BIOPOLYMERS VOL. 13, 217–226 (1974)

Flexibility of DNA

JOHN A. SCHELLMAN, *Department of Chemistry, University of Oregon, Eugene, Oregon 97403*

Synopsis

A model for the flexibility of DNA is proposed that is based on discrete variations in the direction of propagation in going from one subunit to the next. Expansion of the local free energy in terms of the local bending gives a Gaussian distribution function. The assumption of the independence of local bends on neighbors lead to very simple formulae for the persistence length and the characteristic ratio. Emphasis, however, is placed on the application of the formulae for molecules of finite size where the persistence length and C_∞ are not defined. The formulae are worked out for two models, which should serve as limits for the real physical situation.

The manner in which one deals with the hydrodynamical properties of locally stiff, but very long, molecules such as double-stranded DNA depends markedly on chain length. For short lengths (which nevertheless have large axial ratios) the molecule is representable as a rigid cylinder or ellipsoid. Distances between points on the molecule are calculable from the rise per residue and the number of intervening residues. When the molecule is very long, as it is when the molecular weight is 10^7 or greater, the Kuhn limit of a Gaussian distribution is eventually reached, and distances between sequentially widely separated groups are obtainable from random chain formulas.

On the other hand, when the length of the molecule is of the order of one or two persistence lengths, the method of interpretation is far from established. Since this is the molecular weight range that results from the sonication of DNA, the problem has considerable relevance to experimental work. We have encountered this difficulty in interpreting the streaming linear dichroism of DNA systems in the length region 1000–2000 Å. The simple models described below were developed to aid in the interpretation and visualization of such systems. These models deal with discrete units of the DNA helix rather than the continuum models, which are prevalent in the field. The purpose is to develop the hydrodynamic properties in terms of the same kind of structural parameters as are employed in the analysis of the conformation and spectra of DNA.

TWO MODELS FOR THE BENDING OF DNA

The semiflexible chain is assumed to consist of a series of discrete links, which we identify with the base pairs and their ribose phosphate moieties.

© 1974 by John Wiley & Sons, Inc.

The propagation associated with the ith link will be designated by a vector $\mathbf{l}_i$. In the undistorted helix, all of the $\mathbf{l}_i$ are colinear and coincide with the helix axis. Because of thermal fluctuations, there will be local distortions from regularity, which cause a local bending of the molecule with the result that the ith link is no longer in the same direction as the $i - 1$ link but deviates by an angle θ_i.

The question arises as to whether the local direction of propagation has any meaning in a structure that is subject to arbitrary distortions at every linkage. There are a number of ways in which the local sense of propagation of the helix can be defined. The most convenient method will ultimately depend on the mechanical details of the bending mechanism, which is at present unknown. If it is the B helix of DNA, a vector perpendicular to the base planes provides such a definition. If the base planes are twisted, the least square plane can serve instead. More generally, if the bond angles, bond lengths, and dihedral angles of a complete unit of a chain molecule are fixed, a helix is determined with a given direction and rise per residue.[1] For distorted systems, these parameters may be used to determine the local vector $\mathbf{l}_i$. Other ways involving a local principal axis of inertia, etc., are also possible. The main point is that the concept can be made physically meaningful and the appropriate method can be eventually decided by convenience.

We offer two models for the probability of local bending of a DNA molecule. In the first (the hinge model) we assume that between each consecutive pair of subunits there is a preferred axis of bending that is perpendicular to the helix axis. Possible modes of change in direction of the local axis are folding into the large and small grooves, sliding of base pairs on one another, or changes in conformation of the ribose phosphate chain, or combinations of these motions. For small bending about the axis, the local change in free energy caused by the distortion can be expanded in a Taylor series in θ_i

$$g_i = g_{0i} + \frac{dg_i}{d\theta_i}\theta_i + \frac{1}{2}\frac{d^2g_i}{d\theta_i^2}\theta_i^2 \ldots \tag{1}$$

$$\Delta g_i = \frac{1}{2} g_i''\theta_i^2. \tag{2}$$

The disappearance of the first derivative in the second form of Eq. (1) arises from the condition of equilibrium of the helix. Powers of θ_i^2 higher than the second have been dropped. One consequence of this is the automatic assumption that the energy versus θ curve is symmetrical. While this is not a necessary property of the model, the arbitrary addition of skewness to the local distribution would involve parameters for which there is no hope of experimental determination at the present time. Another assumption in Eq. (1) is that the links can be treated as independent of one another. The generalization of the model to include interaction amongst the links will be brought out in the final discussion.

