On the Melting of Deoxyribonucleic Acid as a Heterogeneous Chain

Seiji Tanaka and Akio Nakajima*

Department of Polymer Chemistry, Kyoto University, Kyoto 606, Japan.

Received February 4, 1972

ABSTRACT: Exact thermodynamic analyses of the helix–coil transition have been performed for heterogeneous chains with random or Markoffian sequence distributions, by a Monte Carlo calculation. The transition breadth with respect to temperature has been shown to depend strongly on both the base composition and sequence distribution. It was also shown that the shift of melting temperature is markedly dependent on the sequence distribution in the case of a Markoffian chain.

Helix–coil transitions in polynucleotides and nucleic acids have been studied theroetically by a number of investigators.[1–16] Several authors have extended the theory to heterogeneous chains, such as a DNA molecule consisting of two kinds of nucleotide pairs with different stabilities. The transition breadth with respect to temperature for the helix–coil transition has been shown to depend strongly on the base composition.[9,12,13] Fink, *et al.*,[14] have performed Monte Carlo calculations to obtain the melting curve of heterogeneous polymers.

In the present study, we investigate the helix–coil transition of heterogeneous copolymers by more complete Monte Carlo calculation than those of the previous authors.[12–14] Namely, we use the direct calculation method instead of numerical differentiation to obtain the transition breadth, and in addition bring in an exact treatment for the sequence distribution.

The DNA from an organism, *e.g.*, DNA from *E. coli* or calf thymus, reveals a much wider transition breadth than that theroetically predicted for heterogeneous polynucleotides with random sequences.[15] We discuss the effect of sequence distribution on the melting behavior of DNA by exact treatment of the sequence distribution under the assumption of Markoffian chain.

In treating the melting of polynucleotides in the form of double-stranded helices another important feature may be loop formation in the coil region. But this paper is not concerned with this problem.

Model and Mathematical Treatment

The model of DNA employed here is that proposed by Reiss, McQuarrie, McTague, and Cohen[9] (RMMC), originating from the theory of Zimm and Bragg[17] for helix–coil transitions of polypeptides. In this model, the completely nonbonded state of the polynucleotide (neglecting strand separation) is taken as the standard state. DNA is a copolymer consisting of two types of base pair, *i.e.*, adenine-thymine (A-T) and guanine-cytosine (G-C), which have different free energies of bonding. The model proposed by RMMC represents DNA as a linear sequence of A-T and G-C base pairs. By assuming that A-T and G-C are respectively equivalent to T-A and C-G, DNA can be treated as a copolymer consisting of two kinds of unit and having a sequence such as AABABBBAA . . . , where A and B denote A-T and G-C, respectively. The nucleation parameter σ was, in the original theory of Zimm and Bragg,[17] assumed to be independent of the number of preceding unbonded units. Furthermore, the value of σ has, in the heterogeneous chain model, been assumed to be independent of both the base-pair composition and sequence. The internal partition functions S_{A} and S_{B} are employed for bonded A-T and G-C groups. According to RMMC, we can get the partition function for a DNA molecule consisting of N base pairs with a specified base-pair sequence (neglecting end effects), as

$$Z = \mathbf{e}^*\left[\prod_{i=1}^{N} \mathbf{M}_{\mathrm{m};i}\right]\mathbf{e} \tag{1}$$

where $\mathbf{M}_{\mathrm{m};i}$ is a matrix operator for ith unit, given by

$$\mathbf{M}_{\mathrm{m};i} = \begin{bmatrix} 1 & \sigma S_{\mathrm{m};i} \\ 1 & S_{\mathrm{m};i} \end{bmatrix} = \begin{bmatrix} 1 & \sigma S_{\mathrm{m}} \\ 1 & S_{\mathrm{m}} \end{bmatrix}_i \tag{2}$$

and $\mathbf{e}^*$ and $\mathbf{e}$ are denoted by

$$\mathbf{e}^* = [1 \quad 0] \qquad \mathbf{e} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{3}$$

In eq 2, m denotes the base-pair species, *i.e.*, $S_m = S_{\mathrm{A}}$ or S_{B} for an A-T or G-C group, respectively.

The fraction θ_{h} of bonded bases for a specified base-pair sequence of A-T and G-C is given by

$$\theta_{\mathrm{h}} = N^{-1}\sum_{i=1}^{N} \frac{\partial \ln Z}{\partial \ln S_{\mathrm{m};i}} = N^{-1}Z^{-1}\sum_{i=1}^{N} \frac{\partial Z}{\partial \ln S_{\mathrm{m};i}} \tag{4}$$

In accordance with the mathematical method (matrix multiplication method) applied to the conformational statistics of polymer molecules by Flory[18] and to helix–coil equilibria by Flory and Miller,[19] one obtains the following equation

