The Effect of Intrinsic Curvature on Conformational Properties of Circular DNA

Vsevolod Katritch* and Alexander Vologodskii#
*Institute of Molecular Genetics, Moscow 123182, Russia, and #Department of Chemistry, New York University, New York, New York 10003 USA

ABSTRACT Both thermal fluctuations and the intrinsic curvature of DNA contribute to conformations of the DNA axis. We looked for a way to estimate the relative contributions of these two components of the double-helix curvature for DNA with a typical sequence. We developed a model and Monte Carlo procedure to simulate the Boltzmann distribution of DNA conformations with a specific intrinsic curvature. Two steps were used to construct the equilibrium conformation of the model chain. We first specified the equilibrium DNA conformation at the base pair level of resolution, using a set of the equilibrium dinucleotide angles and DNA sequence. This conformation was then approximated by the conformation of the model chain consisting of a reduced number of longer, straight cylindrical segments. Each segment of the chain corresponded to a certain number of DNA base pairs. We simulated conformational properties of nicked circular DNA for different sets of equilibrium dinucleotide angles, different random DNA sequences, and lengths. Only random sequences of DNA generated with equal probability of appearance for all types of bases at any site of the sequence were used. The results showed that for a broad range of intrinsic curvature parameters, the radius of gyration of DNA circles should be nearly independent of DNA sequence for all DNA lengths studied. We found, however, a DNA property that should strongly depend on DNA sequence if the double helix has essential intrinsic curvature. This property is the equilibrium distribution of the linking number for DNA circles that are 300-1000 bp in length. We found that a large fraction of the distributions corresponding to random DNA sequences should have two separate maxima. The physical nature of this unexpected effect is discussed. This finding opens new opportunities for joined experimental and theoretical studies of DNA intrinsic curvature.

INTRODUCTION

It became clear in the late 1970s that some DNA fragments have an intrinsic equilibrium curvature that provides their unusual electrophoretic behavior (Marini et al., 1977). Intensive studies of the phenomenon have exposed many of its features (see reviews (Trifonov, 1985; Diekmann, 1987; Crothers et al., 1990; Hagerman, 1990)). It is now commonly accepted that short 4–6-bp runs of adenines are mainly responsible for the intrinsic curvature of the double helix. Two basic models were suggested to explain the intrinsic curvature on molecular level. The junction model (Selsing et al., 1979; Wu and Crothers, 1984) relates the curvature to deflections at each junction between the axes of the usual B-DNA and the B′-DNA, which is specific for A tracts. According to the junction model, DNA molecules without A tracts cannot have essential intrinsic curvature. The second model proposes smooth bending along the double helix caused, in the first approximation, by small additive wedges (Trifonov, 1980; Trifonov and Sussman, 1980; Bolshoy et al., 1991). The wedges are composed of the roll and tilt angles of independent base pair steps. All 16 wedge angles have been estimated from experimental data (Bolshoy et al., 1991; De Santis et al., 1992).

There are solid experimental data that support the junction model of DNA intrinsic curvature (see Hagerman, 1990, for a review). However, both the junction and the wedge models predict the basic features of the sequence dependence of DNA electrophoretic mobility, and neither of them is able to describe the entire pool of experimental data quantitatively (Haran et al., 1994). Recent data also showed that intrinsic DNA curvature may depend strongly on solution conditions (Sprous et al., 1995). It was found, in addition, that in the presence of some divalent ions, the GGGCCC sequence also provides essential intrinsic curvature (Brukner et al., 1994; Dlakic and Harrington, 1995).

It seems now that both basic models reflect some essential properties of the phenomenon, and instead of asking which model is correct, we should ask, what is the relative contribution of the two mechanisms to intrinsic curvature? The predictions of the two models for the intrinsic curvature of DNA sequences containing phased A-tracts, usually used to study DNA intrinsic curvature, are similar. Thus such sequences do not allow estimation of the relative contribution of the two models. DNA molecules with typical random sequences, however, can help in the study of this issue.

If the wedge model catches the basic physical nature of the phenomenon, DNA with random sequence also should have an essential intrinsic curvature (one can use a simple random sequence as a first approximation of actual DNA sequences; Arneodo et al., 1995). This is not the case if the intrinsic curvature is a result of junctions of B and B′ DNA

Received for publication 20 May 1996 in final form 10 December 1996.

Address reprint requests to Dr. Alex Vologodski, Department of Chemistry, New York University, 4 Washington Place, New York, NY 10003. Tel.: 212-998-3249; Fax: 212-995-4152; E-mail: avologod@jethro.cs.nyu.edu.

Dr. Katritch's present address is Department of Chemistry, Rutgers University, New Brunswick, NJ 08903.

© 1997 by the Biophysical Society

0006-3495/97/03/1070/10 $2.00

helices, because the frequency of appearance of A tracts 4–6 bases in length is very low in a random sequence. Quantitatively average intrinsic curvature of the double helix with a random sequence can be specified in terms of the static persistence length, a_{st} (Trifonov et al., 1987). The persistence length of a polymer chain, a, specifies a decrease of correlation in the directions between the first segment of the chain and the ith segment when i increases (see Cantor and Schimmel, 1980, for example). For a chain consisting of segments of length l, the persistence length can be defined as follows:

$$a = \left\langle (\mathbf{l}_1/l_1) \cdot \sum_{i=1}^{n} \mathbf{l}_i \right\rangle, \qquad n \to \infty, \tag{1}$$

where $\mathbf{l}_1/l_1$ is the unit vector directed along the first segment and $\langle\ \rangle$ corresponds to the averaging over the equilibrium distribution of conformations. In the usual wormlike chain model, which is commonly used to describe DNA properties (Hagerman, 1988), the minimum energy state of the chain is straight, and only thermal fluctuations provide deflections of the DNA axis when we move along the chain contour. Following Trifonov et al. (1987), we define this persistence length as the dynamic one, a_{dyn}. One can use the same equation (Eq. 1) to define a_{st}. In this case we need to consider only deflections of the DNA axis provided by equilibrium bends and take the average through all random sequences. We can say that a_{st} characterizes the average intrinsic curvature of DNA with a random sequence at zero temperature. For nonzero temperatures, both intrinsic curvature and thermal fluctuations will contribute to the properties of a particular DNA (this point was illustrated by Olson et al., 1993). A typical conformation of DNA with randomly distributed intrinsic curvature matches a typical conformation of the wormlike chain with a persistence length, defined by the equation (Trifonov et al., 1987; Schellman and Harvey, 1995)

$$a^{-1} = a_{st}^{-1} + a_{dyn}^{-1}. \tag{2}$$

Because the measured values depend mainly on the size of the molecules in a solution, most of the determination of DNA persistence length gives the value of a rather than a_{dyn}. This is valid for hydrodynamic techniques, light scattering, and, with some restrictions, cyclization kinetics of DNA fragments (see reviews in Hagerman, 1988; Crothers et al., 1992). So the commonly accepted value of the DNA persistence length, a, reflects the contribution from both static and dynamic bends. There is no direct approach to finding the value of a_{st} from experimental data. A few indirect evaluations concluded that equilibrium curvature is equal to or larger than the curvature provided by thermal fluctuations ($a_{dyn} \geq a_{st}$) (Allison et al., 1989; Song and Schurr, 1990; Porschke et al., 1993). It should be noted, however, that only rapid dynamic bending could be observed by Allison et al. (1989) and Song and Schurr (1990); a contribution of a slow bending component to a_{dyn} also could be essential.

