ingly on the helix side. Water has thus clearly expressed its preference for helix over coil in these media, and the solubility of the same solid phase must therefore be even greater if helix is the solution phase. Thus, it must be true that the aqueous medium is an even better solvent for the acid form of the helix than for the acid form of the coil. Yet, uncharged PLGA helices are insoluble. It is possible that there are important differences in solvent interaction between a -COOH group and a -COO group, *i.e.*, between a carboxyl and a hypothetical *discharged* carboxylate. The A_2 we obtain at $I = \infty$ is for a polymer with the latter as side chain, not the former. We deem it much more likely, however, that the crystallinity of PLGA precipitates lies at the root of the difficulty. The term "good solvent" in polymer parlance compares the dissolved polymer with an amorphous precipitate. A solvent that is "good" to the tune of $A_2 = 5 \times 10^{-4}$ (mol cm^3)/g^2 may still not be good enough to prevail against the blandishments of crystallinity. The partially crystalline character of PLGA is well known, and has thus far foiled attempts to find a Θ solvent for this substance.[39]

Our estimated value of A_2 for infinite ionic strength solutions of randomly coiled PLGA may be compared to the measurements[46] of a related, nonionic water-soluble polypeptide, poly(N^5-(3-hydroxypropyl)·L-glutamine). It is found that at 25° this polymer has $A_2 = 2.8 \times 10^{-4}$, 3.0×10^{-4}, and 7.5×10^{-4} for samples with M_w equal respectively to 333×10^3, 39.1×10^3, and 18.7×10^3. The helix content of these samples under the indicated conditions decreases with decreasing molecular weight, being respectively 27, 20, and 13%. This would indicate that helix formation decreases A_2[47] and that A_2 is substantially positive for the uncharged random coil in water, as we find for PLGA.

The question arises of the consistency of our interpretation of the data on second virial coefficients with our analysis above of the molecular dimensions as obtained through the viscosity. The close connection between the two has been known, at least qualitatively, since the pioneering work of Flory.[33] Clearly, if water at infinite ionic strength is a good solvent (*i.e.*, $A_2 > 0$), then the expansion factor is greater than unity, and the intrinsic viscosity must be greater than it is in a Θ solvent. That is, we should find $[\eta]_\infty > [\eta]_\Theta$. Combining eq 3 and 4, we note that

$$[\eta]_\Theta = K_\Theta M^{1/2} = \Phi_\Theta b_\Theta{}^3 M^{1/2}/M_0{}^{1/2} \tag{15}$$

Using the value of K_Θ reported above (the value giving $b_\Theta \simeq 8$ Å), we indeed always find from eq 15 that $[\eta]_\infty > [\eta]_\Theta$. If an appreciably larger value of b_Θ (~10 Å) is tried, however, the $[\eta]_\Theta$ values are sometimes above and sometimes below the experimental values of $[\eta]_\infty$. If a value as high as 11.5 Å is used, the calculated values of $[\eta]_\Theta$ are in all cases larger than $[\eta]_\infty$. We therefore feel that the lower value of b_Θ (8 Å) is supported by the observed second virial coefficients.

(46) K. Okita, A. Teramoto, and H. Fujita, *Biopolymers*, **9**, 717 (1970).

(47) The sterically excluded volume [estimated according to B. Zimm, *J. Chem. Phys.*, **14**, 164 (1946)] would lead to a second virial coefficient for PLGA random coils that is over four times that expected for the α helix. The experimental second virial coefficient, however, is a result of a manifold of interactions in addition to the sterically excluded volume.

Statistics of Random Copolymers Containing Blocks of One Component

Eugene Helfand

Bell Laboratories, Murray Hill, New Jersey 07974. *Received January 4, 1972*

ABSTRACT: When long sequences (blocks) of one component occur in the molecules of a copolymer system, the blocks may segregate into domains. For a copolymer of A and O, we present a formula for the weight fraction of macromolecules which contain exactly *i* A blocks of length *k* or greater. A Monte Carlo estimate of this weight fraction and an analytical determination of the corresponding mole fraction have been given by Frensdorff. We employ an extension of the latter technique.

Frensdorff[1] has recently made an important contribution to the statistical theory of copolymers of A and O containing blocks (sequences) of one component. He addresses himself to the determination of the fraction of macromolecules which have exactly i A blocks of length k or greater. The problem is of interest because[1] "copolymer properties can be profoundly influenced by the presence of blocks . . . of one of the constituent monomers, especially if these blocks are long and numerous enough to segregate into separate phases of glassy or crystalline domains." A number of specific techniques have been developed to synthesize macromolecules with just such blocks. However, long sequences may form to a greater or lesser extent in random copolymers, too, depending on the kinetics of the addition process.