---

(1) S. A. Rice and A. Wada, *J. Chem. Phys.*, **29**, 233 (1959).
(2) J. H. Gibbs and E. A. DiMarzio, *ibid.*, **30**, 271 (1959).
(3) T. L. Hill, *ibid.*, **30**, 383 (1959).
(4) B. H. Zimm, *ibid.*, **33**, 1349 (1960).
(5) S. Lifson, *Biopolymers*, **1**, 25 (1963).
(6) M. Ozaki, M. Tanaka, and E. Teramoto, *J. Phys. Soc. Jap.*, **18**, 551 (1963).
(7) D. M. Crothers and B. H. Zimm, *J. Mol. Biol.*, **9**, 1 (1964).
(8) S. Lifson and G. Allegra, *Biopolymers*, **2**, 65 (1964).
(9) H. Reiss, D. A. McQuarrie, J. P. McTague, and E. R. Cohen, *J. Chem. Phys.*, **44**, 4567 (1966).
(10) E. Montroll and N. S. Goel, *Biopolymers*, **4**, 855 (1966).
(11) M. Go, *J. Phys. Soc. Jap.*, **23**, 597 (1967).
(12) G. W. Lehman and J. P. McTague, *J. Chem. Phys.*, **49**, 3170 (1968); G. W. Lehman, in "Statistical Mechanics," T. A. Bak, Ed., W. A. Benjamin, New York, N. Y., 1967.
(13) M. Fixman and D. Zeroka, *J. Chem. Phys.*, **48**, 5223 (1968).
(14) T. R. Fink and D. M. Crothers, *Biopolymers*, **6**, 863 (1968).
(15) D. Poland and H. A. Scheraga, "Theory of Helix-Coil Transition in Biopolymers," Academic Press, New York, N. Y., 1970.
(16) Y. U. S. Lazurkin, M. D. Frank-Kamenetskii, and E. N. Trifonov, *Biopolymers*, **9**, 1253 (1970).
(17) B. H. Zimm and J. K. Bragg, *J. Chem. Phys.*, **31**, 526 (1959).
(18) P. J. Flory, "Statistical Mechanics of Chain Molecules," Interscience, New York, N. Y., 1969.
(19) P. J. Flory and W. G. Miller, *J. Mol. Biol.*, **15**, 284 (1966).

$$\theta_{\mathrm{h}} = N^{-1}Z^{-1}[\mathbf{e}^* \ \mathbf{0}]\left[\prod_{i=1}^{N} \mathbf{M}^*_{\mathrm{m},S;i}\right]\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \tag{5}$$

where

$$\mathbf{M}^*_{\mathrm{m},S;i} = \begin{bmatrix}\mathbf{M}_{\mathrm{m}} & \mathbf{M}'_{\mathrm{m},S}\\ \mathbf{0} & \mathbf{M}_{\mathrm{m}}\end{bmatrix}_i \tag{6}$$

and

$$\mathbf{M}'_{\mathrm{m},S} = \partial\mathbf{M}_{\mathrm{m}}/\partial \ln S_{\mathrm{m}} \tag{7}$$

and hereafter **0** means a null vector or matrix of an appropriate order to fill out the expressions.

The derivative of the fraction θ_{h} of bonded base (helical) with respect to temperature is formulated as follows

$$\frac{\mathrm{d}\theta_{\mathrm{h}}}{\mathrm{d}T} = (N^{-1}Z^{-1})[\mathbf{e}^* \ \mathbf{0}]\left[\prod_{i=1}^{N} \mathbf{M}^{**}_{\mathrm{m},S,T;i}\right] \times \begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} - (\theta_{\mathrm{h}}Z^{-1})\frac{dZ}{\mathrm{d}T} \tag{8}$$

Z and θ_{h} in eq 8 are given by eq 1 and eq 4, respectively. The temperature coefficient of the partition function Z is expressed by

$$\frac{\mathrm{d}Z}{\mathrm{d}T} = [\mathbf{e}^* \ \mathbf{0}]\left\{\mathbf{M}'_{\mathrm{m},T;1}\left[\prod_{i=2}^{N} \mathbf{M}_{\mathrm{m};i}\right] + \mathbf{M}_{\mathrm{m};1}\mathbf{M}'_{\mathrm{m},T;2} \times \left[\prod_{i=3}^{N} \mathbf{M}_{\mathrm{m};i}\right] + \cdots + \left[\prod_{i=1}^{N-1} \mathbf{M}_{\mathrm{m};i}\right]\mathbf{M}'_{\mathrm{m},T;N}\right\}\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \tag{9}$$

where

$$\mathbf{M}'_{\mathrm{m},T;i} = \mathrm{d}\mathbf{M}_{\mathrm{m};i}/\mathrm{d}T \tag{10}$$

According to the matrix multiplication method proposed by Flory, the summation in eq 9 is expressed, in more convenient form, as

$$\frac{\mathrm{d}Z}{\mathrm{d}T} = [\mathbf{e}^* \ \mathbf{0}]\left[\prod_{i=1}^{N} \mathbf{M}^*_{\mathrm{m},T;i}\right]\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \tag{11}$$

where

$$\mathbf{M}^*_{\mathrm{m},T;i} = \begin{bmatrix}\mathbf{M}_{\mathrm{m}} & \mathbf{M}'_{\mathrm{m},T}\\ \mathbf{0} & \mathbf{M}_{\mathrm{m}}\end{bmatrix}_i \tag{12}$$

The first term in the right-hand side of eq 8 corresponds to

$$(N^{-1}Z^{-1})[\mathbf{e}^* \ \mathbf{0}]\left\{\mathbf{M}^{*\prime}_{\mathrm{m},S,T;1}\left[\prod_{i=2}^{N} \mathbf{M}^*_{\mathrm{m},S;i}\right] + \mathbf{M}^*_{\mathrm{m},S;1}\mathbf{M}^{*\prime}_{\mathrm{m},S,T;2}\left[\prod_{i=3}^{N} \mathbf{M}^*_{\mathrm{m},S;i}\right] + \cdots + \left[\prod_{i=1}^{N-1} \mathbf{M}^*_{\mathrm{m},S;i}\right]\mathbf{M}^{*\prime}_{\mathrm{m},S,T;N}\right\}\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \tag{13}$$

where

$$\mathbf{M}^{*\prime}_{\mathrm{m},S,T;i} = \mathbf{dM}^*_{\mathrm{m},S;i}/\mathrm{d}T \tag{14}$$

and $\mathbf{M}^*_{\mathrm{m},S;i}$ being given by eq 6.

