Excluded Volume of an Intermediate-Molecular-Weight DNA. A Monte Carlo Analysis

CAROL BETH POST,* *Department of Chemistry, B-017, University of California at San Diego, La Jolla, California 92093*

Synopsis

A Monte Carlo procedure was used to determine the effect of excluded volume on the dimensions of an intermediate-molecular-weight DNA for different Na^+ concentrations. The calculation of α, the parameter for the linear expansion due to excluded volume, was accomplished by generating sets of chains and, for each set, comparing the average radius of gyration for the set of chains that do not overlap to that averaged over the entire set of chains. Each chain was defined by cylinders linked with free rotation and with bend angles generated according to a weighted Gaussian distribution. The chain parameters—contour length, cylinder length and diameter—were fixed in order to resemble published light-scattering experiments on Col E_1 DNA. Values for α were less than 1.08 for Na^+ concentrations between 0.007 and $1.0M$. A previously reported analytical calculation of the excluded-volume correction of intermediate-sized DNA gave results that are closely similar to those from the Monte Carlo analysis.

INTRODUCTION

Numerous experimental measurements have been made to determine the persistence length, a, of DNA (many references are given in Ref. 1), and most recently to determine the dependence of a on ionic strength. In principle, the value of a can be obtained from the radius of gyration (as well as from other experimentally observable quantities related to the polymer dimensions); however, the analysis of the data for intermediate- and high-molecular-weight DNA is complicated by the effects of excluded volume; the interactions of nonsequential segments "perturb" the polymer dimensions relative to those predicted for the statistical wormlike chain. Several theories on excluded volume exist,[2] but most of this work has assumed Gaussian statistics and is therefore valid for wormlike chains only in the limit of infinite molecular weight. One theoretical study that is applicable to intermediate-sized wormlike chains was reported by Yamakawa and Stockmayer.[3]

In contrast to estimating excluded-volume effects using a theoretical approach, we present a Monte Carlo procedure to determine the excluded volume of an intermediate-sized DNA.

---

* Present address: Inorganic Chemistry Labs, South Parks Road, Oxford OX1 3QR, England.

Chains of linked cylinders with free rotation and with bend angles restricted to a certain distribution were generated in free space. The cylinder dimensions were chosen such that the Monte Carlo computation modeled the light-scattering experiments of Borochov et al.,[4] which measured the radius of gyration, R, of Col E_1 DNA ($M_r = 4.3 \times 10^6$) as a function of Na^+ concentration. Manning[5] has also evaluated an excluded-volume correction for the data of Borochov et al.[4] using the Yamakawa and Stockmayer[3] factor for a chain with a finite number of statistical segments. In part, the purpose of this report is to examine the agreement between the two approaches.

Kirste[6] has carried out a similar computer computation to estimate the excluded volume of the stiff-chain polymer cellulose trinitrate as a function of molecular weight, for a maximum of 8.8×10^5. In contrast to our chain model, his model consisted of vectors linked with a fixed angle of bending, and a constant persistence length equal to 17 nm.

CHAIN MODEL

A Monte Carlo procedure was used to generate a chain of n linked vectors of fixed length, l, and oriented according to the scheme described below. Each vector defined the axis of a cylinder of diameter d. The values of n, l, and d were chosen to correspond to a certain Na^+ concentration of the light-scattering data, while the contour length, L, was constant and equaled that of Col E_1 DNA[4], 2230 nm.

In order to evaluate an excluded-volume correction, each chain was tested for the occurrence of intersecting cylinders, allowing the dimensional parameters for a set of chains to be averaged over the entire set of chains, corresponding to the statistical- or unperturbed-chain dimensions, or over the subset containing only chains without an intersection, corresponding to real- or perturbed-chain dimensions. Resemblance to the light-scattering experiments was achieved by matching the Monte Carlo root-mean-square $\langle R^2 \rangle_{MC}^{1/2}$ values for the sets of nonintersecting chains to the light-scattering values R_{LS} for Col E_1 DNA. (We hereafter denote the unperturbed dimensions by zero-subscripted symbols, and the real dimensions by unsubscripted symbols.) The excluded-volume correction to R is the linear expansion parameter, α, defined as follows:

$$\alpha^2 = \langle R^2 \rangle_{MC} / \langle R_0^2 \rangle_{MC} \tag{1}$$

Angle Generation and Cylinder Dimensions

The orientation of the axis of the ith cylinder relative to the axis of the $(i - 1)$st cylinder is defined by the rotation angle, ϕ_i, and the bend angle, θ_i, where θ equals $0°$ when there is no bending. All values of ϕ between zero and 2π are equally probable and are obtained using a random-number generator. Values for θ were generated according to the following distribution function:

$$P(\theta_i) = \sin \theta_i \, \exp(-q\theta_i^2) \qquad (2)$$

The width of the distribution depends on the parameter q. This is the random-ϕ model described by Schellman.[7]

