Equilibrium distributions of topological states in circular DNA: Interplay of supercoiling and knotting

Alexei A. Podtelezhnikov*, Nicholas R. Cozzarelli†‡, and Alexander V. Vologodskii*

*Department of Chemistry, New York University, New York, NY 10003; and †Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Contributed by Nicholas R. Cozzarelli, September 16, 1999

Two variables define the topological state of closed double-stranded DNA: the knot type, *K*, and Δ*Lk*, the linking number difference from relaxed DNA. The equilibrium distribution of probabilities of these states, *P*(Δ*Lk*, *K*), is related to two conditional distributions: *P*(Δ*Lk*|*K*), the distribution of Δ*Lk* for a particular *K*, and *P*(*K*|Δ*Lk*) and also to two simple distributions: *P*(Δ*Lk*), the distribution of Δ*Lk* irrespective of *K*, and *P*(*K*). We explored the relationships between these distributions. *P*(Δ*Lk*, *K*), *P*(Δ*Lk*), and *P*(*K*|Δ*Lk*) were calculated from the simulated distributions of *P*(Δ*Lk*|*K*) and of *P*(*K*). The calculated distributions agreed with previous experimental and theoretical results and greatly advanced on them. Our major focus was on *P*(*K*|Δ*Lk*), the distribution of knot types for a particular value of Δ*Lk*, which had not been evaluated previously. We found that unknotted circular DNA is not the most probable state beyond small values of Δ*Lk*. Highly chiral knotted DNA has a lower free energy because it has less torsional deformation. Surprisingly, even at |Δ*Lk*| > 12, only one or two knot types dominate the *P*(*K*|Δ*Lk*) distribution despite the huge number of knots of comparable complexity. A large fraction of the knots found belong to the small family of torus knots. The relationship between supercoiling and knotting *in vivo* is discussed.

Topological properties of DNA are essential for life. It is simplest to consider the topological properties of circular DNA in which both strands are intact, called closed circular DNA, but linear DNA *in vivo* is also topologically constrained (1, 2). The topological state of closed circular DNA can be described by two variables. One is the knot type, *K*, formed by the double helix axis. In particular, a molecule may be unknotted (unknot, trivial knot) or form a non-trivial knot. The second variable, the linking number of the complementary strands, *Lk*, describes the winding of the strands of the double helix about each other. It is more convenient to use the difference between *Lk* and that of relaxed DNA (Lk_o), $\Delta Lk = Lk - Lk_o$, than *Lk* itself. Circular DNA extracted from cells has negative Δ*Lk* (3).

Random cyclization of linear DNA molecules results in an equilibrium distribution of topological states, *P*(Δ*Lk*, *K*). Studies of the components of this distribution have greatly advanced our understanding of DNA conformational properties. The measurement in 1975 of the equilibrium distribution of Δ*Lk* for unknotted circular DNAs, the conditional distribution *P*(Δ*Lk*|Unknot), led to elegant determinations of the free energy of supercoiling (4, 5). These textbook experiments were elaborated later to include the effect of DNA length, solvent, temperature, and ionic conditions (6–10). A theoretical analysis of *P*(Δ*Lk*|Unknot) allowed the determination of the torsional rigidity of DNA (11–15). The conditional distribution for the simplest knot, *P*(Δ*Lk*|Trefoil), has also been studied theoretically (16) and experimentally (17).

The value of *Lk* is not defined in nicked circular DNA, whose topological state is specified by knot type only. The corresponding equilibrium distribution of knots in torsionally unstressed molecules, *P*(*K*), has been the subject of many theoretical studies (16, 18–23). Experimental measurements of *P*(*K*) for different DNA lengths were performed for the first time in 1993 and allowed an accurate determination of the electrostatic repulsion between DNA segments under different ionic conditions (10, 17, 24).