Utilizing Eq. (2) in the Boltzmann factor for the ith link we find

$$P(\theta_i) = N^{-1/2} \exp\left[-\frac{g_i''\theta_i^2}{2RT}\right] = \left(\frac{a_i}{2\pi}\right)^{1/2} \exp\left[-\frac{a_i\theta_i^2}{2}\right] \tag{3}$$

where $P(\theta_i)$ is the probability of angle θ_i, N is a normalizing coefficient, and $a_i = g_i''/RT = 1/\Delta_i^2$ where Δ_i is the standard deviation of θ_i. From this Gaussian distribution we can calculate average values

$$\langle\theta_i^{2n}\rangle = \frac{(2n)!}{2^n n!}\Delta_i^{2n}$$

$$\langle|\theta_i^{2n+1}|\rangle = \sqrt{\frac{2}{\pi}}\, 2^n n! \Delta_i^{(2n+1)} \tag{4}$$

and

$$\langle\cos\theta_i\rangle = \exp[-{}^1/_2 a_i] = \exp[-\Delta_i^2/2]. \tag{5}$$

The odd moments of θ_i vanish. $\langle\sin\theta_i\rangle$ involves confluent hypergeometric functions but will not be necessary in our analysis though it does appear in matrix methods of chain generation. In particular

$$\langle|\theta_i|\rangle = \sqrt{\frac{2}{\pi}}\,\Delta_i \text{ and } \langle\theta_i^2\rangle = \Delta_i^2. \tag{6}$$

In calculating these averages the range of integration has been extended to infinity, thereby assuming that large angles are sufficiently improbable that they contribute little to the integration. Apart from this mathematical approximation, there is the physical aspect of the problem, which would dictate solvent penetration and an entirely new mechanism for bending through wide angles. This subject will be taken up again at the end of the next section.

In the second model (the random ϕ model), it is assumed that there is no preferred local axis of bending so that for a given θ_i all the values of the longitudinal angles ϕ_i are equally likely. Assuming the same kind of energy dependence on θ_i as before, the probability distribution function takes the form $\exp[-(a_i\theta_i^2)/2]\sin\theta_i$ where $\sin\theta_i$ is the latitudinal weighting function. One can now calculate the moments of θ_i, $\langle|\theta_i|^n\rangle$, $\langle\cos{}^n\theta_i\rangle$, $\langle\sin{}^n\theta_i\rangle$, etc. in terms of confluent hypergeometric functions or incomplete gamma functions. Though these functions have been tabulated, they are not particularly convenient to use. Since the present analysis will only require $\langle\cos\theta_i\rangle$, a formalism will be used that provides this average value in a simple way. Instead of expanding the local free energy g_i in terms of θ_i, we expand it in terms of $\sin(\theta_i/2)$. Temporarily suppressing the subscript i

$$\Delta g = \left(\frac{dg}{d\sin\theta/2}\right)_{\sin\theta/2=0} \sin(\theta/2) + \frac{1}{2}\left(\frac{d^2g}{d(\sin\theta/2)^2}\right)_{\sin\theta/2=0} \sin^2\theta/2 + \ldots \tag{7}$$