However, conformational properties of DNA with any particular sequence depend on a_{dyn} and a_{st} in different ways. Indeed, the DNA sequence does not change during the experiment, and thus the minimum energy conformation is always the same for any particular DNA. The thermal fluctuations take place around this sequence-dependent conformation. Thus the distribution of conformations, as well as measurable DNA properties, must be sequence-dependent. The question is, how large should this dependence be for typical DNA sequences and a particular property?

The purpose of the current paper is to make a quantitative study of the influence of the randomly distributed intrinsic DNA curvature on different DNA conformational properties. Our approach is based on a computer simulation of the equilibrium distribution of DNA conformations. We paid major attention to the equilibrium distribution of linking number, $P(Lk)$ (see Vologodskii, 1992, for example). For actual DNA, $P(Lk)$ can be obtained by the action of topo I enzyme on circular closed DNA or by closing linear molecules with DNA ligase (Depew and Wang, 1975; Pulleyblank et al., 1975). Although the values of linking number, Lk, can only be integers, $P(Lk)$ corresponds to the continuous distribution of the sum of twist, Tw, and writhe, Wr, in the same circular DNA with single-stranded nick. Therefore, one can simulate the equilibrium conformational distribution for nicked circular DNA to calculate the distribution of $P(Lk)$. Comparisons of the results of such simulations with experimental data have given important information about DNA properties (Benham, 1978; Vologodskii et al., 1979; Le Bret, 1980; Frank-Kamenetskii et al., 1985; Levene and Crothers, 1986; Shimada and Yamakawa, 1988; Klenin et al., 1989). We found that the intrinsic curvature can dramatically change the $P(Lk)$ for short circular DNAs. This unexpected effect opens new opportunities for estimating the value of the static persistence length of the double helix from experimental measurements and testing different models of DNA intrinsic curvature.

METHODS OF CALCULATIONS

The model

We developed a DNA model and Metropolis Monte Carlo procedure that make it possible to simulate an equilibrium set of conformations of a circular DNA with intrinsic curvature. Direct analysis of such a set makes it possible to calculate different conformational properties of the model chain. A similar model was introduced earlier by Porschke et al. (1993) for a simpler case of linear molecules with an intrinsic curvature. The construction of the model chain for a particular DNA sequence consists of several steps.

1. *Construction of the minimum energy conformation of a linear chain with intrinsic curvature at the base pair level of resolution.* We used the description of the double helix introduced by Bolshoy et al. (1991). Each base pair of the DNA was represented by a right-handed orthonormal triplet of vectors, (x_i, y_i, z_i). Vector z_i lies along the axis of the double helix, vector y_i is directed to the double-helix major groove. The orientation of

the triplet $i + 1$ relative to the previous triplet i is determined by rotation matrix

$$\mathbf{T}(\theta_i^*, \sigma_i^*, \delta_i^*) = \mathbf{R}(\theta_i^*/2) \cdot \mathbf{W}(\sigma_i^*, \delta_i^*) \cdot \mathbf{R}(\theta_i^*/2), \quad (3)$$

which comprises three consequent rotations: 1) rotation by half of the helical twist $\theta_i^*/2$ about the axis z_i; 2) rotation by the deflection angle σ_i^* about the line on plane (x_i, y_i), perpendicular to the plane of the axis deflection (the direction of the line is given by the angle δ_i^* measured from the new axis x_i); 3) rotation by the angle $\theta_i^*/2$ about the new axis $z_i^* = z_{i+1}$.

Thus we can calculate the sets of triplets as

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{T}(\theta_i^*, \sigma_i^*, \delta_i^*) \times \mathbf{x}_i \\ \mathbf{y}_{i+1} &= \mathbf{T}(\theta_i^*, \sigma_i^*, \delta_i^*) \times \mathbf{y}_i \\ \mathbf{z}_{i+1} &= \mathbf{T}(\theta_i^*, \sigma_i^*, \delta_i^*) \times \mathbf{z}_i. \end{aligned} \quad (4)$$

The minimum energy conformation of the chain consisting of N base pairs is described by N translation vectors $\nu_i = z_i h$, where $h = 0.34$ nm is the length of one step.

We used the wedge model (Bolshoy et al., 1991) to construct the minimum energy conformations. Although other models could be used as well (see Discussion), the wedge model is simple and convenient, as it is based on the assumption that angles θ_i^*, σ_i^*, δ_i^* for each joint are determined only by the dinucleotides flanking this joint (nearest-neighbors model). The values of dinucleotide angles θ_i^*, σ_i^*, δ_i^* were taken from table 3 of Bolshoy et al. (1991) and were constructed randomly to obtain a particular value of a_{st} (see below).

2. *Construction of the minimum energy conformation of a linear chain with extended segments.* To reduce computational time we approximated the original detailed model chain by a chain, consisting of a reduced number of longer, straight cylindrical segments (Fig. 1). A new chain of n vectors $\mathbf{V}_j$ of equal length $l = Nh/n$ was substituted for the original minimum energy trajectory. The new chain started from the same origin, and its vectors $\mathbf{V}_j$ lay on the detail trajectory. To make the end points of the two chains coincide, the vector length l had to be adjusted slightly. We did this by iterations until the distance between the end points of the two chains was less than $l/100$. For each value of l, the ends of vectors $\mathbf{V}_j$ lay on the detail trajectory. Unit vectors $\mathbf{X}_j$,$\mathbf{Y}_j$, which made an orthogonal triplet with vector $\mathbf{V}_j$, were attached to each joint of the chain. We used the following conditions to find rotation matrices $\mathbf{Tr}(\theta_j^0\sigma_j^0\delta_j^0)$, which define the equilibrium chain conformation: 1) matrix $\mathbf{Tr}(\theta_j^0\sigma_j^0\delta_j^0)$ transforms vector $\mathbf{V}_j$ into vector $\mathbf{V}_{j+1}$,

$$\mathbf{V}_{j+1} = \mathbf{Tr}(\theta_j^0, \sigma_j^0, \delta_j^0) \times \mathbf{V}_j; \quad (5)$$

2) helical twist θ_j^0 must be equal to the total helical twist of the corresponding part of the detail chain, including a twist of the first and last base pairs comprising the segment j (according to the share of base pair length related to this segment).