To provide a more complete description of DNA topology, we evaluated the general relationships between *P*(Δ*Lk*, *K*) and the four derivative distributions, *P*(Δ*Lk*|*K*), *P*(*K*|Δ*Lk*), *P*(*K*), and *P*(Δ*Lk*). There is no known method for measuring *P*(Δ*Lk*, *K*) directly, but it can be calculated from a pair of derivative distributions that can be measured or simulated, *P*(*K*) and *P*(Δ*Lk*|*K*). *P*(Δ*Lk*, *K*) can also be simulated directly.

We focused particularly on the conditional distribution *P*(*K*|Δ*Lk*), the equilibrium distribution of knot types in DNA molecules with a particular value of Δ*Lk*. We computed *P*(*K*|Δ*Lk*) in two different ways and obtained the same distributions. These computations showed, in agreement with (25), that beyond very small values of Δ*Lk*, the lowest energy form of DNA for a particular Δ*Lk* is knotted and not plectonemically supercoiled. This preference arises because formation of highly chiral knots minimizes torsional deformation of DNA. Unexpectedly, we found that only a few knots dominated the distribution for a particular Δ*Lk* value and a large fraction of these knots belongs to the small family of torus knots. We discuss the relationship between supercoiling and knot formation inside the cell.

Methods of Calculations

DNA Model. We modeled DNA as a discrete analog of a worm-like chain and accounted for intersegment electrostatic repulsion. A DNA molecule composed of *n* Kuhn statistical lengths is modeled as a closed chain of *kn* rigid cylinders of equal length. Replacement of a continuous worm-like chain with *kn* hinged rigid segments is an approximation that improves as *k* increases. The bending energy of the chain, E_b, is given by

$$E_b = \alpha k_B T \sum_{i=1}^{kn} \theta_i^2, \qquad [1]$$

where the summation extends over all the joints between the elementary segments, θ_i is the angular displacement of segment *i* relative to segment $i - 1$, α is the bending rigidity constant, k_B is the Boltzmann constant, and *T* is the absolute temperature. The value of α is defined so that the Kuhn statistical length corresponds to *k* rigid segments (12). We used $k = 10$, which has been shown to be large enough to obtain accurate results for supercoiled DNA (26). The Kuhn length was set equal to 100 nm (27).

In the simulation of closed circular DNA, we also accounted for the energy of torsional deformation, E_t:

$$E_t = (2\pi^2 C/L)(\Delta Tw)^2, \qquad [2]$$

where *C* is the torsional rigidity constant of DNA, *L* is the length of the DNA chain, and Δ*Tw* is the difference in double helical twist from relaxed DNA (26). The value of Δ*Tw* was not specified in the model directly but was calculated for each

‡To whom reprint requests should be addressed. E-mail: ncozzare@socrates.berkeley.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

conformation using White's equation (28–30), which connects ΔLk and writhe of the DNA axis, Wr, to ΔTw:

$$\Delta Tw = \Delta Lk - Wr. \quad [3]$$

Eq. **3** allows us to use our DNA model to simulate the properties of closed circular DNA with a specified value of ΔLk. The calculation of Wr for a particular conformation was based on Le Bret's algorithm (16).

The excluded volume effect and the electrostatic interactions between DNA segments are taken into account in the model via the concept of effective diameter, d. This is the actual diameter of the impenetrable cylindrical segments of the model chain. We used d = 5 nm throughout this work, which corresponds to a NaCl concentration of 0.2 M (24, 31).

Monte Carlo Simulation Procedure. We used the Metropolis Monte Carlo procedure (32) to generate an equilibrium set of conformations as described in detail elsewhere (33).