Frensdorff[1,2] has used a clever modification of the Markov transition matrix to determine a generating function. From this the (mole) fraction of macromolecules with i A blocks of length $\geq k$ has been analytically calculated. He notes, though, that a more significant quantity may be the corresponding weight fraction. The magnitude of the latter he has determined by a Monte Carlo procedure for a few values of the parameters. In the present paper the weight fraction of macromolecules with i A blocks of length $\geq k$ is expressed in algebraic form.

The notation of Frensdorff has been followed rather closely. For completeness, required quantities are briefly redefined but the reader is advised to peruse ref 1 for a deeper understanding.

(1) H. K. Frensdorff, *Macromolecules*, **4**, 369 (1971).

(2) H. K. Frensdorff and R. Pariser, *J. Chem. Phys.*, **39**, 2303 (1963).

Theory

Consider a copolymer of units A and O. The basic parameters which determine the statistics of the random copolymer are: (1) the four transition probabilities, P_{aa}, P_{ao}, P_{oa}, and P_{oo}, where $P_{\beta\gamma}$ is the conditional probability that a unit of type Γ will follow one of type B in the chain; (2) the termination probabilities, T_a and T_o; and (3) the initiation probabilities, I_a and I_o. These probabilities are related by the normalization conditions

$$\begin{aligned} I_a + I_o &= 1 \\ P_{aa} + P_{ao} + T_a &= 1 \qquad (1) \\ P_{oa} + P_{oo} + T_o &= 1 \end{aligned}$$

The type of average upon which our interest centers can be expressed in terms of the probability, $\psi(i, n, k)$, that a molecule in the system will consist of n units and have exactly i blocks of A that are of length $\geq k$. The (mole) fraction of macromolecules with i A blocks of length $\geq k$ is given by

$$X(i, k) = \sum_{n=1}^{\infty} \psi(i, n, k) \qquad (2)$$

while the weight fraction is

$$W(i, k) = (1/\bar{n}) \sum_{n=1}^{\infty} n\psi(i, n, k) \qquad (3)$$

(Equation 3 is precise if A and O have the same molecular weight. Otherwise, a slight correction must be made for the fact that a molecule with a long sequence of A's has a different average molecular weight than one without this property. This embellishment will not be considered here.) In eq 3, $\bar{n}$ denotes the number-average degree of polymerization

$$\bar{n} = \sum_{n,i} n\psi(i, n, k) \qquad (4)$$

We shall determine a generating function

$$F(s, z, k) = \sum_{n=1}^{\infty} \sum_{i=0}^{\infty} \psi(i, n, k) s^i z^n \qquad (5)$$

This may also be written in the form

$$F(s, z, k) = \sum_{i=0}^{\infty} F_i(z, k) s^i \qquad (6)$$

The $F_i(z, k)$ may be obtained from $F(s, z, k)$ by

$$F_i(z, k) = \frac{1}{i!} \frac{\mathrm{d}^i}{\mathrm{d}s^i} F(s, z, k) \Big|_{s=0} \qquad (7)$$

or, as we shall find convenient, by casting $F(s, z, k)$ into a form with well-known power series expansion. The quantities X and W may be seen to equal

$$X(i, k) = F_i(z = 1, k) \qquad (8)$$

$$W(i, k) = (1/\bar{n})[\mathrm{d}F_i(z, k)/\mathrm{d}z]_{z=1} \qquad (9)$$

$$\bar{n} = [\mathrm{d}F(s = 1, z, k)/\mathrm{d}z]_{z=1} \qquad (10)$$

Following Frensdorff,[1] we label the state of a chain in a special manner. When there is block of A's, the leftmost is labeled A_1, the second A_2, . . . , the $(k - 1)$st A_{k-1}, the kth A_k, and all subsequent ones A_k also. When the system terminates, we affix a state T to the right, and all subsequent states are T. Thus for $k = 3$ the chain[1]

AOAOOAAAOOOAAAAAOAAAAOO

becomes

$A_3A_1OA_1OOA_1A_2A_3OOOA_1A_2A_3A_3A_3OA_1A_2A_3A_3OOTTT\ldots$

T...

Construct a matrix $\mathbf{Q}(s, z, k)$ with rows and columns labeled with the indices B, Γ = O, A_1, A_2, . . . , A_{k-1}, A_k, T; which we abbreviate as $\beta, \gamma = 0, 1, 2, \ldots, k, t$

$$\mathbf{Q}(s, z, k) = \begin{array}{c|ccccccccc|} & 0 & 1 & 2 & 3 & \cdots & k-2 & k-1 & k & t \\ \hline 0 & zP_{oo} & zP_{oa} & 0 & 0 & \cdots & 0 & 0 & 0 & T_o \\ 1 & zP_{ao} & 0 & zP_{aa} & 0 & \cdots & 0 & 0 & 0 & T_a \\ 2 & zP_{ao} & 0 & 0 & zP_{aa} & \cdots & 0 & 0 & 0 & T_a \\ 3 & zP_{ao} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & T_a \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k-2 & zP_{ao} & 0 & 0 & 0 & \cdots & 0 & zP_{aa} & 0 & T_a \\ k-1 & zP_{ao} & 0 & 0 & 0 & \cdots & 0 & 0 & szP_{aa} & T_a \\ k & zP_{ao} & 0 & 0 & 0 & \cdots & 0 & 0 & zP_{aa} & T_a \\ t & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \\ \hline \end{array}. \qquad (11)$$