As angle θ_j^0 is determined by condition 2), it is sufficient to find angles σ_j^0 and δ_j^0 to determine matrix $\mathbf{Tr}(\theta_j^0\sigma_j^0\delta_j^0)$. Deflection angle σ_j^0 is the angle between vectors $\mathbf{V}_j$ and $\mathbf{V}_{j+1}$ by definition. Directional angle δ_j^0 can be found as $\delta_j^0 = \lambda - \theta_j^0/2$, where λ_j is the angle between vector $\mathbf{X}_j$ and the projection of vector $\mathbf{V}_{j+1}$ on the $(\mathbf{X}_j, \mathbf{Y}_j)$ plane. The angles $(\theta_j^0\sigma_j^0\delta_j^0)$ determine the rotation matrix $\mathbf{Tr}(\theta_j^0\sigma_j^0\delta_j^0)$ for vectors $(\mathbf{X}_j, \mathbf{Y}_j, \mathbf{V}_j)$:

$$\begin{aligned} \mathbf{X}_{j+1} &= \mathbf{Tr}(\theta_j^0, \sigma_j^0, \delta_j^0) \times \mathbf{X}_j \\ \mathbf{Y}_{j+1} &= \mathbf{Tr}(\theta_j^0, \sigma_j^0, \delta_j^0) \times \mathbf{Y}_j \\ \mathbf{V}_{j+1} &= \mathbf{Tr}(\theta_j^0, \sigma_j^0, \delta_j^0) \times \mathbf{V}_j. \end{aligned} \quad (6)$$

Thus, choosing the first pair of unit vectors $(\mathbf{X}_1, \mathbf{Y}_1)$ randomly, we can find, by induction, angles $(\theta_j^0\sigma_j^0\delta_j^0)$ and unit vectors $(\mathbf{X}_j, \mathbf{Y}_j)$ for all joints of the minimum energy conformation of the new chain. To calculate the matrix $\mathbf{Tr}(\theta_j^0\sigma_j^0\delta_j^0)$, the linear chain must contain one extra segment, the sequence of which is a repeat of the sequence of the last segment. This algorithm of "resegmentation" makes it possible to build a model chain with any number of base pairs per segment, k.

3. *Construction of circular chain conformation.* One can arbitrarily construct a closed chain of n vectors of equal length $\mathbf{V}_j$, choosing orthogonal vectors $\mathbf{X}_1$ and $\mathbf{Y}_1$ for the first vector $\mathbf{V}_1$ and ascribing a certain value of helical twist θ_j to each joint. This choice makes it possible to find the set of n rotation matrices $\mathbf{Tr}(\theta_j, \sigma_j, \delta_j)$ and $n + 1$ triplets of vectors $(\mathbf{X}_j, \mathbf{Y}_j, \mathbf{V}_j)$ by the procedure described in step 2. To make the chain continuous in the nth joint, we supposed here that the additional "virtual" vector $\mathbf{V}_{n+1}$ (see Eq. 6) coincides with the first vector of the closed chain $\mathbf{V}_1$. In this paper we considered circular DNA with a single-stranded break (nick), which allows free torsional rotation in the base pair stack at the break point. Therefore, vectors $\mathbf{X}_{n+1}$ and $\mathbf{X}_1$, which define the torsional orientations for the ends of the chain, were not forced to coincide. The value of angle Ω between vectors $\mathbf{X}_{n+1}$ and $\mathbf{X}_1$ in our case of nicked chain had no impact on the energy of the chain (see below, Eq. 7).

The set of n triplets of angles θ_j, σ_j, δ_j is now the set of parameters that fully define the chain conformation. The energy of the chain, E, was calculated as the sum of torsional and bending energies over all the joints between the chain segments:

$$E = \sum_{j=1}^{n} \frac{C}{2Nh}(\theta_j - \theta_j^0)^2 + \sum_{j=1}^{n} g\alpha_j^2, \quad (7)$$

where C is the torsional rigidity constant; θ_j^0 and θ_j are the equilibrium and current values of helical twist, respectively; g is the bending rigidity constant; and α_j is the angle between vectors $\mathbf{V}_j$ and $\mathbf{Tr}(\theta_j^0\sigma_j^0\delta_j^0)\cdot\mathbf{V}_{j-1}$. This definition of the bending energy implies isotropic bending rigidity.

Excluded-volume effects were incorporated into the model via the concept of DNA effective diameter d. This is the actual diameter of the cylindrical segments of the model. The value of d takes into account not only the geometrical diameter of DNA segments, but also electrostatic repulsion between them. The value of d depends strongly on ionic condi-

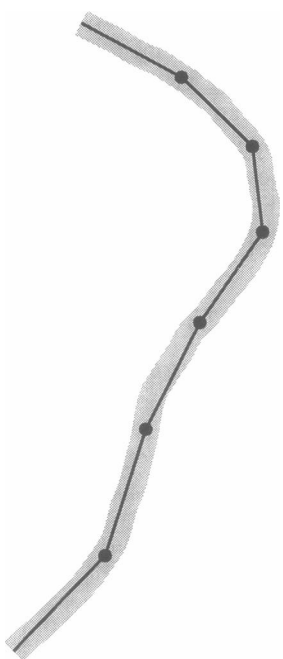


FIGURE 1 Two models of intrinsically curved DNA. Projection of the minimum energy conformation of the model chain is shown by the bold gray line at the base-pair level of resolution; thin black segments display the corresponding conformation of the chain with extended segments. One extended segment corresponds to 30 bp in this diagram.

tions. The accuracy of this approximation of the actual electrostatic inter-segment interactions was considered recently by Vologodskii and Cozzarelli (1995).

Simulation procedure

We used the Metropolis Monte Carlo procedure to obtain the equilibrium set of chain conformations (Metropolis et al., 1953). Three types of moves were used for successive changes of chain conformations:

i) Rotation of a subchain $\{\mathbf{V}_m \ldots \mathbf{V}_{m+p}\}$ with arbitrary numbers m and p about the axis that connects the ends of the subchain by the angle ϕ, randomly chosen in range $[-\phi_{max}, \phi_{max}]$. This movement changes values of wedge angles σ in joints m and $m + p + 1$, all directional angles δ in joints from m to $m + p + 1$, but does not change any value of helical twist θ.

ii) Change of the helical twist θ for an arbitrary segment m by the angle γ, randomly chosen in range $[-\gamma_{max}, \gamma_{max}]$; the movement does not change any value of σ, but it changes values of δ for all joints from m to the end of the chain.

iii) Rotation of vector $\mathbf{X}_1$ about $\mathbf{V}_1$ by angle ρ, randomly chosen in range $[-\rho_{max}, \rho_{max}]$. This movement does not change any values of σ and θ, but changes all directional angles δ by angle ρ, thus allowing the torsional rotation of the chain as a whole. Movement 3) was not obligatory (as values of δ_j changed in 1) and 2)), but it made it possible to make in one step such changes that would take many steps to make by other types of movements.

Although one cannot prove strictly that these moves provide a Boltzmann distribution of conformations for the model chain in real simulation runs, it is reasonable to believe that this is the case. First, a Boltzmann distribution should be reached when the number of steps tends to infinity, because the moves used in our procedure satisfy microscopic reversibility, the only necessary requirement of the Metropolis procedure (Metropolis et al., 1953). This principle states that if a conformation B has a probability P of being chosen as a trial state for the energy test in the Metropolis procedure when a current conformation is A, we must have the same probability P of choosing the conformation A as a trial state for the energy test if a current conformation is B.

The proof of this principle for the moves used can be done by direct calculations of P according to the move descriptions.

Second, we showed that number of steps used in our work was large enough to equilibrate the system. We determined, for many chains with specific intrinsic curvatures, that the simulated distributions of chain conformations were independent of starting conformations or the choice of seed for the random number generator (see Fig. 7 *B*).

In the Monte Carlo procedure, a chain conformation had infinite energy if the distance between two nonadjacent segments was smaller than the value of d. Therefore, the minimum distance between all pairs of nonadjacent segments of a trial conformation was calculated. If any distance was less than d, the trial conformation was rejected.