Control of Topological Variables. Since the chain segments are allowed to pass through each other during successive deformations in the Metropolis procedure, the knot type of the chain can change. The constructed equilibrium set of chain conformations specifies the equilibrium distributions of knots, $P(K)$. To calculate $P(K)$, one needs only to know the topology of each conformation. In the simulations, this is done by calculating the value of the Alexander polynomial, $\Delta(t)$ at $t = -1$ and $t = -2$ (18). Although the values of $\Delta(-1)$ and $\Delta(-2)$ distinguish all knots obtained in this work, to identify complex knots, we also calculated the more powerful invariant, the Jones polynomial (see ref. 34, for example), using a program written by Jenkins (35).

To calculate $P(\Delta Lk|K)$, we prevented a change of knot type during the simulation by rejecting trial conformations for which the values of $\Delta(-1)$ or $\Delta(-2)$ had changed. We calculated first $P(Wr|K)$, the distribution of Wr for a particular knot type. The torsional and bending deformations of DNA are independent to a good approximation (36). This allowed us to calculate $P(\Delta Lk|K)$ as a convolution of $P(Wr|K)$ and the distribution of the torsional fluctuations, $P(\Delta Tw)$. Namely,

$$P(\Delta Lk|K) = \int P(Wr|K)P(\Delta Tw)dWr$$

$$= \int P(Wr|K)P(\Delta Lk - Wr)dWr. \quad [4]$$

We assumed that $P(\Delta Tw)$ is specified by a Gaussian distribution with variance $\langle(\Delta Tw)^2\rangle$ given by ref. 15:

$$\langle(\Delta Tw)^2\rangle = k_BTL/4\pi^2C. \quad [5]$$

We used a value of $3{\cdot}10^{-19}$ erg·cm (1 erg = 0.1 μJ) for C (10, 15, 27). This way of calculating $P(\Delta Lk|K)$ was first suggested by Benham (37) and was realized in ref. 11.

Results

Definitions of *Lk* and Δ*Lk*, and the Classification of Knots. We consider here the equilibrium probability distributions of the linking number difference, ΔLk, and knot type, K. The value of Lk for two closed contours C_1 and C_2 can be defined as (30, 38):

$$Lk = \frac{1}{4\pi}\oint_{C_1}\oint_{C_2}\frac{[d\mathbf{r}_1 \times d\mathbf{r}_2]\mathbf{r}_{12}}{r_{12}^3}, \quad [6]$$

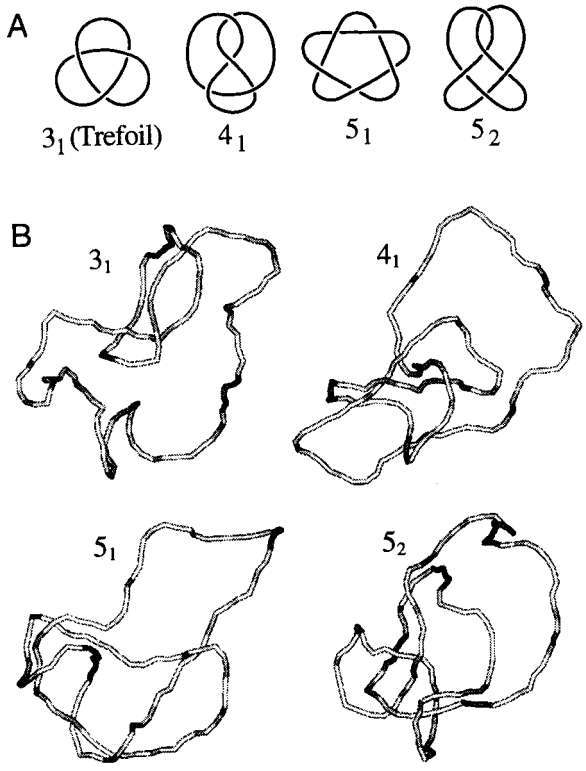

Fig. 1. Simple knots. (*A*) Shown are the standard forms of knots that can be presented with less than six crossings in plane projection. The chiral knots 3_1, 5_1, and 5_2 are represented by only one mirror image. (*B*) Typical simulated knotted conformations of nicked circular DNA, 4 kb in length.

