e.*, $\Delta G_{c\rightarrow h}$ and σ include interactions up to j_{tr} residues away. Now, in the cooperative helix–coil transition, the distribution of helical lengths, L_h, in the vicinity of $s \sim 1$ is a broad function,[33,34] *i.e.*

$$L_h \sim \sigma^{-1/2} \pm \sigma^{-1/2}/2 \qquad (28)$$

This means that almost all of the residues are in helical sequences larger than j_{tr}. Thus, one can treat *all* sequence lengths by eq 27, committing an error for $j < j_{tr}$ but, since these states are improbable anyway, the error is very small. Thus, eq 3 and 4 are actually "pseudo" nearest-neighbor representations; the nearest-neighbor form is valid if there is an average correlation (sequence) length that is large compared with the extent of interresidue interactions.

The assumption that the copolymers are random is not critical. This has been documented by Lehman and McTague[19] who have compared random and Markoffian distributions for the same f_A, finding little difference, and by Poland and Scheraga[33,34] who have computed detailed helical probability profiles for a given random sequence for various degrees of coarse graining of the sequence, the degree of coarse graining making little difference if the grain size is much less than the correlation length. From the above evidence, we have reason to believe that, if there are departures from randomness on a scale (grouping) that is much less than the correlation length ($\sim\sigma^{-1/2}$), then there will be little or no effect on the helix–coil transition. This was also borne out by the data of Figure 4 of the present paper.

Finally, there is the assumption that the statistical weights for A and B are independent of the nature of their neighbors. We stress that this is strictly the simplest assumption that can be made, although it appears to be a very reasonable one.[3–5] We have, in fact, reason to believe from theoretical calculations[25] of s and σ that the neighboring side chains can affect the values of s and σ (in second order) for a given residue. Thus, this assumption may need modification; if guest parameters are determined at a given ratio of guest/host, then these parameters may not apply at a different guest/host ratio. For the copolymers studied in paper II,[23] the assumption of independence seems to be valid, but this may be due to the close similarity of the two side chains (although the transition temperatures differ by 40°). One could introduce a hierarchy of σ and s factors depending on the nature of the nearest neighbors; however, there is not enough experimental sensitivity to determine such a large number of adjustable parameters. Perhaps conformational energy calculations[25] will yield good approximations for such a hierarchy of statistical weights based on a few empirical constants. At the present, we can determine the average σ and s [*i.e.*, ΔH and $T_{TR}(N)$] in water, for example, for a copolymer of up to 20% glycine guest in a particular host. Such a determination is presented in paper III.[24] While the possibility exists that the guest parameters may depend on the nature of the host, the results discussed in ref 3–5, where proteins are considered, suggest that the assumption of independence is a very good first-order approximation.

Acknowledgment. We should like to thank Mrs. Karen B. Platzer for checking some of the computer programs, Mr. Peter N. Lewis for writing the Cohen–Penrose computer program, and Dr. V. S. Ananthanarayanan for helpful discussions.

Appendix A. Dependence of T_{TR} on f_A

Our primary aim in this Appendix (besides deriving eq 11 and 12 of the main text) is to show that the T_{TR} *vs.* f_A curve is linear in a very special case. If there is no entropy of mixing, the curve will be linear, while any (even small) amount of entropy of mixing will lead to a sigmoidal-shaped curve. As will be shown here, an entropy of mixing arises (in a three-dimensional solution) upon mixing of two species and (in a one-dimensional polymeric chain) because of the mixing of h and c states. The entropy of mixing is absent only for the one-dimensional case with $\sigma \rightarrow 0$.

The helix–coil equilibrium in a copolymer is a special (one dimensional) case of equilibria in binary mixtures. Since the A and B units are localized in the one-dimensional case, in contrast to the three-dimensional solution where they can move around with respect to each other, the translational entropy (entropy of mixing of *species*) which is present in the three-dimensional solution is absent in the one-dimensional problem. As a result, the nature of the dependence of T_{TR} on f_A will sometimes differ in the two cases, although the general method for treating equilibria in binary systems is the same for both dimensionalities; a lattice model will be used to treat these mixtures.