After each movement we recalculated angles θ_j, σ_j, δ_j and found the energy of the trial conformation, which was accepted or rejected according to the energy test of the Metropolis procedure (Metropolis et al., 1953). The values of ϕ_{max}, γ_{max}, and ρ_{max} were adjusted so that about half of the corresponding type of moves were successful. The particular type of the movement for each step of the Monte Carlo procedure was chosen randomly with probabilities $P(i) = 0.8$, $P(ii) = 0.1$, and $P(iii) = 0.1$. These values approach the optimum for achieving the equilibrium sampling.

We used the special control and correction procedure to avoid the accumulation of computational error for coordinates of the chain. After each correction all angles θ_j, δ_j, δ_j were recalculated.

The writhe, Wr, the total twist, Tw, and the linking number, Lk, were calculated for every conformation. Our model allows independent calculation of Wr and Tw of the chain. Writhe was calculated by Le Bret's algorithm (Le Bret, 1980). Helical twist is simply the sum of the helical twists of all segments (in units of number of turns):

$$Tw = \frac{1}{2\pi}\sum_{i=1}^{n}\theta_i. \qquad (8)$$

We used the equation

$$Lk = Tw + Wr \qquad (9)$$

to calculate the Lk value. We also used an additional method to calculate a fractional part of Lk (Lk does not necessarily equal an integer in our model). This fractional part simply equals the angle Ω between vectors $\mathbf{X}_{n+1}$ and $\mathbf{X}_1$. The fractional part of Lk calculated by these two ways turned out to be the same up to the fourth character, even for rather long chains. This is a good check for the accuracy of chain construction. It is important that such a level of accuracy can be reached only when double-precision variables (64 bits for a variable) were used for calculations.

Parameters of the model

To specify the intrinsic curvature, we used the set of dinucleotide angles proposed by Bolshoy et al. (1991), set BT, as well as two other randomly generated sets, set 1 and set 2. The value of a_{st} was equal to 168 nm for each of these sets. To generate sets 1 and 2 we ascribed random values to direction angles δ and wedge angles σ for different dinucleotides. Then we adjusted the values of all wedge angles by multiplying them by some constant to get the required value of a_{st}. We also constructed sets of angles with a_{st} values of 100 nm, 250 nm, 400 nm, and 800 nm by multiplying all wedge angles of set BT by the appropriate constant. Twist angles θ for all sets were the same as for the BT set.

The values of the bending rigidity constant, g, were chosen so that the intrinsically straight model chain had a particular value of a_{dyn} for a particular value of k (Frank-Kamenetskii et al., 1985).

We calculated the values of a_{st} and a for a particular set of equilibrium dinucleotide angles by direct simulation of long linear chains and by using the equation

$$a = \frac{\langle r^2\rangle}{2Nh}, \qquad (10)$$

where $\langle r^2\rangle$ is the mean-square end-to-end distance. Thermal fluctuations were included in the algorithm of the chain construction when we calculated the value of a and were not included when we calculated the value of a_{st}. The chain self-intersections were ignored during those calculations to avoid disturbing Eq. 10. We found that Eq. 2 was valid, with an accuracy of better than 1% for the sets of parameters we studied.

We used random sequences of DNA generated with equal probability of appearance for all types of bases at any site of the sequence. The sequences used in the paper are available from the authors to the interested reader on request.

Other parameters of the model were taken close to their usual values at physiological conditions: torsional rigidity, C, was equal to 2.5×10^{-19} erg·cm (Hagerman, 1988), and the DNA effective diameter was equal to 4 nm (Vologodskii and Cozzarelli, 1995).

RESULTS

We studied the effect of the distributed intrinsic curvature on conformational properties of circular DNA for different sets of equilibrium dinucleotide angles. The sets corresponded to the values of a_{st} between 100 nm and 800 nm. For all cases the values of a_{dyn} were chosen so that the value of a, defined by Eq. 2, was equal to 50 nm.

Radius of gyration

It is general practice to specify the size of a polymer chain in solution by the mean square root radius of gyration, R_g, the value of which can be measured by light scattering (see

Flory, 1969, for example). We studied the effect of intrinsic curvature on the value of R_g. Our simulations showed that partitioning between static and dynamic persistence lengths has little impact on the value of R_g of circular DNA. Fig. 2 shows that both intrinsically straight circular DNAs and closed chains with essential randomly distributed intrinsic curvature (a_{st} = 168 nm) should have nearly the same values of R_g as soon as they are characterized by the same value of a. This conclusion is valid for the whole range of DNA lengths studied, from 300 bp to 4000 bp. The scatter of R_g for different sequences was very small, although the value of a_{st} was comparable to the value of a_{dyn} (Fig. 2). We concluded from this result that R_g depends mainly on the value of a.

The equilibrium distribution of linking number

At the beginning of this study we suggested that only thermal fluctuations contribute to the width of the equilibrium distribution of the linking number (see Vologodskii, 1992, for example), which should depend mainly on the value of a_{dyn}. Our simulation results showed that the actual picture is more complex. First, we found that the DNA sequence has a large impact on $P(Lk)$. The scatter of the distribution variance, Var$\{Lk\}$, over different sequences becomes very large for small DNA circles (300–1000 bp in length) if the value of a_{st} is comparable to the value of a_{dyn} (Fig. 3). Because we were looking for a conformational property that would be the most sensitive to the intrinsic curvature and is easily measurable, we used the DNA length of 900 bp in the following calculations.

We also found that a large scatter of the Var$\{Lk\}$ for different sequences is primarily defined by the value of a_{st} rather than the particular set of dinucleotide angles that specify the intrinsic curvature. Three different sets of dinucleotide angles, BT, 1, and 2, which specified the same value of a_{st}, 168 nm, gave approximately the same scatter of Var$\{Lk\}$ for several random sequences 900 bp in length (Fig. 3).

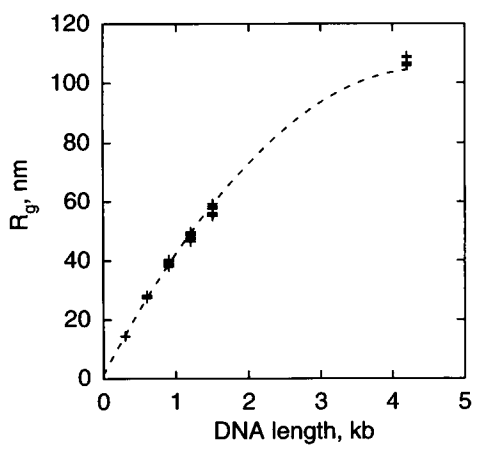


FIGURE 2 Simulated radius of gyration, R_g, of intrinsically curved circular DNAs. Each simulation result (+) corresponds to one randomly generated DNA sequence. The value of a_{st} was equal to 168 nm. The dashed line shows the corresponding dependence for intrinsically straight DNA with the same value of a.