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are vectors that start at a point O and move, upon integration, over C_1 and C_2, respectively; $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$. This definition using the Gauss integral can be applied equally to knotted and unknotted contours. ΔLk can be calculated as

$$\Delta Lk = Lk - N/\gamma, \quad [7]$$

where N is the number of base pairs in the DNA and γ is the number of base pairs per turn of the unstressed double helix. Because the value of N/γ is not integral, it is more convenient to consider ΔLk as a continuous variable even though for any particular DNA its value can differ only in integral amounts. The distribution of discrete values of ΔLk is obtained from the corresponding continuous distribution by simple renormalization. Although most of our calculations were for negative ΔLk, there is no internal chirality in the DNA model used in the simulations, and thus the results can be easily generalized to positive values of ΔLk.

Knots are classified according to the minimum number of intersections in their plane projection. We will refer to such presentation of knots as standard forms. The simplest knot has three intersections in standard form, and there are only four different types of knots with less than six intersections (Fig. 1*A*). As the number of intersections in the standard form increases, the number of knot types grows very fast: there are 1,701,936 knots with less than 17 crossings (39). A knot and its mirror image are rarely topologically equivalent (only the knot 4_1 is equivalent to its mirror image among the four simplest knots shown in Fig. 1*A*), but only one representative of a pair is accounted in the classification (see ref. 34, for example, for more details).

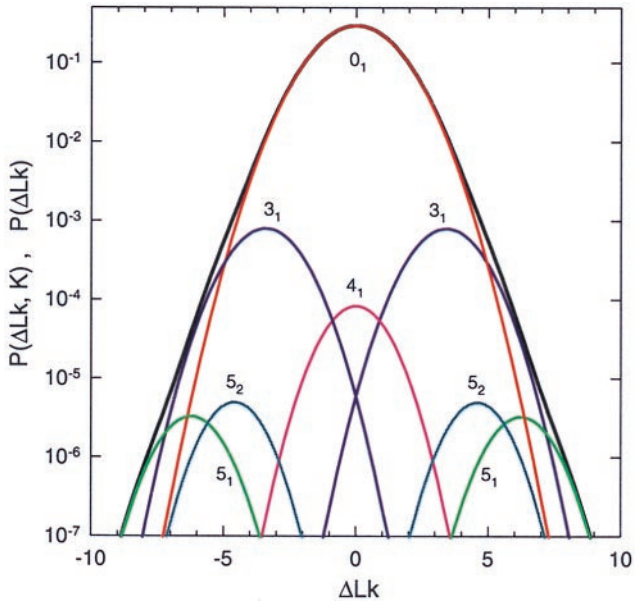

Fig. 2. Simulated distributions of $P(\Delta Lk, K)$ and $P(\Delta Lk)$. The distribution $P(\Delta Lk, K)$ is represented by separate peaks that correspond to different knots: 0_1 (unknotted circles) (red), 3_1 (blue), 4_1 (pink), 5_1 (light green), and 5_2 (turquoise). Contributions from both mirror images of the chiral knots 3_1, 5_1, and 5_2 are shown. Each peak is a Gaussian distribution over ΔLk. The distribution $P(\Delta Lk)$, the sum of $P(\Delta Lk, K)$ over K, is shown by the black line. The simulations were performed for DNA molecules 4 kb in length.