Now

$$\left(\frac{dg}{d \sin \theta/2}\right)_{\sin \theta/2=0} = 2\left(\frac{dg}{d\theta}\right)_{\theta=0} = 0$$

and

$$\left(\frac{d^2g}{d(\sin \theta/2)^2}\right)_{\sin \theta/2=0} = 4\left(\frac{d^2g}{d\theta^2}\right)_{\theta=0} = 4g'',$$

so

$$\Delta g = 2g'' \sin^2 \theta/2 \mp g''(1 - \cos \theta). \tag{8}$$

Calculation shows that the factor $(1 - \cos \theta)$ in Eq. (8) differs from $\theta^2/2$ in Eq. (2) by one part in a thousand at $\theta = 5°$ and three parts in a thousand at $\theta = 10°$. There is a no *a priori* reason for supposing Eq. (2) to be superior to Eq. (8). Using Eq. (8)

$$\langle \cos \theta \rangle = \frac{\int_0^\pi \cos \theta \, e^{a \cos \theta} \sin \theta d\theta}{\int_0^\pi e^{a \cos \theta} \sin \theta d\theta} \tag{9}$$

$$= \coth a - 1/a \equiv L(a)$$

where $L(a)$ is the Langevin function and $a = g''/RT$ as before.

CHAIN STATISTICS

Three experimental quantities that are used to characterize the mode of propagation of flexible chains are the persistence length P_∞, the rms end-to-end distance $h = \langle \mathbf{h}^2 \rangle^{1/2}$, and the radius of gyration R_G. The persistence length is defined as the average sum of the projections of all bonds (including the first) on the first bond of the chain[2]

$$P_n = \mathbf{h} \cdot \frac{\mathbf{l}_1}{l} = l \sum_{k-1}^{n} \cos \theta_{1k} \tag{10}$$

where $\mathbf{h}$ is the end-to-end vector given by $\mathbf{h} = \sum_{l=1}^{n} \mathbf{l}_i$, l is the length of each link, and θ_{1k} is the angle between the kth link and the first link. For Eq. (10) to provide an adequate definition of the persistence length, n must be sufficiently large that there is no correlation between the first and final links so that the series converges before the end of the molecule is reached. Since this condition will frequently not be met in DNA systems of interest, we define the "persistence" of a chain of n links by Eq. (10). The persistence length is then P_∞.

In both of the models discussed in the previous section, the projection of bond $i + 1$ on bond i is in the direction of bond i. That is, all components perpendicular to the previous bond average to zero. For the hinge model this is true because positive and negative values of θ_{i+1} are equally likely; it is true for the random ϕ model because the average value of $\cos \phi$ and $\sin \phi$

are zero. As a result, matrix methods need not be employed and the average value of $\cos\theta_{1k}$ is given by $\langle\cos\theta_{1k}\rangle = \prod_{i=2}^{k}\langle\cos\theta_{i-1,\,i}\rangle$. We have already assumed that the θ_i are independent of one another. We now assume as well that all the g_i'' are the same. If this is not true because of base pair sequence, then g'' must be understood as an average value for the chain. With these assumptions, all the averages in the product are identical so that $\langle\cos\theta_{1k}\rangle = \langle\cos\theta\rangle^{k-1}$. This gives for the persistence P_n

$$P_n = l\left(1 + \sum_{k=2}^{n} \alpha^{k-1}\right) = l\left(\frac{1-\alpha^n}{1-\alpha}\right) \tag{11}$$

where $\alpha = \langle\cos\theta\rangle$. For a long enough chain $\alpha^n \to 0$ and the persistence length is given by

$$P_\infty = l/(1-\alpha). \tag{12}$$

The value of n required for this convergence increases with the stiffness of the chain.