In treating the limiting cases of $\sigma \rightarrow 1$ and $\sigma \rightarrow 0$ here, we recognize that these apply in general to independent equilibria ($\sigma \rightarrow 1$) or to highly cooperative equilibria ($\sigma \rightarrow 0$). For the one-dimensional problem, the equilibria envisaged are, of course, the helix–coil transitions

$$\begin{matrix} A_h \rightleftarrows A_c \\ B_h \rightleftarrows B_c \end{matrix} \qquad (A\text{-}1)$$

---

(33) D. Poland and H. A. Scheraga, *Physiol. Chem. Phys.*, **1**, 389 (1969).

(34) Reference 32, p 135.

For $\sigma \rightarrow 1$, these two equilibria are independent of each other (and of the sequence), and the transition of an A or a B residue from an h to a c state occurs without being influenced by the state of a neighboring residue. For $\sigma \rightarrow 0$, the molecule will be completely helical for $T < T_{TR}$ and completely coil for $T > T_{TR}$ (independent of the sequence), with an infinitely sharp transition at T_{TR}. The three-dimensional analogs of these cases are the following. For $\sigma \rightarrow 1$, eq A-1 could pertain to two unrelated chemical reactions in a solution of two species A and B, where both species interconvert between two forms and neither equilibrium influences the other; for example, A could be 1,2-dichloroethylene undergoing a cis–trans equilibrium and B could be the corresponding fluorinated compound, with T_{TR} being the temperature at which the *total* number of moles of cis form in the system equals the *total* number of moles of trans form. For $\sigma \rightarrow 0$, the system could be a binary liquid solution in equilibrium with the binary vapor, *e.g.*, benzene and toluene; to preserve the analogy (requiring that the behavior of the A units is unaffected by the *kind* of B unit), we consider only an ideal solution for the case of $\sigma \rightarrow 0$, with T_{TR} being the temperature at which the total number of moles of liquid is equal to the total number of moles of gas. In all cases, T_{TR} depends on f_A, and we seek to obtain this relationship.

We designate the total number of A residues by N_A, and those in states 1 and 2 by N_{A1} and N_{A2}, respectively; for the helix–coil transition, 1 and 2 would refer to h and c, respectively. Corresponding designations are assigned to B residues with $N_A + N_B = N$. Since we are dealing only with infinite systems here, we are concerned with T_A, T_B, T_{TR}, rather than the corresponding quantities for finite systems. Then T_A is the temperature at which $N_1 = N_{A1} = N_{A2} = N_2$ for a system of pure A, while T_{TR} is the temperature at which $N_1 = N_{A1} + N_{B1} = N_{A2} + N_{B2} = N_2$ for a system of A and B. The mole fraction of the residues in state 1 which are A's is designated X_A, where $X_A = N_{A1}/N_1$; similarly, the mole fraction of residues in state 2 which are A's is designated Y_A, where $Y_A = N_{A2}/N_2$. The pressure will be held constant.

Consider first the limiting case of $\sigma \rightarrow 1$. The partition functions Q_{1d} and Q_{3d}, for the one- and three-dimensional cases, respectively, are

$$Q_{1d} = \sum_{\substack{N_{A1}=0 \\ (N_A+N_B=N)}}^{\infty} \sum_{N_{B1}=0}^{\infty} q_{A1}^{N_{A1}} q_{A2}^{N_{A2}} q_{B1}^{N_{B1}} q_{B2}^{N_{B2}} \times \frac{N_A!}{N_{A1}!N_{A2}!} \frac{N_B!}{N_{B1}!N_{B2}!} \tag{A-2}$$

and

$$Q_{3d} = \sum_{\substack{N_{A1}=0 \\ (N_A+N_B=N)}}^{\infty} \sum_{N_{B1}=0}^{\infty} q_{A1}^{N_{A1}} q_{A2}^{N_{A2}} q_{B1}^{N_{B1}} q_{B2}^{N_{B2}} \times \frac{N!}{N_{A1}!N_{A2}!N_{B1}!N_{B2}!} \tag{A-3}$$

respectively. The two differ by the factor $N!/N_A!N_B!$, which takes into account the number of ways that the A's and B's can be arranged when A and B can be interchanged, as they can be in the three-dimensional case. Application of the binomial theorem in eq A-2 and A-3 leads to

$$Q_{1d} = (q_{A1} + q_{A2})^{N_A}(q_{B1} + q_{B2})^{N_B} \tag{A-4}$$

and

$$Q_{3d} = (q_{A1} + q_{A2})^{N_A}(q_{B1} + q_{B2})^{N_B}\frac{N!}{N_A!N_B!} \tag{A-5}$$