General Relationships Between $P(\Delta Lk, K)$ and the Derivative Distributions. We use in our calculations four general relationships among $P(\Delta Lk, K)$ and the derivative distributions $P(\Delta Lk|K)$, $P(K|\Delta Lk)$, $P(K)$, and $P(\Delta Lk)$. These relationships, valid for any two-dimensional distribution, are:

$$P(\Delta Lk, K) = P(\Delta Lk|K)P(K); \quad [8]$$

$$P(\Delta Lk, K) = P(K|\Delta Lk)P(\Delta Lk); \quad [9]$$

$$P(K) = \sum_{Lk} P(\Delta Lk, K); \quad [10]$$

$$P(\Delta Lk) = \sum_{K} P(\Delta Lk, K). \quad [11]$$

The Distribution of Knots and ΔLk, $P(\Delta Lk, K)$. Fig. 1*B* illustrates typical conformations of the simplest knots obtained in the simulation of DNA molecules 4 kb in length. We calculated the equilibrium distribution, $P(\Delta Lk, K)$, using Eq. **8** and simulated distributions of $P(\Delta Lk|K)$ and $P(K)$ (Fig. 2). Because $P(K)$ decreases sharply as knot complexity grows, we were able to calculate $P(K)$ with reasonable accuracy only for the four simplest knots, 3_1, 4_1, 5_1, and 5_2. Knots

Table 1. Properties of the simplest knots

Knot type, K	c_K	σ_K^2	$P(K)$
0_1	0	1.59	0.995324
3_1	−3.42	1.21	0.002210
4_1	0	1.08	0.000204
5_1	−6.23	1.00	0.000008
5_2	−4.59	0.99	0.000012

The parameters of the distribution $P(\Delta Lk|K)$, specified by Eq. **12**, were computed for DNA molecules 4 kb in length. c_K equals the average value of writhe, $\langle Wr\rangle$, of a torsionally unstressed DNA molecule forming the knot K. For chiral knots, the values of c_K for negative knots are shown.

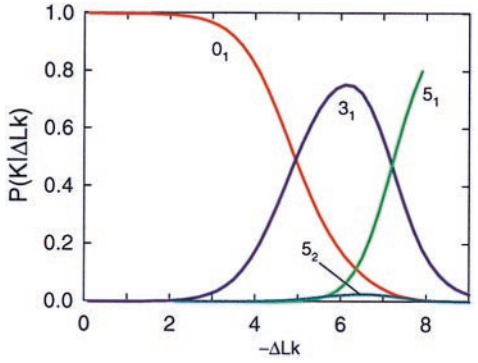

Fig. 3. Simulated distribution of $P(K|\Delta Lk)$ for 4-kb DNA. The distribution was obtained by using Eqs. **9** and **10** and the data from Table 1. Each curve corresponds to a particular knot: 0_1 (red), 3_1 (blue), 5_1 (light green), and 5_2 (turquoise).

3_1, 5_1, and 5_2 are chiral, and therefore both of their mirror images are presented in Fig. 2; knot 4_1 is achiral. The values of $P(K)$ are shown in Table 1. In agreement with experimental data (4, 5, 17), we found that for all these knots $P(\Delta Lk|K)$ is approximated well by the Gaussian distribution:

$$P(\Delta Lk|K) = \frac{1}{\sigma_K\sqrt{2\pi}} \exp[-(\Delta Lk - c_K)^2/2\sigma_K^2]. \quad [12]$$

The values of the distribution variance, σ_K^2, and of c_K are shown in Table 1. The simulation data did not deviate by more than 15% from the best fitted Gaussian curve in the interval $(-4\sigma_K, 4\sigma_K)$. The values of c_K correspond to the average values of writhe, $\langle Wr\rangle$, over the distribution of equilibrium conformations of nicked circular DNA forming a knot K. Our values of c_K are in full agreement with those calculated by Katritch *et al.* (40).