The mean square chain length $\langle h^2\rangle$ is given by

$$\begin{aligned}\langle h^2\rangle &= \langle \mathbf{h}\cdot\mathbf{h}\rangle = \langle(\Sigma\mathbf{l}_i\cdot\Sigma\mathbf{l}_j)\rangle \\ &= nl^2\left(\frac{1+\alpha}{1-\alpha} - \frac{2\alpha}{n}\frac{(1-\alpha^n)}{(1-\alpha)^2}\right) \equiv nl^2C_n\end{aligned} \tag{13}$$

where C_n is the characteristic ratio for n links (Flory, 1969). For n large this becomes

$$\langle h^2\rangle = nl^2\frac{1+\alpha}{1-\alpha} = nl^2C_\infty. \tag{14}$$

Finally for the radius of gyration

$$\begin{aligned}R_G^2 = \frac{nl^2}{6}\Bigg[&\left(\frac{n+2}{n+1}\right)\left(\frac{1+\alpha}{1-\alpha}\right) - \frac{\alpha}{(n+1)(1-\alpha)^2} \\ &+ \frac{2\alpha^2}{(n+1)^2(1-\alpha)^3} - \frac{2\alpha^3(1-\alpha^n)}{n(n+1)^2(1-\alpha)^4}\Bigg]\end{aligned} \tag{15}$$

and for n large

$$R_G^2 = \frac{nl^2}{6}\frac{1+\alpha}{1-\alpha} = \frac{\langle h^2\rangle}{6}. \tag{16}$$

Equations (11)–(16) are now standard in the field. (See the monographs of Volkenstein[3] or Flory[2] for original references.) α is normally a fixed bond angle rather than a small angle produced by structural fluctuations.

As an initial attempt at applying the results of the previous section, we will assume that the only operative mechanism for DNA flexibility is small local bending between adjacent links. (This hypothesis probably does not apply to long DNA molecules as will be discussed shortly.) With this

assumption $\langle \cos \theta \rangle$ can be evaluated via Eqs. (12), (14), and (16) from the experimental determination of P_∞ or C_∞. We have

$$\frac{P_\infty}{l} = \frac{1}{1-\alpha} = \frac{1}{1-e^{-\Delta^2/2}} \qquad \text{hinge model} \tag{17a}$$

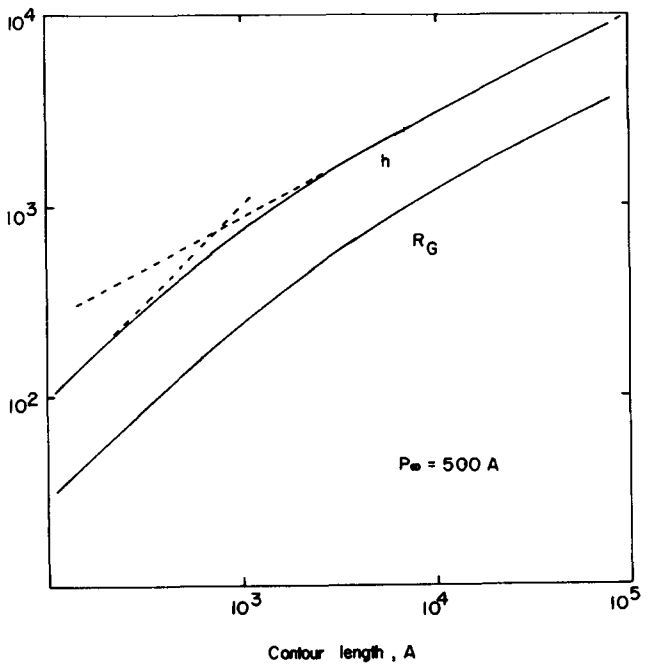
$$= \frac{1}{1-L(a)} \qquad \text{random } \phi \tag{17b}$$

or

$$C_\infty = \frac{1+\alpha}{1-\alpha} = \coth(\Delta^2/4) \qquad \text{hinge model}$$