There is some ambiguity for $k = 1$. The correct structure is

$$\mathbf{Q}(s, z, 1) = \begin{array}{c|ccc|} & 0 & 1 & t \\ \hline 0 & zP_{oo} & szP_{oa} & T_o \\ 1 & zP_{ao} & zP_{aa} & T_a \\ t & 0 & 0 & 0 \\ \hline \end{array} \qquad (12)$$

The matrix $\mathbf{Q}(1, 1, k)$ is the transition matrix of Markov chain theory. The $\beta\gamma$ element of $[\mathbf{Q}(1, 1, k)]^m$ is the probability of going from state β to γ in m steps with units O, A, or T. The factor s on $[\mathbf{Q}(s, z, k)]_{k-1,k}$ serves as a marker and introduces a power of s every time a chain passes from a sequence of $(k - 1)$ A's to k A's. The z's help count the number of A or O units in the chain. Thus, the element βt of the matrix $\mathbf{Q}^m(s, z, k)$ is a polynomial in s and z. The coefficient of $s^i z^{n-1}$ is the probability of going from state β to termination with n A or O units and with i A blocks of length $\geq k$. However, n cannot be greater than m. To consider all n, the limit of $m \rightarrow \infty$ must be taken. Since the chain can start only in state O or A_1, with probabilities I_o and I_a, respectively, the generating function is given by

$$F(s, z, k) = \lim_{m \to \infty} \{I_o z[\mathbf{Q}^m(s, z, k)]_{0t} + I_a z[\mathbf{Q}^m(s, z, k)]_{1t}\}, \quad k \neq 1 \qquad (13)$$

$$F(s, z, I) = \lim_{m \to \infty} \{I_o z[\mathbf{Q}^m(s, z, 1)]_{ot} + I_a z s[\mathbf{Q}^m(s, z, 1)]_{1t}\} \tag{14}$$

As $m \to \infty$, one finds that $\mathbf{Q}^m$ has elements only in the column t. (Physically, this corresponds to the fact that the set of chains which do not terminate is immeasurable.) We label the elements of this column as vector **b**

$$\mathbf{b} = (b_0, b_1, \ldots, b_{k-1}, b_k, b_t) \tag{15}$$

Since a chain can go only to T from T, $b_t = 1$. The **b** vector is an eigenvector of **Q** corresponding to eigenvalue unity

$$\mathbf{Qb} = \mathbf{b} \tag{16}$$

The elements needed are ($k > 1$)

$$b_0 = \frac{T_o + z(P_{oa}T_a - P_{aa}T_o) - z^k P_{oa}P_{aa}{}^{k-1}T_a(1 - s)}{[1 - z(P_{oo} + P_{aa}) + z^2(P_{oo}P_{aa} - P_{oa}P_{ao}) + z^{k+1}P_{oa}P_{ao}P_{aa}{}^{k-1}(1 - s)]} \tag{17}$$

$$b_1 = \frac{[T_a - z(P_{oo}T_a - P_{ao}T_o)][1 - z^{k-1}P_{aa}{}^{k-1}(1 - s)]}{[1 - z(P_{oo} + P_{aa}) + z^2(P_{oo}P_{aa} - P_{oa}P_{ao}) + z^{k+1}P_{oa}P_{ao}P_{aa}{}^{k-1}(1 - s)]} \tag{18}$$

One could proceed to a calculation of the weight and mole fractions using

$$F(s, z, k) = I_o z b_0 + I_a z b_1 \qquad k > 1 \tag{19}$$

but the formulas soon become cumbersome. We leave the algebra involved in determining X and W to the concerned reader, and specialize to a simpler case, as Frensdorff[1] does.