The Distribution $P(K|\Delta Lk)$. Without extra simulation, two other derivative distributions, $P(K|\Delta Lk)$ and $P(\Delta Lk)$, can be obtained from $P(\Delta Lk, K)$ by using Eqs. **9** and **11**. $P(K|\Delta Lk)$, the calculated distribution of knots as a function of ΔLk, is shown in Fig. 3. Although only unknotted circular molecules (0_1) and four knots, 3_1, 4_1, 5_1, and 5_2, were taken into account during the calculation, further results showed (see Fig. 4) that only these knots compete for appearance in the range of ΔLk between 0 and −7.5. It is easy to understand why the knot 4_1 does not appear in this distribution. The average value of Wr for the amphichiral 4_1 equals zero, and thus it competes for appearance with the trivial knot in the range of ΔLk around zero but loses out because $P(\Delta Lk, 0_1) \gg P(\Delta Lk, 4_1)$. It is more interesting that there is a very small amount of the knot 5_2 in Fig. 3. This is because $P(\Delta Lk, 3_1) \gg P(\Delta Lk, 5_2)$, even though the absolute value of $\langle Wr\rangle$ is lower for 3_1 than for 5_2 (see Fig. 2), for all values of $-\Delta Lk$ less than 8. For $-\Delta Lk > 7.2$, knot 5_1 makes a major contribution to the conditional distribution.

It is difficult to obtain $P(K|\Delta Lk)$ for larger values of ΔLk using Eqs. **9** and **11**, because we are not able to calculate $P(K)$ for more complex knots. We can, however, simulate $P(K|\Delta Lk)$ directly for a range of values of ΔLk. The algorithm we use allows us to restrict the equilibrium set of conformations to certain values of K or ΔLk or to remove one or both of these restrictions (see *Methods of Calculations*). To calculate $P(K|\Delta Lk)$, we constructed equilibrium sets of conformations for particular values of ΔLk but allowed any change of knot type during the simulation.

The distributions $P(K|\Delta Lk)$ obtained by direct simulation for two DNA lengths, 2.4 and 4 kb, are shown in Fig. 4. The distribution for 4-kb DNA in Fig. 4*B* is nearly identical with the calculated distribution in Fig. 3 over the same range of ΔLk and for DNA of the same size. This agreement demonstrates the consistency of the computations. A remarkable feature of the distribution is readily

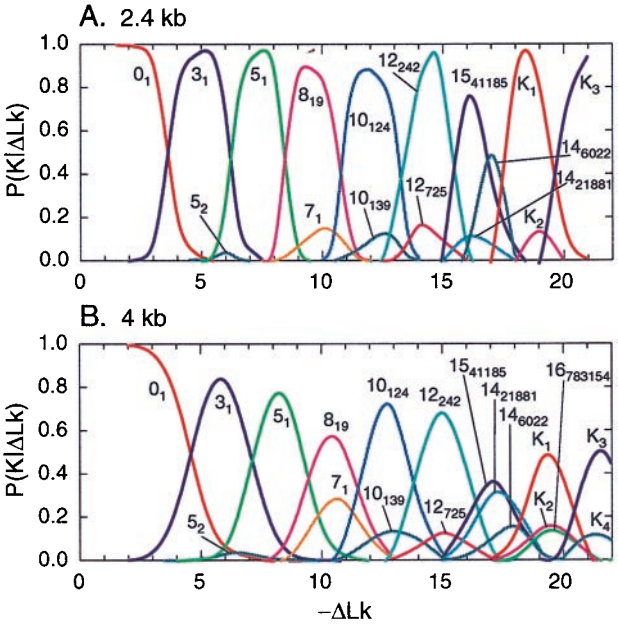

Fig. 4. Simulated distribution of $P(K|\Delta Lk)$ for 2.4-*kb* (*A*) and 4-kb (*B*) DNAs. The data were obtained by direct simulation of this conditional distribution. Each curve corresponds to a particular knot. Curves are shown only for those knots for which $P(K|\Delta Lk)$ exceeds 0.1 (with the exception of knot 5_2). The standard form of these knots is shown in Fig. 5, and some of their features are listed in Table 2.