The first step to produce the distribution in Eq. (2) was the generation of random numbers normally distributed about a mean of zero and with a standard deviation equal to $(2q)^{-1/2}$ using the ALGOL program[8] NORRF4. To assign a value to q, we refer to the expression for a chain with restricted bending that relates the step length and angle to the chain persistence, p.[9] We remark that p does not equal a since the chain model consists of finite steps and, consequently, large bend angles.

$$l/p = 1 - \langle \cos \theta \rangle \qquad (3)$$

l/p was set equal to either 0.5 or 0.25. Assignment of l/p determines the value $\langle \cos \theta \rangle$ and, by expansion of the cosine, a certain sum of the moments of θ. These moments can be expressed in terms of q using the distribution function in Eq. (2):

$$\langle \theta^r \rangle = \int_0^\infty \theta^r \sin \theta \exp(-q\theta^2) \, d\theta \Big/ \int_0^\infty \sin \theta \exp(-q\theta^2) \, d\theta \qquad (4)$$

The results of the integration for the first three terms in the expansion of the cosine are given:

$$\langle \theta^2 \rangle = \frac{1}{q} \left(\frac{480q^2 - 40q - 1}{480q^2 + 40q + 3} \right) \qquad (5)$$

$$\langle \theta^4 \rangle = \frac{2}{q^2} \left(\frac{480q^2 - 120q + 3}{480q^2 + 40q + 3} \right) \qquad (6)$$

$$\langle \theta^6 \rangle = \frac{6}{q^3} \left(\frac{480q^2 - 200q + 15}{480q^2 + 40q + 3} \right) \qquad (7)$$

The expansion was evaluated for an assumed value of q in Eqs. (5)–(7), varying q until $1 - \langle \cos \theta \rangle$ became equal to the desired value. The standard deviation for NORRF4 appropriate to l/p equal to 0.5 and 0.25 was found to be 0.913 and 0.550, respectively.

NORRF4 produces two independent values, normally distributed, from which θ_i was taken to be

$$\theta_i^2 = \theta_{i,1}^2 + \theta_{i,2}^2 \qquad (8)$$

The result was a displacement in two orthogonal planes according to the distribution $P(\theta_i) = \theta_i \exp(-q\theta_i^2)$. The second step to get Eq. (2) was to modify this distribution by the factor $(\sin \theta_i)/\theta_i$. Hence, θ_i was accepted with a probability of $(\sin \theta_i)/\theta_i$ by generating a random number, x_i, between zero and one, and comparing x_i to $(\sin \theta_i)/\theta_i$. If x_i was less than $(\sin \theta_i)/\theta_i$, then θ_i was accepted. To be certain that the angular distribution produced by this procedure was the proper one, a set of 7×10^3 θ_i values was generated and was found to compare well with the analytical curve from Eq. (2). In addition, $\langle \theta_i \rangle$ and $\langle \theta_i^2 \rangle$ obtained from the generated distribution were in good agreement with the same quantities calculated using Eq. (2).

The values ϕ_i and θ_i determined the direction of the x-axis of the coordinate system of the ith cylinder. These local coordinates were transformed to the $(i - 1)$st coordinate system by the rotation matrix in Eq. (1) of Ref. 10 and, by successive application of the matrix, to the coordinate system of the first cylinder. This rotation matrix multiplied a vector of length l, and the resulting vector was then added to the $(i - 1)$st vector.

For a given Na^+ concentration, the cylinder diameter was assigned the value for the effective diameter of DNA that Stigter[11] determined by evaluating the repulsion between charged cylinders as a function of the solvent ionic strength. The Monte Carlo analysis did depend on the value assumed for d; however, the change in the excluded-volume correction over a reasonable range for d was small. The effect of d on the evaluation of α is also discussed in the Comments.