In the one-dimensional case, it can be shown that the chemical potential of A1 in a mixture, μ_{A1} (which is $-RT\partial \ln Q/\partial N_{A1}$ on a molar basis), is

$$\mu_{A1} = \mu_{A1}' + RT \ln \frac{N_{A1}}{N_A} \tag{A-6}$$

where μ_{A1}' is the chemical potential of pure A in state 1 at the same temperature and pressure. Correspondingly

$$\mu_{A2} = \mu_{A2}' + RT \ln \frac{N_{A2}}{N_A} \tag{A-7}$$

At any temperature T where the two states of A coexist at equilibrium, $\mu_{A1} = \mu_{A2}$; hence, at the temperature T_{TR} we obtain

$$\Delta\mu_A'(T_{TR}) = \mu_{A1}' - \mu_{A2}' = RT_{TR} \ln \frac{N_{A2}}{N_{A1}} = RT_{TR} \ln \frac{Y_A}{X_A} \tag{A-8}$$

It can easily be shown in a similar manner that, whereas eq A-6 and A-7 have a different form in the three-dimensional case, eq A-8 nevertheless holds. The Gibbs–Helmholtz equation is used to relate the change in the free energy of pure A at T_{TR} to that at T_A, *i.e.*

$$\mathrm{d}(\Delta\mu'/T)/\mathrm{d}T = -\Delta H/T^2 \tag{A-9}$$

Hence

$$\frac{\Delta\mu_A'(T_{TR})}{T_{TR}} = \frac{\Delta\mu_A'(T_A)}{T_A} - \int_{T_A}^{T_{TR}} \frac{\Delta H_A}{T^2}\mathrm{d}T \tag{A-10}$$

For pure A at T_A, $\Delta\mu_A'(T_A) = 0$; hence we obtain

$$\frac{\Delta\mu_A'(T_{TR})}{T_{TR}} = -\int_{T_A}^{T_{TR}} \frac{\Delta H_A}{T^2}\mathrm{d}T \tag{A-11}$$

On combining eq A-8 and A-11, we obtain

$$X_A = Y_A \exp\left[\int_{T_A}^{T_{TR}} \frac{\Delta H_A}{RT^2}\mathrm{d}T\right] \tag{A-12}$$

In a similar way, it can be shown that

$$X_B = Y_B \exp\left[\int_{T_B}^{T_{TR}} \frac{\Delta H_B}{RT^2}\mathrm{d}T\right] \tag{A-13}$$

Since $X_A + X_B = 1$ and $Y_A + Y_B = 1$, eq A-12 and A-13 lead to an expression for Y_A, which can then be substituted into eq A-12 to give X_A. Since $f_A = (X_A + Y_A)/2$ at T_{TR}, these derived expressions for X_A and Y_A give

$$f_A = \frac{\tanh \int_{T_{TR}}^{T_B} \frac{\Delta H_B}{2RT^2}\mathrm{d}T}{\tanh \int_{T_{TR}}^{T_B} \frac{\Delta H_B}{2RT^2}\mathrm{d}T + \tanh \int_{T_A}^{T_{TR}} \frac{\Delta H_A}{2RT^2}\mathrm{d}T} \tag{A-14}$$

which is eq 11. As seen from the derivation, eq A-14 applies to both the one- and three-dimensional cases for $\sigma \rightarrow 1$; *i.e.*, the relation between f_A and T_{TR} is the same even though the partition functions, chemical potentials, and entropies all differ.

Turning to the case of $\sigma \rightarrow 0$, in which the molecules in state 1 and state 2 exist in two physically distinct phases, the partition functions for the one- and three-dimensional cases are

$$Q_{1d} = q_{A1}^{N_{A1}} q_{B1}^{N_{B1}} \text{ and } q_{A2}^{N_{A2}} q_{B2}^{N_{B2}} \tag{A-15}$$

and

$$Q_{3d} = q_{A1}^{N_{A1}} q_{B1}^{N_{B1}} \frac{N_1!}{N_{A1}!N_{B1}!} \text{ and } q_{A2}^{N_{A2}} q_{B2}^{N_{B2}} \frac{N_2!}{N_{A2}!N_{B2}!} \tag{A-16}$$