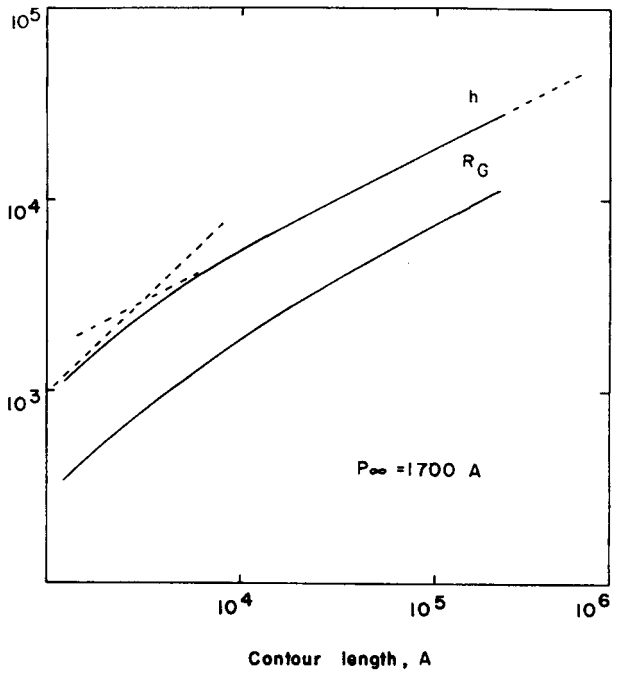
$$= \frac{1+L(a)}{1-L(a)} \qquad \text{random } \phi. \tag{18}$$

Recent estimates of P_∞ for DNA range from 500 Å[4] to 1700 Å.[5] For persistence lengths as long as 500 Å, $P_\infty/3.45$ is in the neighborhood of 150; $\langle \cos \theta \rangle$ is sufficiently close to unity that the linear expansion of the exponential in Eq. (17a) produces an error of less than 1%. The Langevin function in Eq. (17b) can be replaced by $(1 - 1/a)$ to very high accuracy. Thus



(*a*)

Fig. 1 (*continued*)



(b)

Fig. 1. Log–log plot of rms end-to-end distance h and radius of gyration R_G as function of contour length for DNA's with persistence lengths of 500 and 1700 Å. Dashed line starting on upper right of curve for h has slope of $^1/_2$ and represents Kuhn limit of Gaussian behavior. Dashed line on lower left has slope of unity and represents limit of completely rigid rod. Only mechanism considered in constructing figures is small local bending. Length of link taken to be 3.45 Å.

$$\frac{P_\infty}{l} = \frac{1}{1 - (1 - \Delta^2/2)} = \frac{2}{\Delta^2} = \frac{2g''}{RT} \qquad \text{hinge model}$$

$$\frac{P_\infty}{l} = \frac{1}{1 - (1 - 1/a)} = a = \frac{g''}{RT} \qquad \text{random } \phi. \tag{19}$$

For the hinge model g'' is calculated to be 43.5 and 148 kcal/rad^2 at 300°K for a persistence length of 500 and 1700 Å, respectively. For the random ϕ model g'' is doubled, i.e., 87 and 296 kcal/rad^2. The factor of 2 represents the fact that with identical g'' the random ϕ model will have larger values of $\langle \cos \theta \rangle$ because of the extra degree of freedom. $\theta = 0$ is the most probable value for the hinge model; it is not for the random ϕ model.

Corresponding to the cases of Eq. (19), we find that

$$C_\infty = \coth \Delta^2/4 \cong 4/\Delta^2 = \frac{4g''}{RT} \tag{20a}$$

$$C_\infty = \frac{1 + (1 - 1/a)}{1 - (1 - 1/a)} \cong 2a = \frac{2g''}{RT} \tag{20b}$$

If g'' is purely energetic, P_∞ and C_∞ should diminish with increasing T. In the unlikely event that it is purely entropic, there would be no temperature coefficient. On the basis of temperature studies, Gray and Hearst[6] have concluded that the enthalpy of bending is positive and accompanied by a small negative entropy.