The cylinder length was chosen such that the value $\langle R^2 \rangle_{MC}^{1/2}$ corresponded to R_{LS}.[4] An initial estimate for l was determined using the equation[12] for R_{rb}, the unperturbed mean-square radius of gyration of a chain with restricted bending, where R_{rb} is set equal to R_{LS}:

$$\frac{R_{rb}^2}{l^2} = \frac{n(n + 2)\langle 1 + \cos\theta \rangle}{6(n + 1)\langle 1 - \cos\theta \rangle} - \frac{\langle \cos\theta \rangle}{\langle 1 + \cos\theta \rangle^4}$$
$$\times \left[\frac{n(n + 1) - 2n(n + 2)\cos\theta + (n + 1)(n + 2)\langle \cos\theta \rangle^2}{(n + 1)^2} \right] \quad (9)$$

Monte Carlo chains were generated using this initial l. From the approximate α determined by $\langle R^2 \rangle_{MC}$ and $\langle R_0^2 \rangle_{MC}$, a new R_{rb} was estimated. Again solving Eq. (9) for l and repeating the Monte Carlo procedure, a new value of $\langle R^2 \rangle_{MC}^{1/2}$ closer to R_{LS} was obtained. The final l gave $\langle R^2 \rangle_{MC}^{1/2}$ values within less than 2% of R_{LS}.

Testing for Chain Overlap

Upon generation of each vector, the new cylinder was tested for overlap with previously added cylinders by first calculating the center-to-center distance of the cylinder pair. If this distance was greater than l, then overlap was not possible and a new vector was generated; otherwise, the perpendicular distance between the projections of the two cylinder axes, i.e., the position of closest approach of the projections, was determined. A necessary, but not sufficient, condition for intersection is that the perpendicular distance be less than or equal to the cylinder diameter. For a perpendicular distance less than or equal to d, the two cylinders may intersect if the points on the projections of the axes at the position of closest approach of the projections are either within a cylinder or close to a cylinder. The possible configurations for a perpendicular distance less than or equal to d are illustrated in Fig. 1, with the perpendicular distance normal to the page, as indicated by ($\bullet$). If the points of closest approach did not lie within the cylinders, the perpendicular distance between a certain cylinder
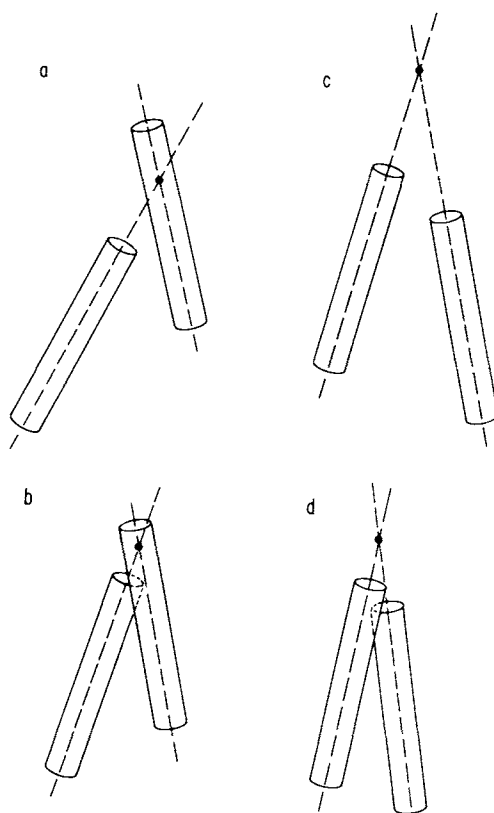
Fig. 1. Possible positions of a cylinder pair relative to the distance of closest approach of the projections of the axes (●), except for the case for which each end of this minimum-length line lies on one of the cylinders. If the closest approach distance is less than the cylinder diameter, then intersection may occur. (a) One end of the minimum-distance line lies on a cylinder, but the cylinders do not intersect. (b) Same as (a) but with an intersection. (c) Neither end of the minimum-distance line lies on a cylinder and no intersection occurs. (d) Same as (c), but with an intersection.

end and the other cylinder was calculated in order to determine if the pair intersected.

The algorithm for detecting an intersection was tested using cylinders with specified coordinates that correspond to the configurations in Fig. 1. One cylinder was systematically revolved around a second cylinder, with different chosen distances along the two axes from the point of intersection.