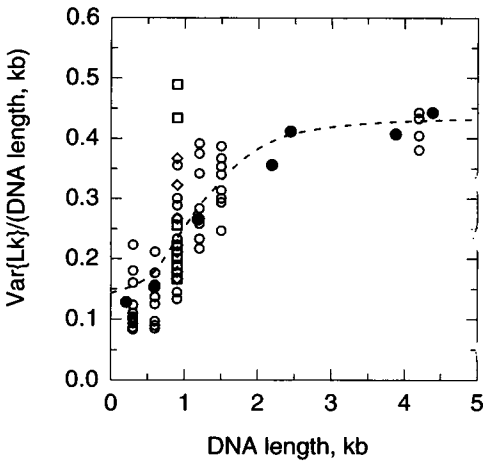


FIGURE 3 Simulated variance of the equilibrium distributions of the linking number, Var$\{Lk\}$, of intrinsically curved circular DNAs. Each simulation result (○, □, ◇) corresponds to one randomly generated DNA sequence. Three sets of equilibrium dinucleotide angles, set BT (○), set 1 (□), and set 2 (◇), were used in the simulations. For each set the value of a_{st} was equal to 168 nm. Experimental data from published works (Depew and Wang, 1975; Pulleyblank et al., 1975; Horowitz and Wang, 1984) are also shown (●). The dashed line shows the corresponding dependence for intrinsically straight DNA with the same value of a.

We also studied how the value of a_{st} has to affect the sequence dependence of Var$\{Lk\}$. Fig. 4 presents the values of Var$\{Lk\}$ simulated for different sequences 900 bp in length and for different values of a_{st}. As well as for all other results, the value of a was equal to 50 nm and the value of

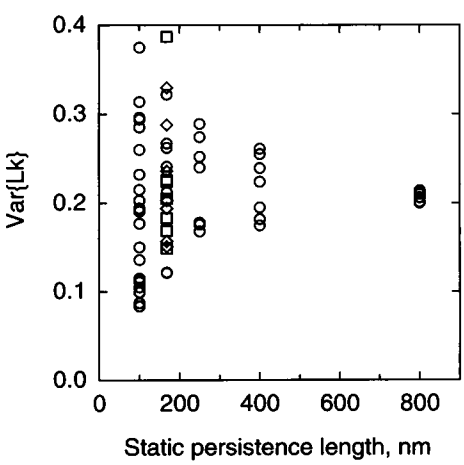
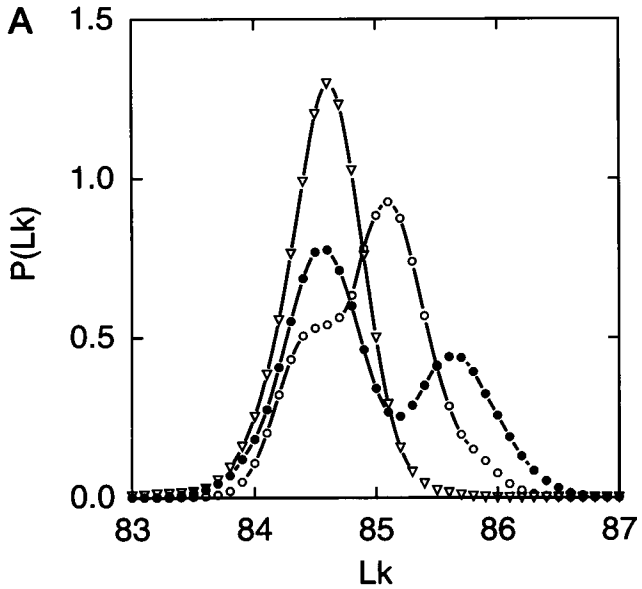
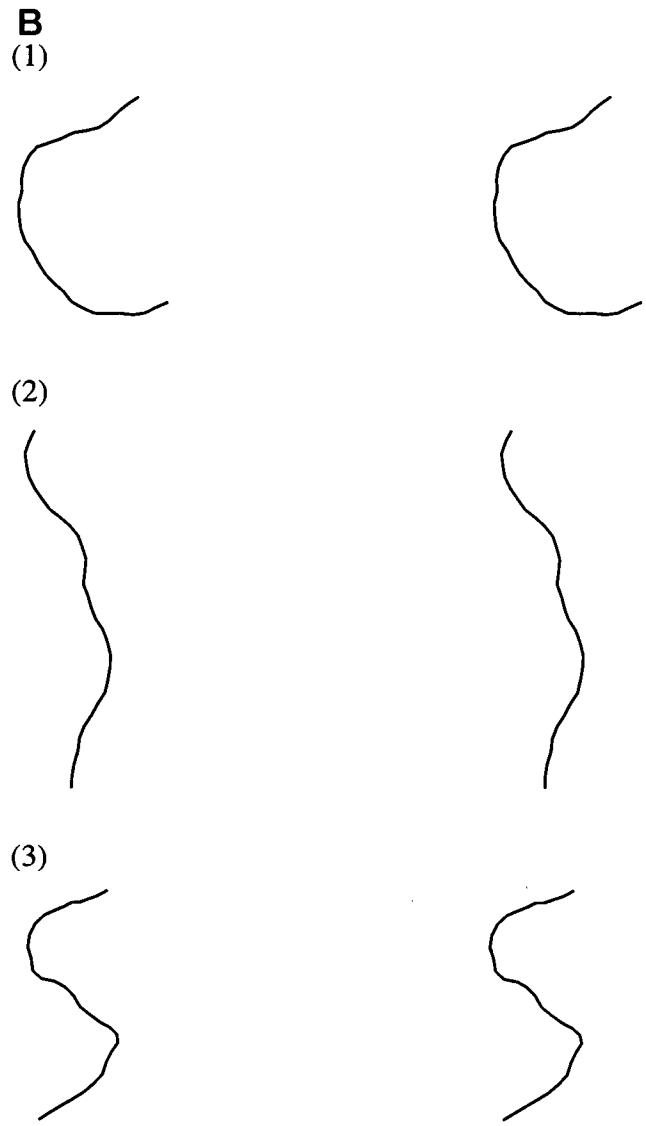


FIGURE 4 Dependence of the variance of the equilibrium distributions of the linking number, Var$\{Lk\}$, on the value of the static persistence length, a_{st}. The model chains with randomly generated sequences corresponded to a DNA length of 900 bp. Three sets of equilibrium dinucleotide angles, set BT (○), set 1 (□), and set 2 (◇), were used in the simulations for a_{st} of 168 nm. The sets of dinucleotide angles for other values of a_{st} were constructed by multiplying all wedge angles of set BT by an appropriate constant. The value of a was equal to 50 nm in all cases.





a_{dyn} changed according to Eq. 2. Therefore, the figure shows that the average values of Var{Lk} over different sequences must depend mainly on the value of the total persistence length, a, rather than on partitioning between a_{dyn} and a_{st}. The figure shows that the sequence dependence of Var{Lk} must be essential, even for rather large values of a_{st} (small intrinsic curvature).

The simulated distributions of Lk, $P(Lk)$, for intrinsically straight DNA were always found to be Gaussian with high accuracy (Vologodskii et al., 1979; Le Bret, 1980; Frank-Kamenetskii et al., 1985; Levene and Crothers, 1986; Shimada and Yamakawa, 1988; Klenin et al., 1989). Our simulation showed that this must be different for intrinsically curved DNA. The simulated shapes of $P(Lk)$ depended strongly on the DNA sequences and a set of equilibrium dinucleotide angles. Typical distributions $P(Lk)$ for three different randomly generated DNA sequences 900 bp in length are presented in Fig. 5, together with the equilibrium (minimum energy) conformations of these model chains in linear form. We found for such chains that 25–35% distributions had a more or less regular shape close to a Gaussian distribution; 15–25% had two distinct maxima; and other distributions (40–60%) had intermediate shapes, not regular but with one maximum. Approximately the same picture was observed for the set of dinucleotide angles that corresponded to the value of a_{st} = 100 nm, but we did not observe distributions with more than one maximum for $a_{st} \geq 250$ nm.