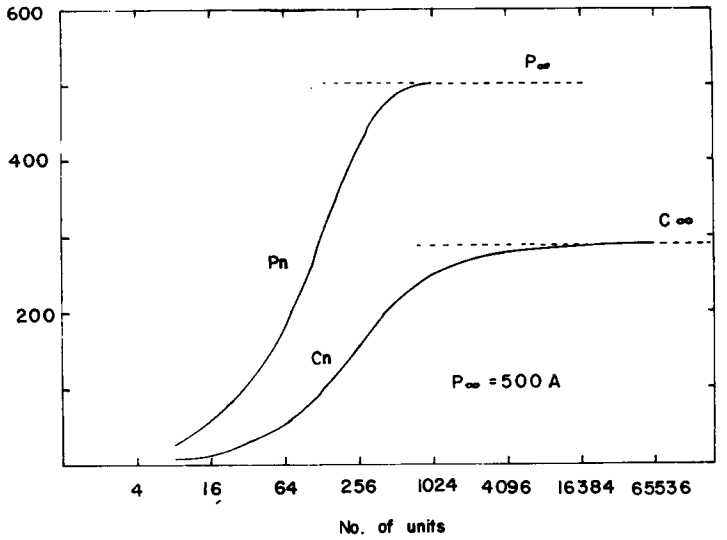
There may be restrictions on the range of applicability of the above formulae. Theoretical and experimental results are in opposition to the uninterrupted propagation of DNA helices as flexible rods. On the experimental side the exchange of interior hydrogen atoms for deuterium[7] or the reaction of the interior hydrogen of adenine[8] with formaldehyde indicates a transient rupture of the ideal order of the Watson–Crick helix. These internal relaxations have been called breathing modes by von Hippel and coworkers. Their presence is in complete consistency with the prediction of Landau and Lifschitz[9] that ordered phases of linear structures should be finite in length.

The extent to which these breathing modes provide an additional mechanism for DNA flexibility is unknown at present, though their contribution should be negligible for low molecular weight. Quantitative estimates of the frequency of breathing modes along the chain and of their inherent flexibility will be required to settle this problem. Since measurements of the persistence length of DNA still vary, we have calculated $\langle h^2 \rangle^{1/2}$ and R_G as a function of contour length for $P_\infty = 500$ and 1700 Å, which appear to be the lower and upper limits of contemporary investigations. The results are shown in Figure 1. Using the values of $\langle \cos \theta \rangle$ appropriate for these persistence lengths, P_n and C_n have been plotted as a function of n in Figure 2. Extensive experimental results with relatively short DNA molecules together with comparison with curves of this kind should eventually provide an estimate of the flexibility of DNA in the absence of highly flexible loops. The values of $\langle \cos \theta \rangle$ and $\langle \theta \rangle$ associated with these curves are given in Table I.

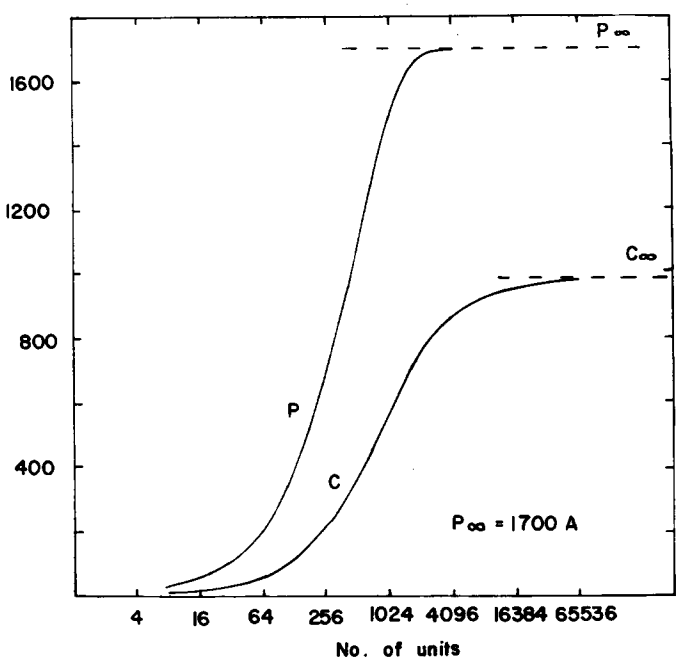
DISCUSSION

It is rather likely that a realistic model for DNA bending should include a dependence of the restoring force g'' on ϕ as well as on θ. If g'' is twofold symmetric in ϕ so that the restoring force for $\phi + \pi$ equals the restoring force for ϕ, transverse contributions of propagation vectors will average to zero. In this case, the value for P_∞/l and C_∞ should lie within the twofold variation in models depicted by Eqs. (19) and (20), since these span situations from total restriction of ϕ to a uniform distribution of ϕ. Consequently, $RT(P_\infty/l)$ is a measure of an effective restoring force averaged over ϕ. For the random ϕ model or hinge model, this is equal to the value of g'' or twice its value, respectively.