By constructing a new matrix from eq 6 and 14, as given by

$$\mathbf{M}^{**}_{\mathrm{m},S,T;i} = \begin{bmatrix}\mathbf{M}^*_{\mathrm{m},S} & \mathbf{M}^{*\prime}_{\mathrm{m},S,T}\\ \mathbf{0} & \mathbf{M}^*_{\mathrm{m},S}\end{bmatrix}_i \tag{15}$$

we can rewrite the summation in eq 13 in the form of matrix multiplication as expressed by the first term on the right-hand side of eq 8. Thus, we are able to obtain all the terms appearing in eq 8.

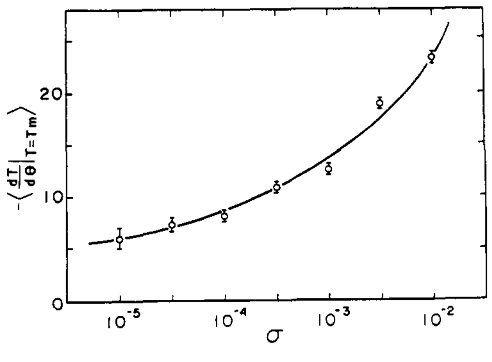

Figure 1. Dependence of transition breadth $\langle \mathrm{d}T/\mathrm{d}\theta|_{T=T_{\mathrm{m}}}\rangle$ on the cooperativity parameter σ for equimolar base composition. Error bars denote the standard deviation for ten lattices consisting of 5001 lattice sites. Sequence distribution is random.

The average number $\bar{N}_{\mathrm{h}}$ of bonded bases for a specified base-pair sequence is given by

$$\bar{N}_{\mathrm{h}} = \sum_{i=1}^{N} \frac{\partial \ln Z}{\partial \ln S_{\mathrm{m};i}} = N\theta_{\mathrm{h}} \tag{16}$$

and the average number $\bar{N}_{\mathrm{hc}}$ of bonded base-pair sequences is

$$\bar{N}_{\mathrm{hc}} = \sum_{i=1}^{N} \frac{\partial \ln Z}{\partial \ln \sigma_i} \tag{17}$$

Equation 17 is also rewritten by using the method of Flory, *et al.*,[18,19] as

$$\bar{N}_{\mathrm{hc}} = Z^{-1}[\mathbf{e}^* \ \mathbf{0}]\left[\prod_{i=1}^{N} \mathbf{M}^*_{\mathrm{m},\sigma;i}\right]\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \tag{18}$$

where

$$\mathbf{M}^*_{\mathrm{m},\sigma;i} = \begin{bmatrix}\mathbf{M}_{\mathrm{m}} & \mathbf{M}'_{\mathrm{m},\sigma}\\ \mathbf{0} & \mathbf{M}_{\mathrm{m}}\end{bmatrix}_i \tag{19}$$

and

$$\mathbf{M}'_{\mathrm{m},\sigma} = \partial\mathbf{M}_{\mathrm{m}}/\partial \ln \sigma \tag{20}$$

By using eq 16 and 18, we can obtain the average length $\bar{L}_{\mathrm{h}}$ of bonded base-pair sequences, for a molecule with a specified base-pair sequence, as

$$\bar{L}_{\mathrm{h}} = \bar{N}_{\mathrm{h}}/\bar{N}_{\mathrm{hc}} \tag{21}$$

The treatment referred to above is a statistical mechanical averaging for a single molecule with a specified base sequence. Experimentally, we observe melting phenomena for a large number of DNA molecules with intramolecular heterogeneity. Hence an ensemble average (configurational average) over all the molecular species in the system must be carried out when we calculate the average fraction $\langle\theta\rangle$ of bonded bases and other thermodynamic quantities for polymers with intramolecular heterogeneity. In this paper, we use $\bar{N}_{\mathrm{h}}$ and $\bar{L}_{\mathrm{h}}$, respectively, in order to designate the statistical mechanical averages of N_{h} and L_{h} over a molecule in the system, while $\langle\theta_{\mathrm{h}}\rangle$ and $\langle\bar{L}_{\mathrm{h}}\rangle$, respectively, designate the ensemble averages of θ_{h} and $\bar{L}_{\mathrm{h}}$ in the system.

The transition behavior of heterogeneous copolymers differs from that of homopolymers in some respects. One of the marked differences is the breadth of the helix-coil transition. The transition breadth is defined in terms of $\mathrm{d}T/\mathrm{d}\theta$, the reciprocal of the temperature coefficient in eq 8, to avoid divergency, at the melting point $T = T_{\mathrm{m}}$ or $\theta_{\mathrm{h}} = 0.5$. The average breadth for heterogeneous polymers is given, after an ensemble averaging, as