We made detailed analysis of the distributions $P(Lk)$ for a DNA length of 900 bp. The analysis showed that most of the distributions can be fitted by a sum of two Gaussians with various weights, positions, and widths (a small fraction of the distributions were fitted by a single Gaussian). Such composite distributions fitted the original ones with extremely high accuracy. For set 3 of dinucleotide angles (a_{st} = 100 nm), only 4 of 25 distributions could be fitted by a single Gaussian, 4 consisted of three Gaussians, and the remaining 17 were accurately fitted by two Gaussians. These findings led us to the hypothesis that intrinsically curved closed chains can possess two (or sometimes more) local minima of elastic energy (unlike straight chains, which always have only one minimum energy conformation in the form of a regular circle).

To test this hypothesis we developed a special variant of the Monte Carlo procedure with slowly (exponentially) decreasing temperature from 1000 K to 0.01 K. Ten cycles of heating-annealing were performed for each of a few chains to find their minimum energy conformations. We found that chain conformations gradually converged to one or two distinct conformations when temperature decreased.

FIGURE 5 Simulated distributions of the linking number, $P(Lk)$, for intrinsically curved DNAs. The distributions for three randomly generated DNA sequences 900 bp in length, 1 (▽), 2 (○), 3 (●), are shown in *A*. A set of dinucleotide angles corresponding to a_{st} of 168 nm was used in the simulations. Stereoscopic views of the minimum energy conformations of the corresponding linear chains are shown in *B*.

Fig. 6 shows an example of two conformations of one chain that correspond to different minima of the elastic energy. The values of Lk that corresponded to the minima of the elastic energy were very close to the maxima of the Gaussians that fitted the corresponding distributions $P(Lk)$ at room temperature. We found a few examples in which the number of minimum energy conformations was less than the number of Gaussians required to fit $P(Lk)$. We think that the extra minima of elastic energy were not deep enough to be found by the annealing procedure.

The difference in Lk values for the conformations that corresponded to minima of elastic energy were about 1 for typical distributions with two maxima (see Figs. 5 and 7). It is interesting that the greater part of this difference was provided by the difference in the twist values of those conformations, and the difference in the values of writhe was rather small. All minimum energy conformations were relatively close to flat circles. This explains why all of these short circles, including those with two minima of elastic energy, have very close values of R_g, and we observed only a small scatter of R_g for different sequences.

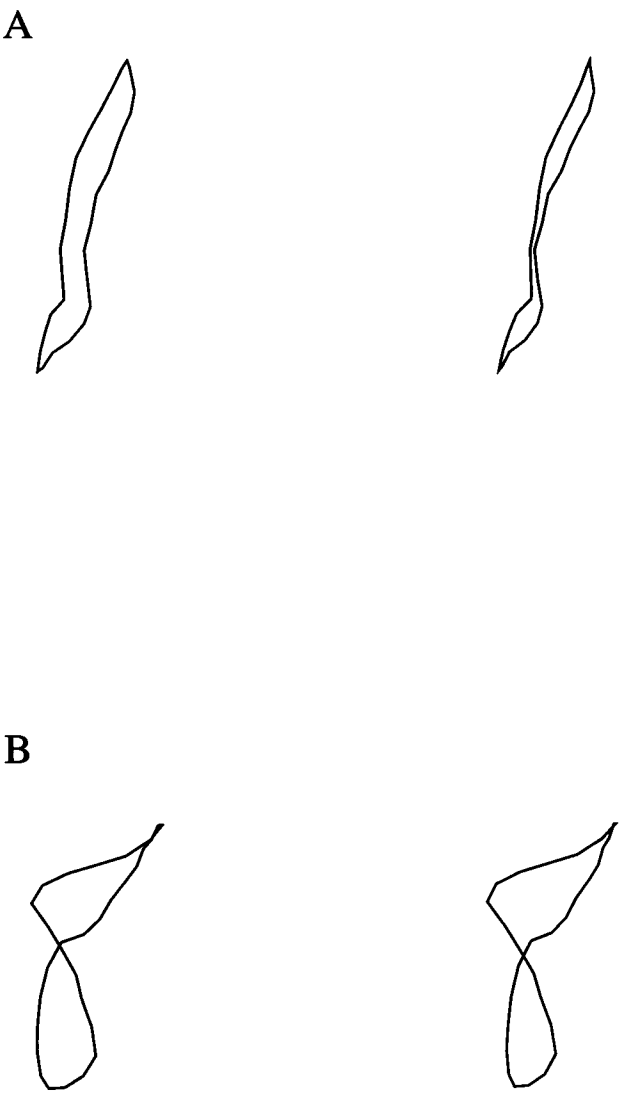


FIGURE 6 Stereoscopic views of two conformations of one model chain 900 bp in length, which correspond to minimum elastic energy. The intrinsic curvature of the chain corresponded to a value of a_{st} of 100 nm. The elastic energy of conformation A is $0.93RT$ less than the energy of conformation B. The values of Lk and writhe for these conformations are equal to 84.02 and -0.062 (A), and 85.12 and 0.302 (B).

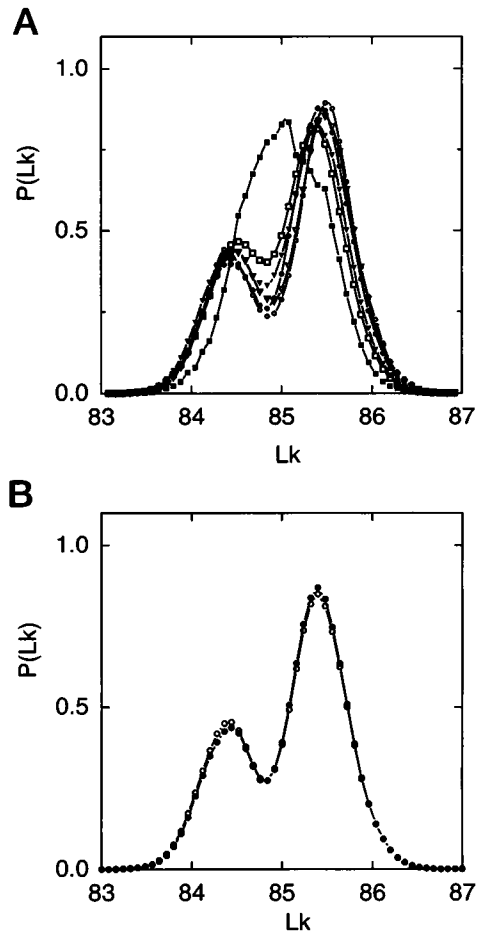


FIGURE 7 Effect of segmentation of the model chain and reproducibility of the simulated distributions $P(Lk)$. All distributions were simulated for one particular sequence 900 bp in length. The set of dinucleotide angles corresponded to a value of a_{st} of 168 nm. The number of base pairs per one model segment, k, was equal to 5 (○), 10 (●), 30 (▽), 60 (▼), 100 (□), and 150 (■) (A). For k = 30, the same distribution was obtained in two independent simulation runs (B).