Table 2. Knots that make the major contributions to $P(K|\Delta Lk)$

Knot type	Alexander polynomial, $\lvert\Delta(-1)\rvert$, $\lvert\Delta(-2)\rvert$	$\langle Wr\rangle$	p, q for torus knots
3_1	3, 7	−3.42	−3,2
5_1	5, 31	−6.23	−2,5
7_1	7, 127	−9.03	−2,7
8_{19}	3, 91	−8.59	−3,4
10_{124}	1, 331	−11.17	−3,5
10_{139}	3, 259	−11.38	
12^n_{242}	1, 1291	−13.57	
12^n_{25}	5, 1147	−13.68	
15^n_{41185}	5, 6355	−15.74	−4,5
14^n_{21881}	1, 5419	−15.93	−3,7
14^n_{6022}	3, 5131	−16.14	
K_1	11, 26611	−18.07	
K_2	19, 29059	−18.24	
16^n_{783154}	3, 21931	−18.39	−3,8
K_3	13, 106483	−20.37	
K_4	21, 115843	−20.50	

Notations for knots 3_1, 5_1, 7_1, 8_{19}, 10_{124}, and 10_{139} are explained in ref. 42; those for 12^n_{242}, 12^n_{725}, 15^n_{41185}, 14^n_{6022}, 14^n_{21881}, and 16^n_{783154} are presented in ref. 39. Knots K_1–K_4 have more crossings than knots in any tables available to us; their structure is shown in Fig. 5. The largest odd numbers that divide $\lvert\Delta(-2)\rvert$ rather than the values of $\lvert\Delta(-2)\rvert$ themselves are shown in the table.

seen in Fig. 4: for any particular value of ΔLk, only very few knots dominate. Indeed, there are 1,701,936 different knots with less than 17 crossings in their standard form (39), but only 12 of them appear in the conditional distributions with probability more than 0.1! Four other knots that appear in Fig. 4 have more than 16 crossings in the standard form and therefore are not among these 1,701,936. Comparison of the distributions calculated for the two DNA lengths shows that the same knots make the major contributions in both cases, although they appear at slightly different values of ΔLk. Also slightly more knots contribute to the distribution for longer DNA. This weak dependence of $P(K|\Delta Lk)$ on DNA length is due to the fact that the average Wr of knotted molecules is nearly length-independent (16, 40). The knots that make the major contribution to $P(K|\Delta Lk)$ are shown in Fig. 5*A* in standard form. Typical conformations for some of these knots are shown in Fig. 5*B*; they are quite similar to the standard presentations and contain barely any extraneous crossings.

Why do so few knots make the major contribution to $P(K|\Delta Lk)$? To address this question, we calculated the average Wr for these knots in the absence of the torsional stress. The results presented in Table 2 show that the values of $|\langle Wr\rangle|$ of the represented knots are very large and exceed the number of crossings in their standard