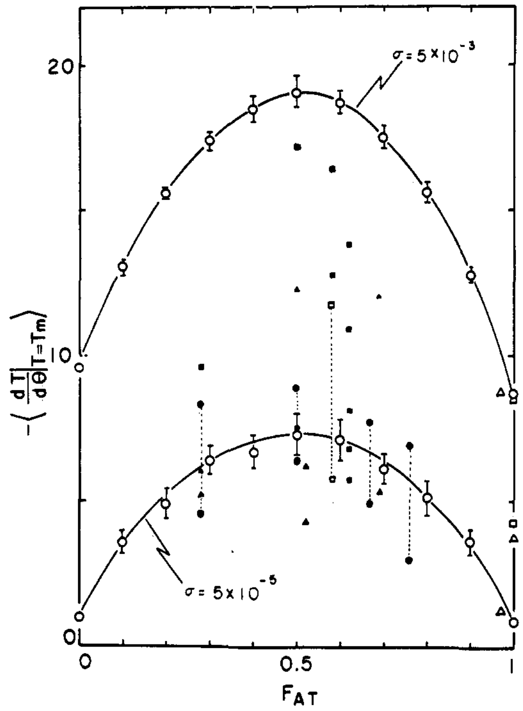

Figure 2. Dependence of transition breadth on the base composition. Upper curve (○) is for $\sigma = 5 \times 10^{-3}$ and lower one (○) for $\sigma = 5 \times 10^{-5}$. Error bars denote the standard deviation of ensemble average for ten lattices consisting of 5001 lattice sites with random sequence distribution. Experimental points are taken from the works of Owen, *et al.* (●---●); the range shown by broken line was obtained under various salt concentrations); Davidson, *et al.* (△); Cruenwedel, *et al.* (□ and □---□); and Dove, *et al.* (■); and Vinograd, *et al.* (▲).

$$\left\langle \frac{dT}{d\theta}\Big|_{T=T_m} \right\rangle = \left\langle \begin{matrix}\text{the right-hand side of eq 8} \\ \text{at } T = T_m \text{ and } \theta_h = \theta_{T=T_m}\end{matrix} \right\rangle \quad (22)$$

where $\theta_{T=T_m}$ is the average fraction of bonded bases at $T = T_m$ (T_m is the actual melting temperature). This temperature T_m, at which θ_h must be 0.5, is different from $T_m{}^0$ (calculated from eq 28) as will be discussed in the following.

Method of Calculation

Choice of Parameters. We employ the set of parameters proposed by Lehman[12] for illustrative calculation to compare the results of machine calculation with those of experiment on the thermodynamic properties of DNA. According to Lehman, the melting temperatures (T_{AT} and T_{GC}) of pure AT and GC polymers, and enthalpies (ΔH_{AT} and ΔH_{GC}) of formation of pure AT and GC polymers are

$$\begin{aligned} \Delta H_{AT} &= -7600 \text{ cal/mol} \\ T_{AT} &= 342.5\text{°K} \\ \Delta H_{GC} &= -8600 \text{ cal/mol} \\ T_{GC} &= 383.5\text{°K} \end{aligned} \quad (23)$$

With these parameters, the equilibrium constants S_{AT} and S_{GC} are given, by the use of the van't Hoff equation, as

$$S_m = \exp\{(\Delta H_m/R)(1/T_m) - (1/T)\} \quad (24)$$

where R is the gas constant and the subscript m denotes AT or GC.

Construction of $\mathbf{M}'_{m,T;i}$ and $\mathbf{M}^{*\prime}_{m,S,T;i}$ (eq 10 and 14) requires the temperature coefficient of S_m, represented by S'_m; this is obtained from eq 24 by differentiating S_m with respect to temperature. After Zimm and Bragg,[17] we assume that σ does not depend on temperature[20] as a first-order approximation. This assumption yields

$$d(\sigma S_m)/dT = \sigma S'_m \quad (25)$$

Hence, for example, eq 10 becomes

$$\mathbf{M}'_{m,T;i} = \begin{bmatrix} 0 & \sigma S'_m \\ 0 & S'_m \end{bmatrix} \quad (26)$$

Monte Carlo Calculation. Heterogeneous lattices consisting of N binary lattice sites are realized by a random sampling technique using a digital computer. The lattice with a Markoffian distribution is exactly generated according to the sequence probability (conditional probability) matrix **P** as discussed in detail in our previous treatment[21]

$$\mathbf{P} = \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix} \quad (27)$$

by assuming that **P** is independent of i. Using eq 1, 5, 8, and 21, the partition function, the helical content, and the average helical length are calculated in a straightforward fashion for the heterogeneous lattice thus generated by means of an appropriate vector-matrix multiplication method. This Monte Carlo procedure must be repeated until the deviation from the mean value falls within an acceptable statistical limit. It should be noted that the fences denoting the ensemble average do not mean the value obtained from the average of each term but the average value of those calculated for the individual lattice. In other words, $\langle \bar{L}_h \rangle$ is not given by $\langle \bar{N}_h \rangle / \langle \bar{N}_{hc} \rangle$, but by $\langle \bar{N}_h / \bar{N}_{hc} \rangle$. These procedures for configurational averaging are similar to our previous calculation of the unperturbed dimensions of stereoirregular diene polymers.[22]

Calculations were carried out with double precision using a FACOM 230-60 computer at the Kyoto University Computer Center. Unfiorm pseudorandom numbers were generated by using the subroutine examined by this center. Setting the parameters to those of homopolymer AT or GC as given in eq 23, the transition breadth was calculated for $N = 5001$ in order to check the accuracy of the computational program. Comparing the above results with the values calculated by an exact equation at infinite N, we found that both results agreed if the effect of lattice size was considered. Calculations were performed for 100–10 separate lattices consisting of 401–6001 lattice sites. The standard deviation α of $\langle \theta_h \rangle$ for the sample size $N = 5001$ and 10 lattices did not exceed 0.02. Computation time to obtain $\langle dT/d\theta|_{T=T_m} \rangle$ for a single set of σ and mole fraction of base pair (F_{AT} or F_{GC}) with random distribution varied from 15 to 20 min depending on the sample size.