Sensitivity of the results to segment length and model parameters

Our DNA model approximated the double helix by a chain of straight segments. One segment of the model chain corresponded to a certain number of DNA base pairs, k. The value of k should be small enough to make sure that chain characteristics under study are the same as for the detail model with k = 1. Therefore, we studied how the choice of k affects the conformational properties of circular DNA. The simulated distributions $P(Lk)$ for one particular DNA

sequence 900 bp in length and a particular set of dinucleotide angles are shown in Fig. 7 *A* for different values of k. The distributions are nearly the same for $k = 5$, 10, 30, taking into account the statistical error of the simulations (Fig. 7 *B*). Although the depth of the minima between two peaks decreases for $k = 60$ and 100, the distribution still contains its major features. For this particular sequence, only $k = 150$ essentially changed the distribution (Fig. 7 *A*). Only minor changes of the average radius of gyration were found for $k < 100$. We observed a similar picture for model chains with other sequences. Therefore, we used the value $k = 30$ for most of the calculations. For chains 600 and 300 bp in length we used the value $k = 15$.

Fig. 7 *A* shows that our model with extended segments gives a very good approximation of the detailed DNA model ($k = 1$). It is not so surprising if we take into account that the model with extended segments is practically never off by more than one DNA geometrical radius from a conformation of the detailed model (see Fig. 1). At the same time, the model with extended segments provides a great savings in computer time. The computer time needed for a simulation increases approximately as the square of the number of chain segments. Thus the computer time is reduced by a factor of 1000 when we reduce the number of chain segments by a factor of 30. It would be impossible to obtain results shown above by using the detailed model. Dynamic simulations of short supercoiled molecules (Tan and Harvey, 1990) that use a detailed DNA model give a good illustration of the problem.

We also checked that the effect of intrinsic curvature is rather stable within the reasonable range of other physical parameters of our model, such as torsional rigidity C, a_{dyn}, and the effective diameter of the model chains, d. We found that the distribution $P(Lk)$ for the chain 900 bp in length contained a specific shape, and the value of $\mathrm{Var}\{Lk\}$ changed by less than 10% for values of C from 2×10^{-19} to 3×10^{-19} erg·cm, a_{dyn} from 50 to 80 nm, and d from 0 to 8 nm.

DISCUSSION

Our results open new opportunities to study the effect of the DNA intrinsic curvature, comparing the thermal distributions of topoisomers, $P(Lk)$, for short circles with different sequences. There are two features of $P(Lk)$ that make its use very promising. First, $P(Lk)$ for short circles is very sensitive to the specific pattern of the intrinsic curvature. The second point is that we can calculate $P(Lk)$ for any specific pattern of intrinsic bends. We are not able to make such calculations for the mobility of DNA molecules in a gel, and this is a serious limitation of the gel retardation method, a major technique in the study of DNA intrinsic curvature. The approach based on the cyclization rate of DNA fragments was the only solution technique so far that made it possible to connect curvature with a measurable effect (Crothers et al., 1992). All quantitative analysis of intrinsic curvature is based on this approach. We found another property of intrinsically curved DNA for which a measurable characteristic is a known function of the equilibrium geometry of the DNA axis.

We showed that short circular DNA molecules with random sequences can often have more than one minimum energy conformation if their intrinsic curvature is comparable to or larger than the curvature provided by thermal fluctuations. One can observe such minima by playing with curved steel wire closed into a circle. These energy minima strongly affect the conformational properties of such short molecules, 300-1000 bp in length, and result in a strong dependence of simulated distributions $P(Lk)$ on DNA sequence. For DNA circles 900 bp in length, about 20% of the distributions should have two maxima if the values of static and dynamic persistence lengths are close to one another. The scatter of $\mathrm{Var}\{Lk\}$ over different DNA sequences must be large for such cases: among 10 randomly taken sequences there were usually two for which the values of $\mathrm{Var}\{Lk\}$ differed by a factor of 3 (see Fig. 3). The effect gradually disappears with increasing chain length because minimum energy conformations of DNA molecules longer than 1500 bp hardly affect the Boltzmann distribution of conformations.

Fig. 5 *B* illustrates one of the possible features of intrinsic shape that result in bimodal $P(Lk)$. Trajectory 3, which corresponds to bimodal $P(Lk)$ in Fig. 5 *A*, has the shape of a spatially distorted character S. The same S-shaped features were found for all studied sequences that gave two minimum energy conformations and strongly bimodal distributions $P(Lk)$. On the other hand, C-shaped molecules (like trajectory 1 in Fig. 5 *B*) give very narrow distributions $P(Lk)$, narrower than intrinsically straight molecules. It looks like such C-shaped intrinsic curvature stabilizes the shape of a circular chain. To confirm this relation between the shape of linear DNA and $P(Lk)$, we also did simulations for specially constructed chains of 900-bp length that had regular S and C shapes (data not shown).

Of course, this classification is very rough and cannot represent all of the diversity of equilibrium shapes of intrinsically curved molecules with random sequences. We use it here only to illustrate one of the simplest methods of obtaining different conformational behaviors of short circular DNA molecules with intrinsic curvature. Other features of the intrinsic shape of DNA molecules may be important as well.

It is important that the strong dependence of $P(Lk)$ on the DNA sequence of short circles is not a specific result of a particular set of dinucleotide angles that defines the intrinsic curvature. We tested the different sets that were characterized by the same value of a_{st} and observed about the same changes in $P(Lk)$ for different sequences. Although we used the simplest model of intrinsic curvature, the nearest-neighbor model, it is clear enough that more complex models would give the same results if they provide the same values of a_{st}.

We also found that the distributed intrinsic curvature must have very little impact on the chain size. The values of R_g obtained in the simulations depended, to a good approximation, only on the value of a. Maybe even more unexpected was the observation that Var$\{Lk\}$ for DNA circles a few thousand base pairs in length also depended, to a good approximation, on the value of a only. These results strongly support the suggestion of Trifonov, Tan, and Harvey that conformational properties of long DNA molecules should depend only on the value of a rather than on its static and dynamic components (Trifonov et al., 1987).

There are a few experimental works in which the Var$\{Lk\}$ of equilibrium distributions of Lk were measured (Depew and Wang, 1975; Pulleyblank et al., 1975; Shure et al., 1977; Lee et al., 1981; Shore and Baldwin, 1983; Horowitz and Wang, 1984). However, most of the available experimental data were obtained for DNA molecules a few thousand base pairs in length and thus are not relevant to the effect we found. We plotted available experimental data in Fig. 3 together with the simulation results. The simulated data correspond to the value of $a_{st} = 168$ nm and the value of $a_{dyn} = 71$ nm. It seems that experimental points can be described by one smooth dependence close to the simulation results for intrinsically straight DNA (also shown in Fig. 3). If future experimental measurements confirm this, we will have to conclude that the value of the static persistence length is essentially larger than the value of the dynamic persistence length. However, more experimental data will be needed to come to any certain conclusion.

We thank the National Institutes of Health (grant GM54215 to AV) for their support.