RESULTS

Chain Statistics

For each Na^+ concentration, the mean persistence and the root-mean-square R and end-to-end length, h, were calculated for the entire set of

chains, containing 800–1100 chains, and for the subset of nonintersecting chains, containing 383–659 chains ($l/p = 0.5$). Monte Carlo values for the unperturbed dimensions from the entire set, $\langle p_0 \rangle_{\text{MC}}$, $\langle R_0^2 \rangle_{\text{MC}}$ and $\langle h_0^2 \rangle_{\text{MC}}$, agreed within the statistical error limits with the values calculated for a chain with restricted bending using Eqs. (3), (9), and (10), respectively. Equation (10) for the square end-to-end length of the statistical chain, h_{rb}^2, is[2,9]

$$\frac{h_{\text{rb}}^2}{l^2} = \frac{n - \langle \cos \theta \rangle (2 + n \langle \cos \theta \rangle)}{\langle 1 - \cos \theta \rangle^2} \tag{10}$$

The Monte Carlo values for R and h, along with d and α from Eq. (1), are shown in Table I for four Na^+ concentrations, ranging from 0.007 to $1.0M$. The excluded-volume correction for the intermediate-sized DNA is less than 8%.

The Effect of l/p on the Monte Carlo Value for α

Being a rigid molecule, DNA is best modeled by a chain obeying wormlike statistics; thus, a rigorous simulation would employ a latitudinal distribution such that $\langle \cos \theta \rangle$ would be close to unity. In the majority of this work, however, $1 - \langle \cos \theta \rangle$ equals 0.5, resulting in a rather large degree of bending between cylinders. This condition was chosen to reduce the computational costs by minimizing the number of cylinders per chain and, therefore, the number of comparisons in testing for intersections (a larger $\langle \theta_i \rangle$ requires a larger l and a smaller n to generate the same $\langle R^2 \rangle_{\text{MC}}$). Since this choice for $\langle \cos \theta \rangle$ is not the wormlike limit, it is necessary to establish that the value for l/p used in the Monte Carlo computation does not change α.

The results using two values of l/p, 0.25 and 0.5, to find α corresponding to the light-scattering conditions for 0.05 and $0.2M$ Na^+ are in Table I. For both values of l/p, d was set equal to either 7.4 or 4.4 nm, and l was varied until $\langle R^2 \rangle_{\text{MC}}^{1/2}$ was approximately equal to R_{LS}. In order to obtain agreement between $\langle R^2 \rangle_{\text{MC}}^{1/2}$ and R_{LS} for the two l/p values, it was necessary to change the chain persistence. As shown in Table I for 0.05 and $0.2M$ Na^+, the cylinder length for $l/p = 0.5$ was not twice the cylinder length for $l/p = 0.25$.

Both values of α obtained for different l/p agree within experimental error; therefore, the computation of α is unaltered by the value chosen for the ratio; the excluded-volume correction is the same for the cases of either two or four cylinders corresponding to the average chain persistence. Setting $\langle 1 - \cos \theta \rangle = 0.5$ appears to be a valid approximation in the calculation of α for intermediate-sized DNA, since reducing the average bending angle and increasing the number of cylinders had no effect on the small excluded-volume correction described here.

TABLE I

Monte Carlo RMS Radius of Gyration and End-to-End Length Averaged Over the Set of Chains Without Overlap (R,h) and Over the Complete Set of Chains (R_0, h_0)

[Na⁺] (M)	l/p^a	d^b (nm)	l^c (nm)	n^c	No-Overlap Set Complete Set (No. of chains)	$\langle R^2 \rangle^{1/2}_{MC}$ $\langle R_0^2 \rangle^{1/2}_{MC}$ (nm)	$\langle h^2 \rangle^{1/2}_{MC}$ $\langle h_0^2 \rangle^{1/2}_{MC}$ (nm)	α
0.007	0.5	18.9	50.7	44	506 900	243.2 ± 2.3 227.0 ± 1.8	604.8 ± 11.5 559.4 ± 8.3	1.072 ± 0.014
0.05	0.5	7.4	37.2	60	532 900	210.1 ± 2.2 199.1 ± 1.7	532.4 ± 10.1 500.3 ± 7.4	1.055 ± 0.013
	0.25	7.4	15.9	140	563 900	207.7 ± 1.9 196.5 ± 1.6	522.9 ± 7.4 488.5 ± 7.1	1.057 ± 0.013
0.2	0.5	4.40	29.7	75	659 1100	186.5 ± 1.9 179.4 ± 1.6	458.5 ± 7.8 440.5 ± 5.9	1.040 ± 0.013
	0.25	4.40	13.0	172	193 300	186.0 ± 3.1 178.3 ± 2.5	460.3 ± 13.1 441.1 ± 10.3	1.043 ± 0.022
1.0	0.5	2.95	21.1	106	383 800	157.2 ± 1.9 150.5 ± 1.3	388.6 ± 8.8 371.8 ± 5.9	1.045 ± 0.015

[a] $l/p = 1 - \langle \cos \theta \rangle$, where θ is the angle of bending.
[b] Effective diameter of DNA reported by Stigter (Ref. 11).
[c] $L = nl = 2230$ nm.