Determination of Actual Melting Temperature

From the works of Marmur and Doty[23] and other authors,[9,12] it was known that the mean melting temperature $T_m{}^0$ for DNAs of different base compositions is a roughly linear function of the fraction of base pairs. This linear relation is expressed as

$$T_m{}^0 = F_{AT}T_{AT} + F_{GC}T_{GC} \quad (28)$$

---

(20) Recently, M. Go, N. Go, and H. A. Scheraga, *J. Chem. Phys.*, **52**, 2060 (1970), derived the temperature dependence of σ by theoretical calculation for parameters S and σ. The effect of the temperature dependence of σ on the helix–coil transition will be discussed elsewhere.

(21) S. Tanaka and A. Nakajima, *Polym. J.*, **2**, 725 (1971).

(22) S. Tanaka and A. Nakajima, *ibid.*, in press.

(23) J. Marmur and P. Doty, *J. Mol. Biol.*, **5**, 109 (1962).

TABLE I

$T_m{}^0$, T_m, AND NUMERICAL VALUES OF PARAMETERS AT $T_m{}^0$ AND T_m, FOR HETEROGENEOUS COPOLYMER WITH RANDOM SEQUENCE DISTRIBUTIONS, AND WITH $F_{GC} = 0.2$ AND $\sigma = 10^{-2}$ [a]

Parameter	Numerical value
$T_m{}^0\ (=F_{GC}T_{GC} + F_{AT}T_{AT})$, °K	350.70
$\langle\theta_h\rangle$ at $T_m{}^0$	0.487 (0.011)[b]
$\langle dT/d\theta\rangle$ at $T_m{}^0$	−19.62 (0.32)[b]
$\langle\Delta T_m\rangle$ at $T_m{}^0$	−0.26
$T_m\ (=T_m{}^0 + \langle\Delta T_m\rangle)$, °K	350.44
$\langle\theta_h\rangle$ at T_m	0.500 (0.011)[b]
$\langle dT/d\theta\rangle$ at T_m	−19.52 (0.31)[b]

[a] Sample size is ten lattices consisting of 5001 units. [b] Numerals in parentheses are standard deviations.

where F_{AT} and F_{GC} are the mole fractions of AT and GC, respectively. Equation 28 is effective only as a first-order approximation. The actual melting temperature T_m differs from the mean melting temperature $T_m{}^0$ obtained from eq 28. To calculate the transition breadth from eq 8 and 22, we must not use $T_m{}^0$ but T_m, as has been noted in the preceding section. Equation 8 is also useful to determine the actual melting temperature T_m as described below.

In the first place, the fraction of bonded bases $\theta_{T=T_m{}^0}$ and its temperature coefficient $d\theta/dT|_{T=T_m{}^0}$ at $T_m{}^0$ were calculated from eq 5 and 8. The deviation of $\theta_{T=T_m{}^0}$ from $\theta_h = 0.5$ is denoted by $\Delta\theta$

$$\Delta\theta = 0.5 - \theta_{T=T_m}{}^0 \qquad (29)$$

The melting curve may be regarded as being linear within a narrow range of $\Delta\theta$. Shift of melting temperatures $\Delta T_m(=T_m - T_m{}^0)$ is

$$\Delta T_m = \Delta\theta/(d\theta/dT|_{T=T_m}) \qquad (30)$$

After repeated computations for other heterogeneous lattices, we obtain the ensemble average, $\langle\Delta T_m\rangle$, of ΔT_m. Then the actual melting temperature averaged over all configurations is given by

$$\langle T_m\rangle = T_m{}^0 + \langle\Delta T_m\rangle \qquad (31)$$

It is necessary to repeat the above procedure if $\langle\theta_h\rangle$ calculated from eq 5 at $T = \langle T_m\rangle$ deviates from 0.5 beyond an acceptable statistical limit. But, practically, we obtain T_m, as shown in Table I, with enough accuracy without repetition. Less cooperativity (larger σ) yielded larger $\langle\Delta T_m\rangle$. According to another approximation method proposed by Fixman and Zeroka,[13] ΔT_m has a maximum value at a mole fraction of base pair close to 0.2 when the sequence distribution is random. In Table I the computational results for $F_{GC} = 0.2$ and $\sigma = 10^{-2}$ were shown, at which larger deviation in θ_h and T_m must be expected. As shown in Table I, the transition breadth is corrected by only 0.5% which is within the statistical accuracy of a Monte Carlo calculation. We may, therefore, regard $\langle dT/d\theta|_{T=T_m{}^0}\rangle$, instead of $\langle dT/d\theta|_{T=T_m}\rangle$, as the transition breadth under conditions of larger cooperativity ($\sigma < 10^{-2}$) and of random distribution (but not for the case of a Markoffian distribution, as discussed later).

Results and Discussion

Results for Random Sequence Distribution. As described in an early section of this paper, the transition of heterogeneous copolymers has been found to be broader than that of homopolymers. The temperature coefficient of helical content $d\theta/dT$ for homopolymers is proportional to $\sigma^{-1/2}$ at infinite chain length. However, $\langle dT/d\theta|_{T=T_m}\rangle$ is an approximately logarithmic function of the cooperativity parameter σ for heterogeneous copolymer with random sequence distribution as shown in Figure 1. The conditions of the calculation are indicated in the figure caption.