REFERENCES

Allison, S., R. Austin, and M. Hogan. 1989. Bending and twisting dynamics of short DNAs. Analysis of the triplet anisotropy decay of a 209 base pair fragment by Brownian simulation. *J. Chem. Phys.* 90:3843–3854.

Arneodo, A., E. Bacry, P. V. Graves, and J. F. Muzy. 1995. Characterizing long-range correlations in DNA sequences from wavelet analysis. *Phys. Rev. Lett.* 74:3293–3296.

Benham, C. J. 1978. The statistics of superhelicity. *J. Mol. Biol.* 123: 361–70.

Bolshoy, A., P. McNamara, R. E. Harrington, and E. N. Trifonov. 1991. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. USA.* 88:2312–2316.

Brukner, I., S. Susic, M. Dlakic, A. Savic, and S. Pongor. 1994. Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J. Mol. Biol.* 236:26–32.

Cantor, C. R. and P. R. Schimmel. 1980. Biophysical Chemistry. W. H. Freeman and Company, New York.

Crothers, D. M., J. Drak, J. D. Kahn, and S. D. Levene. 1992. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.* 212:3–29.

Crothers, D. M., T. E. Haran, and J. G. Nadeau. 1990. Intrinsically bent DNA. *J. Biol. Chem.* 265:7093–7096.

Depew, R. E., and J. C. Wang. 1975. Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. USA.* 72:4275–4279.

De Santis, P., A. Palleschi, M. Savino, and A. Scipioni. 1992. Theoretical prediction of the gel electrophoretic retardation changes due to point mutations in a tract of SV40 DNA. *Biophys. Chem.* 42:147–152.

Diekmann, S. 1987. DNA curvature. *In* Nucleic Acids and Molecular Biology. F. Eckstein and D. Lilley, editors. Springer, Berlin. 138–156.

Dlakic, M., and R. Harrington. 1995. Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. *J. Biol. Chem.* 270:29945–29952.

Flory, P. J. 1969. Statistical Mechanics of Chain Molecules. Wiley, New York.

Frank-Kamenetskii, M. D., A. V. Lukashin, V. V. Anshelevich, and A. V. Vologodskii. 1985. Torsional and bending rigidity of the double helix from data on small DNA rings. *J. Biomol. Struct. Dyn.* 2:1005–1012.

Hagerman, P. J. 1988. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* 17:265–286.

Hagerman, P. J. 1990. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* 59:755–81.

Haran, T. E., J. D. Kahn, and D. M. Crothers. 1994. Sequence elements responsible for DNA curvature. *J. Mol. Biol.* 244:135–143.

Horowitz, D. S., and J. C. Wang. 1984. Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J. Mol. Biol.* 173:75–91.

Klenin, K. V., A. V. Vologodskii, V. V. Anshelevich, V. Y. Klisko, A. M. Dykhne, and M. D. Frank-Kamenetskii. 1989. Variance of writhe for wormlike DNA rings with excluded volume. *J. Biomol. Struct. Dyn.* 6:707–714.

Le Bret, M. 1980. Monte Carlo computation of supercoiling energy, the sedimentation constant, and the radius of gyration of unknotted and circular DNA. *Biopolymers.* 19:619–637.

Lee, C. H., H. Mizusawa, and T. Kakefuda. 1981. Unwinding of double-stranded DNA helix by dehydration. *Proc. Natl. Acad. Sci. USA.* 78: 2838–2842.

Levene, S. D., and D. M. Crothers. 1986. Topological distributions and the torsional rigidity of DNA. A Monte Carlo study of DNA circles. *J. Mol. Biol.* 189:73–83.

Marini, J. C., R. Weisberg, and A. Landy. 1977. The isolation of restriction fragments containing the primary and secondary (galT) bacterial att sites of phage lambda. *Virology.* 83:254–270.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Olson, W. K., N. L. Marky, R. L. Jernigan, and V. B. Zhurkin. 1993. Influence of fluctuations on DNA curvature. A comparison of flexible and static wedge models of intrinsically bent DNA. *J. Mol. Biol.* 232: 530–554.

Porschke, D., E. R. Schmidt, T. Hankeln, G. Notle, and J. Antosiewicz. 1993. Structure and dynamics of curved DNA fragments in solution: evidence for slow modes of configurational transitions. *Biophys. Chem.* 47:179–191.

Pulleyblank, D. E., M. Shure, D. Tang, J. Vinograd, and H. P. Vosberg. 1975. Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: formation of a Boltzmann distribution of topological isomers. *Proc. Natl. Acad. Sci. USA.* 72:4280–4284.

Schellman, J. A., and S. C. Harvey. 1995. Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature. *Biophys. Chem.* 55:95–114.

Selsing, E., R. D. Wells, C. J. Alden, and S. Arnott. 1979. Bent DNA: visualisation of a base-paired and stacked A-B conformational junction. *J. Biol. Chem.* 254:5417–5422.

Shimada, J., and H. Yamakawa. 1988. Moments for DNA topoisomers: the helical wormlike chain. *Biopolymers.* 27:657–673.

Shore, D., and R. L. Baldwin. 1983. Energetics of DNA twisting. II. Topoisomer analysis. *J. Mol. Biol.* 170:983–1007.

Shure, M., D. E. Pulleyblank, and J. Vinograd. 1977. The problems of eukaryotic and prokaryotic DNA packaging and in vivo conformation posed by superhelix density heterogeneity. *Nucleic Acids Res.* 4:1183–1205.

Song, L., and J. M. Schurr. 1990. Dynamic bending rigidity of DNA. *Biopolymers.* 30:229–237.

Sprous, D., W. Zacharias, Z. A. Wood, and S. C. Harvey. 1995. Dehydrating agents sharply reduce curvature in DNAs containing A tracts. *Nucleic Acids Res.* 23:1816–1821.

Tan, R. K. Z., and S. C. Harvey. 1990. Succinct macromolecular models: application to supercoiled DNA. *In* Theoretical Biochemistry and Molecular Biophysics. D. L. Beveridge and R. Lavery, editors. Adenine Press, Guilderland, NY. 125–137.

Trifonov, E. N. 1980. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Res.* 8:4041–4053.

Trifonov, E. N. 1985. Curved DNA. *CRC Crit. Rev. Biochem.* 19:89–106.

Trifonov, E. N., and J. L. Sussman. 1980. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. USA.* 77: 3816–3820.

Trifonov, E. N., R. K. Z. Tan, and S. C. Harvey. 1987. Static persistence length of DNA. *In* DNA Bending and Curvature. W. K. Olson, M. H. Sarma, R. H. Sarma, and M. Sundaralingam, editors. Adenine Press, New York. 243–253.

Vologodskii, A. V. 1992. Topology and Physics of Circular DNA. CRC Press, Boca Raton, FL.

Vologodskii, A. V., and N. R. Cozzarelli. 1995. Modeling of long-range electrostatic interactions in DNA. *Biopolymers.* 35:289–296.

Vologodskii, A. V., A. V. Lukashin, V. V. Anshelevich, and M. D. Frank-Kamenetskii. 1979. Fluctuations in superhelical DNA. *Nucleic Acids Res.* 6:967–982.

Wu, H. M., and D. M. Crothers. 1984. The locus of sequence-directed and protein-induced DNA bending. *Nature.* 308:509–513.