Of models previously proposed in the literature, the present work bears the strongest resemblance to that of Landau and Lifschitz[9] who considered a continuous thin rod subject to curvature as a result of thermal fluctuations. The local energy of bending was considered to be proportional to the square of the local curvature vector, i.e., proportional to the square of



(a)



(b)

Fig. 2. Plot of C_n and P_n as function of number of units in chain for same conditions as Fig. 1. P_n is in angstroms and C_n is dimensionless. Dashed lines indicate asymptotic limits for infinite chains. Note that C_n is changing significantly for quite long chains.

TABLE I
Values of Parameters for the Two Models and Two Values for P_∞

	P_∞ = 500 Å				P_∞ = 1700 Å			
	$\langle\cos\theta\rangle$	$\langle\lvert\theta\rvert\rangle$ (deg)	C_∞	g''/RT	$\langle\cos\theta\rangle$	$\langle\lvert\theta\rvert\rangle$ (deg)	C_∞	g''/RT
Hinge Model	0.993	5.4	289	72.5	0.998	2.9	985	246
Random ϕ	0.993	6.0	289	145	0.998	3.2	985	493

the reciprocal of the local radius of curvature. The random ϕ model discussed above is essentially a discrete version of the Landau–Lifschitz model. The introduction of realistic molecular features does not add any mathematical complexity to the treatment.

It has not seemed desirable to propose models that go too far beyond the refinement of experimental results. When and if it becomes necessary, a number of refinements can be considered. If the angles are not independent but there are cross terms in the energy involving products such as $k\theta_i\theta_{i+1}$, the resulting quadratic form can be diagonalized using standard formulae. The shape of the molecule will then be given as a linear combination of normal modes of distortion. Also, there is no reason why the equilibrium value of θ need be 0. For other types of chain, the local free energy is given by $(k/2)(\theta - \theta_0)^2$ in the parabolic approximation. If θ_0 is not zero, the results come out in terms of hypergeometric functions but this is no serious handicap since a table appropriate for this application could be easily constructed. For the case of DNA, the chief obstacle in the way of application of the formulae is the necessity for unscrambling contributions of local small distortions and breathing modes as factors that determine molecular length and shape. It is hoped that the application of the formulae given above to experiments on short chains, not long enough for the presence of breathing modes, will help in the resolution of the two effects.

This research was supported in part by Grant GB-27399X2 of the National Science Foundation and Grant GM-20195 of the National Institutes of Health.

References

1. T. Miyazawa, *J. Polymer Sci.*, **55**, 215 (1961).
2. P. Flory, in *Statistical Mechanics of Chain Molecules*, Interscience, New York, 1969.
3. M. V. Volkenstein, in *Configurational Statistics of Polymeric Chains*, Interscience, New York, 1963.
4. H. Triebel, K. Reinert, and J. Strassburger, *Biopolymers*, **10**, 2619 (1971).
5. H. Eisenberg, *Biopolymers*, **8**, 545 (1969).
6. H. B. Gray and J. E. Hearst, *J. Mol. Biol.*, **35**, 111 (1968).
7. M. P. Printz and P. H. von Hippel, *Proc. Nat. Acad. Sci.*, Washington, D. C., **53**, 363 (1965).
8. J. Engel, J. McGhee and P. von Hippel, unpublished results.
9. L. D. Landau and E. M. Lifschitz, in *Statistical Physics*, Addison Wesley, Reading, Mass. 1958, sec. 148, 149.

Received August 6, 1973
Revised October 11, 1973