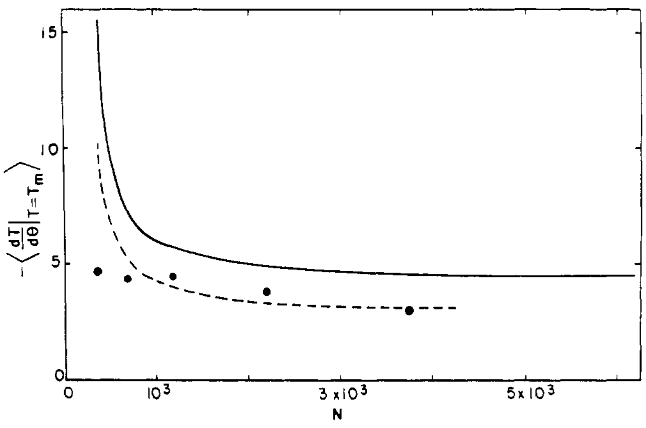

Figure 3. Dependence of transition breadth on degree of polymerization N. The curves are calculated for the base composition of T2 bacteriophage and the cooperativity parameter $\sigma = 10^{-6}$. The solid curve is for a random distribution and the broken one for a Markoffian distribution with sequence probability $P_{GC-GC} = 0.2$. Experimental data points by Crothers are also indicated by ●.

Another feature of the transition behavior of the copolymer has been found in the dependence of its breadth on the base composition. The result of the calculation for a random sequence distribution is shown in Figure 2 as a function of base composition. Calculation was carried out for two σ's by regarding them as adjustable parameters. Experimental points of several authors[24–28] are also shown in Figure 2. The experimental points occur over a wide range of σ due to salt concentration effects. Lehman and McTague[12] have found that the interfacial free energy is an approximately linear function of the logarithm of the ionic strength. The implication to be drawn from Figure 2 is that the breadth predicted by theory is a maximum for $F_{AT} = 0.5$, in qualitative agreement with the experimental data. This means that a simple physical interpretation of the broadening of the transition can be given by regarding σ as an adjustable parameter.

Similar agreement with the transition breadth has been found by Lehman and McTague[12] using another calculation method.

The dependence of the transition breadth on chain length was also investigated. The breadth calculated for T2 bacteriophage assuming a random sequence distribution is given as a function of degree of polymerization N by the solid line in Figure 3. Sample size was taken appropriately as 10 lattices for $N = 6001$ to 100 for $N = 401$. The experimental results by Crothers, *et al.*,[29] are also indicated in Figure 3.

Results for a Markoffian Sequence Distribution. DNA from organisms, for example, *E. coli* and calf thymus, has been found to have a much broader transition than that predicted theoretically for copolymers with a random sequence distribu-

---

(24) T. J. Owen, L. R. Hill and S. P. Lapaga, *Biopolymers*, **7**, 503 (1969).

(25) N. Davidson, *et al.*, *Proc. Nat. Acad. Sci. U. S.*, **53**, 111 (1965).

(26) D. W. Cruenwedel and C.-H. Hsu, *Biopolymers*, **7**, 557 (1969).

(27) W. F. Dove and N. Davidson, *J. Mol. Biol.*, **5**, 467 (1962).

(28) J. Vinograd, J. Leibowitz, and R. Watson, *ibid.*, **33**, 173 (1968).

(29) D. M. Crothers, N. R. Kallenbach, and B. H. Zimm, *J. Mol. Biol.*, **11**, 802 (1965).

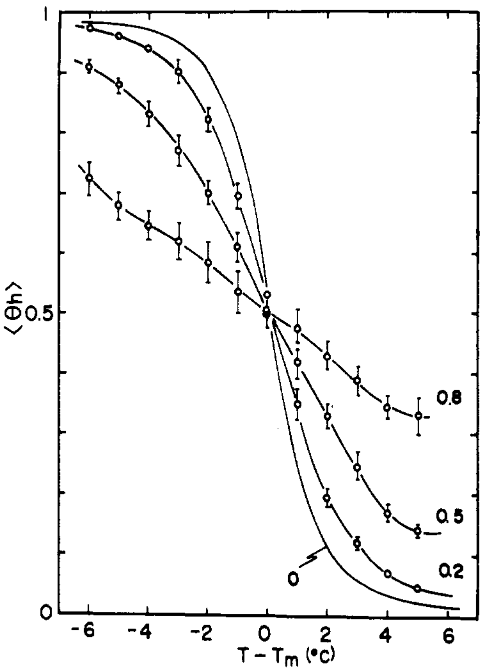

Figure 4. Theoretical denaturation curves for an equimolar base composition with a Markoffian distribution. Sequence probability $f = P_{\mathrm{GC-GC}} = P_{\mathrm{AT-AT}}$ is given by the numerals in the figure. Sample size is ten lattices consisting of 5001 lattice sites, and $\sigma = 5 \times 10^{-4}$. Error bars indicate the standard deviations.