TABLE II

Comparison of Persistence Lengths, a, Corrected for Excluded Volume by Different Methods

[Na$^+$] (M)	R_{LS}[a] (nm)	d[b] (nm)	α	a (nm)
0.007	244.8	18.9	1.072	77.8[c]
		17.2	1.06	82.0[d]
		18.9		72.5[e]
				68.4[f]
0.05	210.2	7.4	1.055	57.5[c]
		6.1	1.04	60.0[d]
		7.4		55.8[e]
0.20	186.0	4.4	1.040	45.7[c]
		3.5	1.04	46.5[d]
		4.4		44.1[e]
				40.8[f]
1.0	159.5	2.95	1.045	32.7[c]
		2.33	1.04	33.0[d]
		2.95		32.3[e]
				30.3[f]

[a] Radius of gyration from light-scattering data of Borochov et al. (Ref. 4).

[b] Effective cylinder diameter.

[c] Determined from Eq. (11), setting L = 2230 nm and $R_w = R_{LS}/\alpha$, where α was obtained by the Monte Carlo procedure.

[d] As reported by Manning (Ref. 5).

[e] Determined by Stigter's uniform ellipsoidal model, Eq. (14) of Ref. 13.

[f] As reported by Kam et al. (Ref. 14).

COMMENTS

This study provides an excluded-volume correction for the value of R obtained from the light-scattering data of Borochov et al.,[4] in that $\langle R^2 \rangle_{MC}^{1/2}$ for nonintersecting chains is matched to the observed R_{LS} by the proper choice for the cylinder length. Although our model was allowed a degree of bending higher than that of the usual wormlike model for DNA, the chain of cylinders is an appropriate model to evaluate an excluded-volume correction, since the resulting α was not sensitive to the choice of $1 - \langle \cos \theta \rangle$.

The Monte Carlo values for α determined from Eq. (1) were used to calculate the persistence length, a, of DNA as a function of Na$^+$ concentration. The equation[12] for the squared radius of gyration of a wormlike chain, R_w^2, is

$$R_w^2 = \frac{La}{3} - a^2 + \frac{2a^3}{L} - \frac{2a^4}{L^2}(1 - e^{-L/a}) \qquad (11)$$

Equation (11) was solved for a with L equal to 2230 nm and with R_w equated to the light-scattering value R_{LS} divided by α. Examination of Table II finds that the persistence length of the wormlike chain, a, obtained in this way does not equal the persistence of the chain with restricted
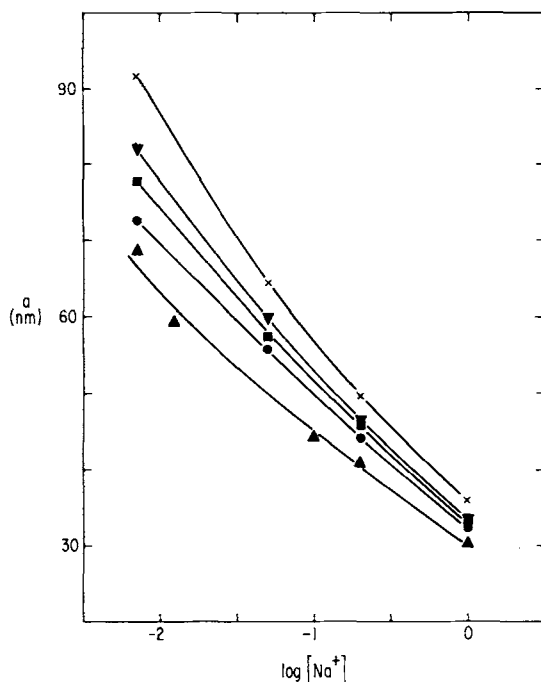
Fig. 2. Persistence length, a (also given in Table II), as a function of the logarithm of the Na$^+$ concentration. The value for a is determined from Eq. (11) for R_{LS} ($\times$) uncorrected, ($\blacktriangledown$) corrected according to Manning (Ref. 5), ($\blacksquare$) corrected by the Monte Carlo analysis, ($\blacktriangle$) corrected according to Kam et al. (Ref. 4), and ($\bullet$) determined from Eq. (14) of Ref. 13.