respectively. In the one-dimensional case at some temperature

$$\mu_{A1} = \mu_{A1}' \tag{A-17}$$

and

$$\mu_{A2} = \mu_{A2}' \tag{A-18}$$

The free energy per mole in state 1 at some temperature is

$$\frac{G_1}{N} = \frac{N_A\mu_{A1} + N_B\mu_{B1}}{N} = f_A\mu_{A1}' + f_B\mu_{B1}' \tag{A-19}$$

with a similar expression for state 2. Since the change in free energy per mole is zero at T_{TR}

$$0 = \frac{G_1 - G_2}{N} = f_A\Delta\mu_A' + f_B\Delta\mu_B' \tag{A-20}$$

From eq A-11 and A-20, and the relation $f_B = 1 - f_A$, we obtain

$$f_A = \frac{\displaystyle\int_{T_{TR}}^{T_B} \frac{\Delta H_B}{T^2}\mathrm{d}T}{\displaystyle\int_{T_{TR}}^{T_B} \frac{\Delta H_B}{T^2}\mathrm{d}T + \int_{T_A}^{T_{TR}} \frac{\Delta H_A}{T^2}\mathrm{d}T} \tag{A-21}$$

which is eq 12. In the three-dimensional case, when $\sigma \rightarrow 0$

$$\begin{aligned}\mu_{A1} &= \mu_{A1}' + RT\ln\frac{N_{A1}}{N_1}\\ \mu_{A2} &= \mu_{A2}' + RT\ln\frac{N_{A2}}{N_2}\end{aligned} \tag{A-22}$$

and hence, at the transition temperature

$$\Delta\mu_A'(T_{TR}) = \mu_{A1}' - \mu_{A2}' = RT_{TR}\ln\frac{N_{A2}}{N_{A1}} = RT_{TR}\ln\frac{Y_A}{X_A} \tag{A-23}$$

This expression is identical with eq A-8. The derivation then follows the same steps from eq A-8 to eq A-14, leading to the result that the relation between f_A and T_{TR} is described by eq A-14 for three-dimensional systems as $\sigma \rightarrow 0$.

In summary, eq A-14 applies to the one-dimensional case with $\sigma \rightarrow 1$ and to the three-dimensional cases with $\sigma \rightarrow 0$ and $\sigma \rightarrow 1$; eq A-21 applies to the one-dimensional case with $\sigma \rightarrow 0$. The dominant factor in determining whether the relation between f_A and T_{TR} should be described by eq A-14 (sigmoidal) or by eq A-21 (linear) in the four cases considered here is the presence or absence of an entropy of mixing in the system, arising from the combinatorial expressions in eq A-2, A-3, and A-16. Whether the entropy is that of mixing the states 1 and 2 (in one dimension, for $\sigma \rightarrow 1$), of mixing the species A and B (in three dimensions, for $\sigma \rightarrow 0$), or of mixing both states and species (in three dimensions, for $\sigma \rightarrow 1$) seems to be immaterial. The only case completely devoid of an entropy of mixing is the one-dimensional one (with $\sigma \rightarrow 0$), for which the T_{TR} *vs.* f_A curve is linear.

Appendix B. Calculations with the LAPS Hierarchy for Orders of Approximation m = 0 and 1 in Finite Chains

In this Appendix, we present the details of the calculation of the melting curves of random copolymers of finite chain length N for the two lowest orders (m = 0 and 1) of the LAPS hierarchy. For these orders, the function $Z_1^\dagger$ can be represented, in terms of the following modification of the 2 × 2 matrix of eq 22

$$\mathbf{W} = \begin{pmatrix}(\alpha s_A + s_B) & (\alpha + 1)\\ (\alpha\sigma_A s_A + \sigma_B s_B) & (\alpha + 1)\end{pmatrix} \tag{B-1}$$

For the case of $m = 0$, α is p_A/p_B. For the case of $m = 1$ (adjustable α), we have omitted any normalization factors from the "denominators" of eq 22, since these only alter the value of α, which is adjusted anyway to satisfy eq 26. Thus, for $m = 0$ and 1

$$Z_1^\dagger = \mathbf{e}\mathbf{W}^N\mathbf{e}^+ = C_1\lambda_1^N + C_2\lambda_2^N \tag{B-2}$$

where

$$\mathbf{e} = (0, 1) \qquad \mathbf{e}^+ = \begin{pmatrix}1\\1\end{pmatrix} \tag{B-3}$$

and λ_1 and λ_2 are the eigenvalues of the matrix of eq B-1, *viz.*

$$\lambda_{1,2} = (^1/_2)\{(\alpha + 1) + (\alpha s_A + s_B) \pm [\{(\alpha + 1) + (\alpha s_A + s_B)\}^2 - 4(\alpha + 1)\{(\alpha s_A + s_B) - (\alpha\sigma_A s_A + \sigma_B s_B)\}]^{1/2}\} \tag{B-4}$$