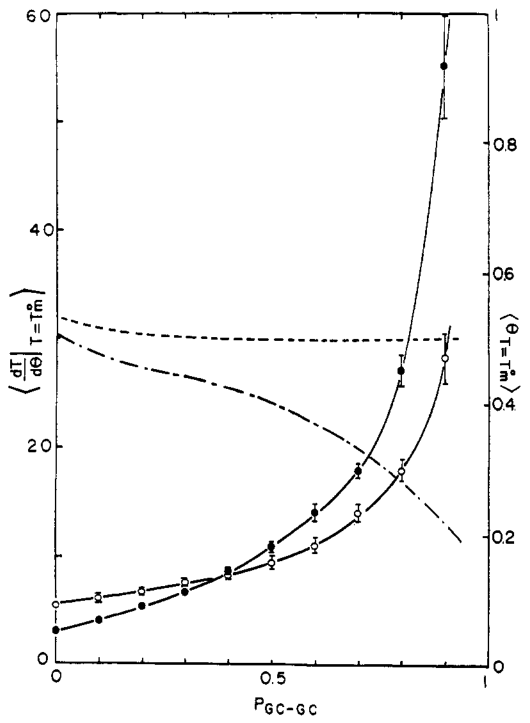

Figure 5. $\langle dT/d\theta|_{T=T_m{}^0}\rangle$ was plotted against a $P_{\mathrm{GC-GC}}$ for a Markoffian chain with $F_{\mathrm{AT}} = 0.9$ (○), and $F_{\mathrm{AT}} = 0.5$ (●): (————) $\langle\theta_{T=T_m{}^0}\rangle$ *vs.* $P_{\mathrm{GC-GC}}$ for $F_{\mathrm{AT}} = 0.9$, (----) $F_{\mathrm{AT}} = 0.5$. Calculations were performed for $\sigma = 5 \times 10^{-4}$ and ten lattices consisting of 5001 lattice sites. Error bars denote the standard deviation of ensemble average.

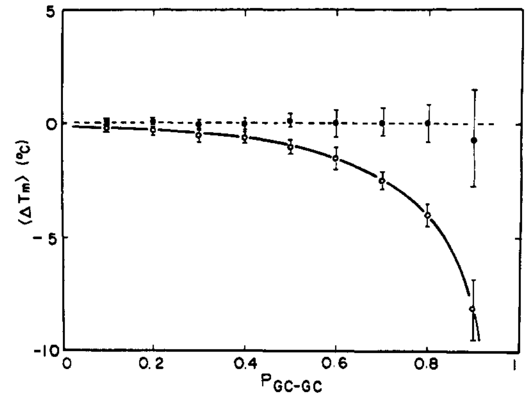

Figure 6. $\langle\Delta T_m\rangle$ *vs.* sequence probability $P_{\mathrm{GC-GC}}$ for Markoffian chain with $F_{\mathrm{AT}} = 0.5$ (●), and $F_{\mathrm{AT}} = 0.9$ (○). Sample size and σ are the same as those of Figure 5. · Error bars denote the standard deviation.

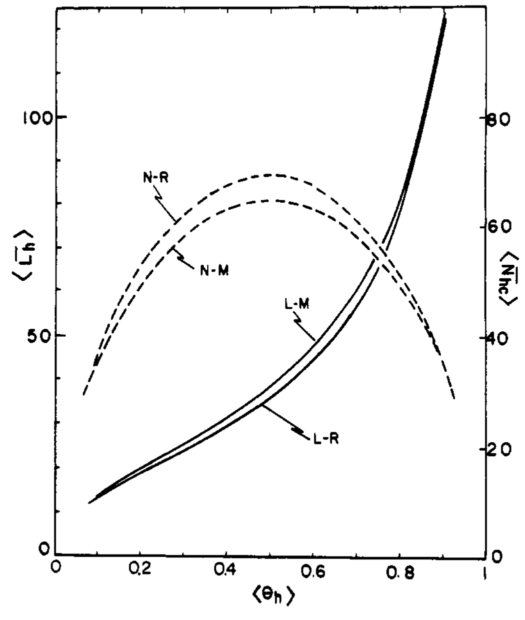

Figure 7. The average number $\langle\bar{N}_{\mathrm{hc}}\rangle$ of helical sequences (---) and the average length $\langle\bar{L}_{\mathrm{h}}\rangle$ of helical sequences (—) plotted against the average helical fraction $\langle\theta_{\mathrm{h}}\rangle$. The broken lines are for a random distribution (N-R) and for a Markoffian distribution with $P_{\mathrm{GC-GC}} = 0.2$ (N-M). The solid lines are for random (L-R) and for Markoffian distribution with $P_{\mathrm{GC-GC}} = 0.2$ (L-M). Sample size is ten lattices consisting of 5001 lattice sites, and σ is 5×10^{-4}.

tion.[16] In fact, it has been reported[30] that the sequence distribution of several types of DNA differed from the random distribution. Therefore, we next investigated the effect of correlated sequence distribution on the melting behavior. Base-pair sequence distribution was dealt with directly and exactly assuming a Markoffian chain which can be characterized by the transition probability given in eq 27. Theoretical denaturation profiles for selected sequence probabilities are plotted in Figure 4. The melting curve for $f = 0$ is that for an alternating copolymer and that for $f = 0.5$ is for a random one.

In Figure 5, the breadth $\langle dT/d\theta|_{T=T_m{}^0}\rangle$ is plotted against $P_{\mathrm{GC-GC}}$, the sequence probability for a GC base pair followed by a GC base pair. It should be noted that the temperature used in the calculation was not the actual melting temperature T_m but the melting temperature $T_m{}^0$. The fraction $\langle\theta_{T=T_m{}^0}\rangle$ of bonded base pairs at $T_m{}^0$ is also indicated in the same figure. As seen from the figure, the curve for $F_{\mathrm{AT}} = 0.5$ gives the

---

(30) J. Josse, A. D. Kaiser, and A. Kornberg, *J. Biol. Chem.*, 236, 864 (1961).

transition breadth at $\theta = 1/2$ in good approximation, but not for $F_{AT} = 0.9$. One can see a marked effect of base-pair distribution on the transition breadth.