bending, p (p can be determined from l and l/p in Table I). An approximate relationship between a and p can be found by equating R_w^2 and R_{rb}^2. Considering only the first terms in Eqs. (9) and (11) for the case of large L, and using Eq. (3), we find a to be less than p by the factor $\langle 1 + \cos\theta \rangle/2$; a and p are equivalent only in the limit $\langle \cos\theta \rangle \rightarrow 1.0$. As expected, calculated values of a (Table II) and of p (Table I) differ by approximately $\langle 1 + \cos\theta \rangle/2$.

The linear expansion of DNA as a result of excluded volume ranged approximately from 5% at $1.0M$ Na$^+$ to 8% at $0.007M$ Na$^+$, as shown in Table I. The Monte Carlo estimate of α does change slightly with the choice of d. An 18% increase in d (from 18.9 to 22.3 nm for $0.007M$ Na$^+$) resulted in about a 22% increase in the difference between α and unity (from 0.072 to 0.088) and, consequently, a 4% decrease in a (from 77.8 to 75.3 nm).

The effect of the assumed value for d could not be evaluated from the generation of one set of chains and varying d in the test for overlapping cylinders; a larger d preferentially eliminates chains with smaller R from the subset of nonintersecting chains and must be compensated by a decrease in the cylinder length in order to maintain the condition that $\langle R^2 \rangle_{MC}^{1/2}$ is approximately equal to R_{LS}.

The Monte Carlo-determined values of α and a are compared in Table II to values determined by various analytical procedures. Manning's[5] values for α and a, obtained using the factor derived by Yamakawa and Stockmayer[3] to correct the excluded-volume parameter for chains smaller than the coil limit, agree well with the Monte Carlo results. Had the Monte Carlo and Manning's analyses assumed the same effective diameter, the agreement would be even closer, since a decrease in d leads to an increase in a. Also included in the table are values for a that we have calculated from an expression obtained by Stigter[13] assuming a uniform ellipsoidal segment distribution. Equation (14) of his work was solved for a, assuming a radius of gyration equal to R_{LS} and d equal to the values shown in Table II. These values for a are comparable to the Monte Carlo results. Although the agreement with these two analytical results is encouraging, a comparison under these conditions may not be a good test of the theories, since α is very near unity. The excluded-volume corrections reported by Kam et al.,[14] which were obtained from their data in a recent analysis based on Daniel's distribution function, are somewhat larger than the other values for α in Table II. Correspondingly, their values for a are smaller than those obtained by the other procedures, as shown in Fig. 2.

By comparison with the Monte Carlo results, it is concluded that Manning's procedure[5] provides accurate values for α and a. In addition, values for a determined from an equation obtained by Stigte[13] using a different theoretical approach gave surprisingly similar results.

References

1. Hagerman, P. J. (1981) *Biopolymers* **20,** 1503–1535.

2. Yamakawa, H. (1971) *Modern Theory of Polymer Solutions*, Harper & Row, New York.

3. Yamakawa, H. & Stockmayer, W. H. (1972) *J. Chem. Phys.* **57,** 2843–2854.

4. Borochov, N., Eisenberg, H. & Kam, Z. (1981) *Biopolymers* **20,** 231–235.

5. Manning, G. S. (1981) *Biopolymers* **20,** 1751–1755.

6. Kirste, R. G. (1970) *Disc. Faraday Soc.* **49,** 51–59.

7. Schellman, J. A. (1974) *Biopolymers* **13,** 217–226.

8. Knuth, D. (1968) *Art of Computer Programming*, Vol. 2, Addison-Wesley, Reading, Mass., p. 104.

9. Bloomfield, V. A., Crothers, D. & Tinoco, I. (1974) *Physical Chemistry of Nucleic Acids*, Harper & Row, New York.

10. Eyring, H. E. (1932) *Phys. Rev.* **39,** 745–748.

11. Stigter, D. (1977) *Biopolymers* **16,** 1435–1448.

12. Benoit, H. & Doty, P. (1953) *J. Phys. Chem.* **57,** 958–963.

13. Stigter, D. (1982) *Macromolecules* **15,** 635–641.

14. Kam, Z., Borochov, N. & Eisenberg, H. (1981) *Biopolymers* **20,** 2671–2690.