The C's are factors given, according to a recipe formulated elsewhere,[35] by

$$C_{1,2} = \frac{(\alpha\sigma_A s_A + \sigma_B s_B)\lambda_{1,2}}{[\lambda_{1,2} - (\alpha + 1)]^2 + (\alpha + 1)(\alpha\sigma_A s_A + \sigma_B s_B)} \tag{B-5}$$

The fraction of helix states is then given explicitly by eq B-6,

$$\theta_h = (1/N)(\partial\ln Z_1^\dagger/\partial\ln s) = \frac{\dfrac{1}{NC_1}\left[\dfrac{\partial C_1}{\partial\ln s} + \dfrac{\partial C_2}{\partial\ln s}\left(\dfrac{\lambda_2}{\lambda_1}\right)^N\right] + \dfrac{\partial\ln\lambda_1}{\partial\ln s} + \dfrac{\partial\ln\lambda_2}{\partial\ln s}\dfrac{C_2}{C_1}\left(\dfrac{\lambda_2}{\lambda_1}\right)^N}{1 + \dfrac{C_2}{C_1}\left(\dfrac{\lambda_2}{\lambda_1}\right)^N} \tag{B-6}$$

where $\partial\ln\lambda_{1,2}/\partial\ln s$ and $\partial C_{1,2}/\partial\ln s$ are given by eq B-7.

The theoretical fraction of A present is given explicitly by eq B-8, and the derivatives of C and λ with respect to α are obtained in a similar straightforward manner.

There are two ways of treating α; it can either be given by eq 24 (the Lifson[13] approximation; $m = 0$) or else be adjusted (the Allegra[17] approximation; which is equivalent to the case of $m = 1$ of the LAPS hierarchy) to whatever value is necessary to have the theoretical fraction of A, F_A, become equal to the experimental fraction of A, f_A. The accuracy of the approximation of order $m = 0$ and $m = 1$, of course, is determined by examining the convergence of the LAPS hierarchy when calculations are carried out for larger values of m. This question is considered in the text of the paper.

$$\begin{aligned}\frac{\partial\ln\lambda_{1,2}}{\partial\ln s} &= \frac{1}{2\lambda_{1,2}}\left\{(\alpha s_A + s_B) \pm \frac{(\alpha s_A + s_B)[(\alpha s_A + s_B) - (\alpha + 1)] + 2(\alpha\sigma_A s_A + \sigma_B s_B)(\alpha + 1)}{[\{(\alpha s_A + s_B) - (\alpha + 1)\}^2 + 4(\alpha + 1)(\alpha\sigma_A s_A + \sigma_B s_B)]^{1/2}}\right\}\\ \frac{\partial C_{1,2}}{\partial\ln s} &= C_{1,2}\left\{1 - \frac{(\alpha + 1)}{\lambda_{1,2}}C_{1,2} + \frac{\partial\ln\lambda_{1,2}}{\partial\ln s} - 2\,\frac{C_{1,2}[\lambda_{1,2} - (\alpha + 1)]}{(\alpha\sigma_A s_A + \sigma_B s_B)}\,\frac{\partial\ln\lambda_{1,2}}{\partial\ln s}\right\}\end{aligned} \tag{B-7}$$

$$F_A = (1/N)(\partial\ln Z_1^\dagger/\partial\ln\alpha) = \frac{\dfrac{1}{NC_1}\left[\dfrac{\partial C_1}{\partial\ln\alpha} + \dfrac{\partial C_2}{\partial\ln\alpha}\left(\dfrac{\lambda_2}{\lambda_1}\right)^N\right] + \dfrac{\partial\ln\lambda_1}{\partial\ln\alpha} + \dfrac{\partial\ln\lambda_2}{\partial\ln\alpha}\dfrac{C_2}{C_1}\left(\dfrac{\lambda_2}{\lambda_1}\right)^N}{1 + \dfrac{C_2}{C_1}\left(\dfrac{\lambda_2}{\lambda_1}\right)^N} \tag{B-8}$$

---

(35) Reference 32, p 304.