The shift in the melting temperature expressed in eq 31 is illustrated in Figure 6 as a function of sequence probability. For a copolymer with a random sequence distribution, the actual melting temperature T_m differed from the mean melting temperature $T_m{}^0$ by an experimentally insignificant and negligible amount, as pointed out by Fixman and Zeroka.[13] For a copolymer with a Markoffian sequence distribution with $F_{AT} = 0.5$, $\langle \Delta T_m \rangle$ does not depend on the sequence probability. But, on the other hand, for $F_{AT} = 0.9$, a marked effect on the shift was found, as indicated in Figure 6.

The molecular weight dependence of the transition breadth for T2 DNA assuming a Markoffian distribution ($P_{GC-GC} = 0.2$) and $\sigma = 10^{-6}$ is given by the broken curve in Figure 3. Figure 3 indicates that the theoretical curve obtained by assuming a Markoffian distribution is more satisfactorily in accord with experimental results than that obtained by assuming a random distribution.

We are concerned with other interesting quantities that can be calculated from the partition function. The average number of helical sequence $\langle \bar{N}_{hc} \rangle$ and the average length of a helical sequence $\langle \bar{L}_h \rangle$ are plotted in Figure 7 as a function of the average helical fraction $\langle \theta_h \rangle$. Calculations were carried out for molecules having a base composition corresponding to T2 DNA, by assuming two typical sequence distributions, *i.e.*, random and Markoffian ($P_{GC-GC} = 0.2$). The effect of sequence distribution on these quantities is found to be appreciable but not large.

Lastly, we will again mention that comparison of theroetical and experimental results should be confined to general trends, because the coil size dependence of σ may indicate a considerable effect on the transition of a heterogeneous copolymer, although it seems not to be so drastic as shown by Zimm[4] for homogeneous polynucleotides.

A Study of Films of Poly(γ-methyl L-glutamate) Adsorbed on Water Using Wave Damping and Other Methods

R. L. Shuler* and W. A. Zisman

Laboratory for Chemical Physics, Naval Research Laboratory, Washington, D. C. 20390. *Received April 3, 1972*

ABSTRACT: Simultaneous measurements of wave damping coefficient, film pressure, and surface potential difference as a function of area per monomer have given much information concerning the behavior and structure of a monolayer of poly(γ-methyl L-glutamate) (PMG) adsorbed at the air–water interface. At large monomer areas the PMG film is composed of widely separated polypeptide units in the α-helical conformation but, at about 18 Å^2 per monomer, the molecules come into close contact to form a uniform and nearly incompressible film. Upon further compression the film molecules are packed forcibly into their most compact state, where the average area per monomer is about 16.5 Å^2. Up to that point the film is completely reversible on expansion. The sole effect of compressing the film further is one of displacing segments of the polymer chain from the interface while the packing and orientation of the molecules remaining in the surface remain constant. If the external pressure is relieved at any point after molecules are expelled from the interface, the film exhibits hysteresis as it enters a lower energy state. After that state is attained, displaced molecules again enter the surface and, when all of the molecules are back on the surface once more, the film returns to its original state in which the monomer area is about 18 Å^2. Therefore, the gradual near-equilibrium compression and expansion behavior of the film is such as to exclude the possibility of a permanent change occurring in the polymer molecules as a result of compression.

The behavior of synthetic polypeptides when adsorbed at the air–water interface has been studied during the last decade to help elucidate the way the more complex proteins function in aqueous environments. Interesting studies at three laboratories have been carried out in this area on poly(γ-methyl L-glutamate) (PMG) which have led to a divergence of opinions about the form this molecule assumes when it is spread and compressed on aqueous substrates.[1–5] Isemura and Hamaguchi concluded from surface pressure[1] and viscosity[2] measurements that PMG was in the α_{II} conformation proposed by Ambrose and Hanby[6] when adsorbed on water. Malcolm's vigorous and persistent investigations have combined the traditional methods of surface chemistry[3] with spectral and deuterium-exchange studies.[4] He rationalized the results by proposing that: (a) PMG was in the α-helical structure of Pauling and Corey[7] when spread on water, (b) compression beyond a certain point caused the polymer molecules to form a multilayer, and (c) the structure of the film molecules was fixed on the surface of the water and was independent of the degree of compression. In addition, Malcolm[3] suggested that the difference in the spreading solvent used by him (chloroform) and by Isemura and Hamaguchi (pyridine) might account for the discordant results obtained in the two studies. Loeb's subsequent investigation[5] confirmed this suspicion by showing that the relative concentration of pyridine and chloroform in the spreading solvent can influence considerably the properties and structure of PMG films on water.

---

(1) T. Isemura and K. Hamaguchi, *Bull. Chem. Soc. Jap.*, **25**, 40 (1952).
(2) T. Isemura and K. Hamaguchi, *ibid.*, **27**, 125 (1954).
(3) B. R. Malcolm, *Polymer*, **7**, 595 (1966).
(4) B. R. Malcolm, *Proc. Roy. Soc., Ser. A*, **305**, 363 (1968).
(5) G. I. Loeb, *J. Colloid Interface Sci.*, **27**, 38 (1968).
(6) E. J. Ambrose and W. E. Hanby, *Nature (London)*, **163**, 483 (1949).
(7) L. Pauley and R. B. Corey, *Proc. Nat. Acad. Sci. U. S.*, **37**, 235 (1951).