The Case Where $T_a = T_o = T$

In the case where

$$T_a = T_o = T \tag{20}$$

a great deal of cancellation occurs. We also introduce as initiation probabilities

$$I_o = P_{ao}/\rho$$
$$I_a = P_{oa}/\rho$$
$$\rho = P_{oa} + P_{ao} \tag{21}$$

One finds

$$F(s, z, k) = \frac{Tz}{1 - z(1 - T)} - \frac{G(z)}{1 - z(1 - T)} + \frac{G(z)}{1 - z(1 - T) + E(z)(1 - s)} \tag{22}$$

$$G(z) = T(1 - zP_{aa})[1 + z(1 - T - P_{oo} - P_{aa})]/zP_{ao}\rho \tag{23}$$

$$E(z) = P_{oa}P_{ao}P_{aa}{}^{k-1}z^{k+1}/[1 + z(1 - T - P_{oo} - P_{aa})] \tag{24}$$

which happens to be correct for all k, including unity. The coefficient of s^i, called $F_i(z, k)$ in eq 6, is

$$F_0(z, k) = F(0, z, k) \tag{25}$$

$$F_i(z, k) = \frac{G(z)}{E(z)}\left[1 + \frac{1 - z(1 - T)}{E(z)}\right]^{-(i+1)} \qquad i \geq 1 \tag{26}$$

By setting $z = 1$, we obtain Frensdorff's formulas

$$X(0, k) = 1 - (G/T) + [G/(T + E)] \tag{27}$$

$$X(i, k) = (G/E)[1 + (T/E)]^{-(i+1)} \qquad i \geq 1 \tag{28}$$

$$G \equiv G(1) = T(1 - P_{aa})(\rho + T)/P_{ao}\rho \tag{29}$$

$$E \equiv E(1) = P_{oa}P_{ao}P_{aa}{}^{k-1}/(\rho + T) \tag{30}$$

TABLE I
WEIGHT FRACTION WITH ZERO AND WITH ONE A BLOCK OF LENGTH $\geq k$. COMPARISON OF EQ 32 AND 33 WITH FRENSDORFF'S[a] MONTE CARLO (MC) ESTIMATES

			$100W(0, k)$		$100W(1, k)$	
$\bar{n}$	k	X_a	Eq 32	MC[a]	Eq 33	MC[a]
100	1	0.8	0.00311	0.0032	0.317	0.315
100	2	0.8	0.0346	0.0323	0.541	0.554
100	3	0.8	0.119	0.122	0.907	0.915
100	4	0.8	0.295	0.273	1.48	1.42
100	5	0.8	0.619	0.604	2.35	2.34
100	10	0.8	7.98	7.83	14.2	14.0
100	20	0.8	67.4		24.8	
100	30	0.8	95.8		4.08	
100	40	0.8	99.56		0.435	
100	50	0.8	99.956		0.0443	
100	3	0.20	37.1		29.5	
100	3	0.35	6.29	6.34	10.3	10.2
100	3	0.50	1.30	1.34	2.96	2.95
100	3	0.65	0.360	0.377	1.22	1.21
100	3	0.70	0.246	0.250	1.02	1.05
100	3	0.75	0.170	0.174	0.913	0.940
100	3	0.80	0.119	0.119	0.907	0.902
100	3	0.88	0.0686	0.0702	1.26	1.28
100	3	0.95	0.0424	0.0414	3.71	3.69
10	3	0.8	8.29	8.29	33.2	33.1
30	3	0.8	1.20	1.20	7.61	7.58
100	3	0.8	0.119	0.122	0.907	0.915
300	3	0.8	0.0137	0.0136	0.110	0.109
1000	3	0.8	0.00124	0.0012	0.0102	0.0101
3000	3	0.8	0.000137	0.0001	0.00114	0.0011
10000	3	0.8	0.0000112	<0.0001	0.000103	0.0001

[a] Reference 1.

From eq 10, one obtains the well-known result

$$\bar{n} = 1/T \tag{31}$$

Finally, by eq 9, we find the new results

$$W(0, k) = 1 + G\left[1 + \frac{1}{\rho + T} + \frac{P_{aa}}{1 - P_{aa}} - \frac{1}{T}\right] + \frac{GT}{T + E}\left[-k - \frac{2}{\rho + T} - \frac{P_{aa}}{1 - P_{aa}} + \frac{1 - T + kT + T(\rho + T)^{-1}}{T + E}\right] \tag{32}$$

$$W(i, k) = TX(i, k)\left[-k - \frac{2}{\rho + T} - \frac{P_{aa}}{1 - P_{aa}} + (i + 1)\frac{1 - T + kT + T(\rho + T)^{-1}}{T + E}\right] \qquad i \geq 1 \tag{33}$$

Frensdorff[1] has made some determinations of $W(0, k)$ and $W(1, k)$ by a Monte Carlo generation of random copolymers. He studies the special model for which

$$P_{aa} = P_{oa} = (1 - T)X_a$$
$$P_{ao} = P_{oo} = (1 - T)(1 - X_a) \tag{34}$$

where X_a is the mole fraction of A. In Table I we contrast his results with eq 32 and 33 and find they are remarkably accurate. The reader will want to consult ref 1 for plots of W and X as a function of X_a and $\bar{n}$.

The simple technique for determining a generating function which keeps track of the chain length with z should find application in a number of other copolymer statistics problems.