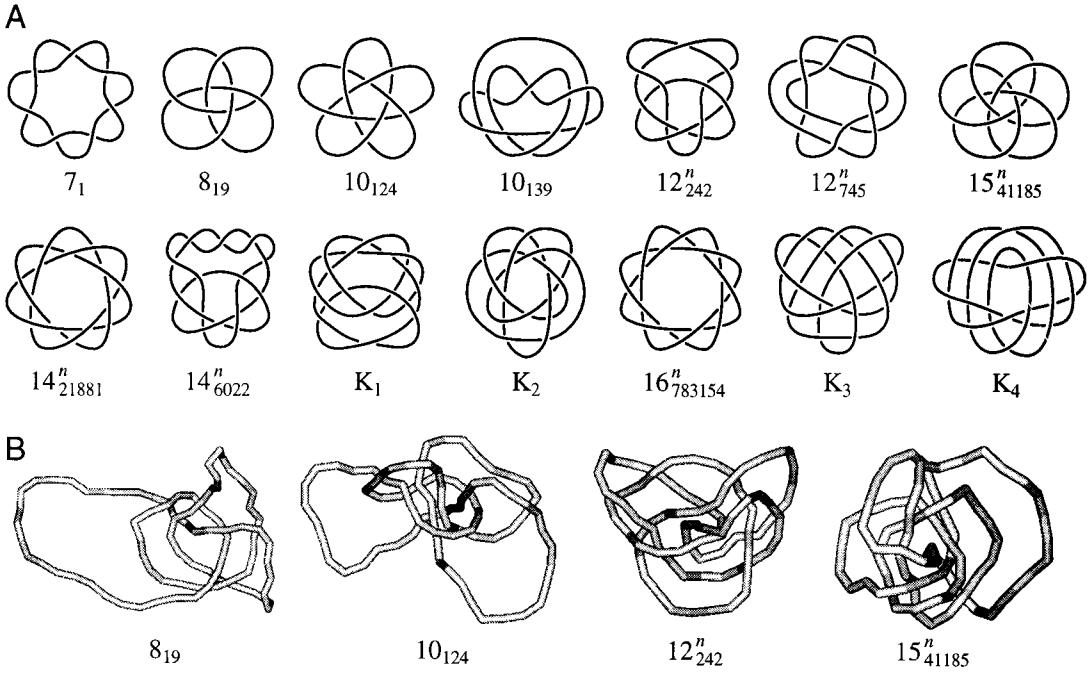

Fig. 5. Knots that make the major contributions to $P(K|\Delta Lk)$. Shown are the standard forms of these knots (*A*) and four examples of simulated conformations of knots (*B*). Knot notations are explained in the legend to Table 2.

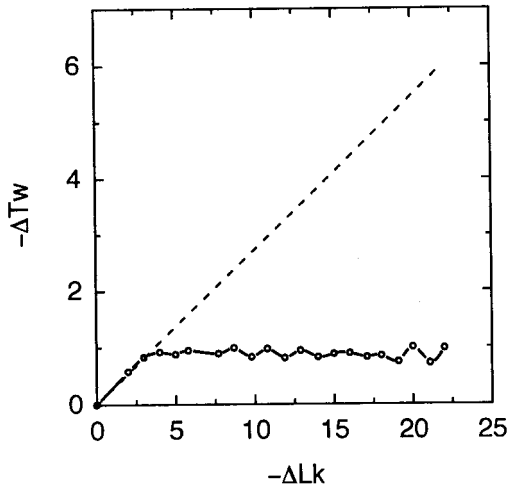

Fig. 6. Dependence of the average torsional deformation, ΔTw, of DNA on ΔLk. Points connected by the solid line are derived from the knot distributions shown in Fig. 4*B*. The dependence of ΔTw on ΔLk for unknotted supercoiled DNA is shown by the dashed line for comparison (41).

form. When $\langle Wr \rangle$ is near ΔLk, the torsional deformation, ΔTw, of the double helix is minimized. This in turn decreases the free energy of torsionally stressed DNA and makes the appearance of the knots thermodynamically favorable. We monitored the average values of ΔTw during the calculation of $P(K|\Delta Lk)$ and found that it does not increase over a wide range of ΔLk (Fig. 6). This is very different from the properties of unknotted supercoiled DNA, for which torsional deformation grows linearly with ΔLk (Fig. 6 and ref. 41). Thus, it seems clear that high chirality, that is a high value of $|\langle Wr \rangle|$ in the torsionally unstressed state, is a necessary condition for the appearance of a knot in the distribution. This high chirality is manifested for the knots shown in Fig. 5 by the fact that all the crossings in standard form have the same sign. High chirality is not sufficient, however. For example, knot 9_1 is not present in the distribution, although the absolute value of its average writhe equals 12.07 (40), and all nine of its crossings have the same sign. There are at least two other chiral knots, 10_{124} (in the notation of ref. 42) and 12^n_{242} (in the notation of ref. 39), for which $P(K|\Delta Lk)$ is larger in the corresponding range of ΔLk (Fig. 4).

Another surprising result is that a large fraction of these knots belong to the family of torus knots (see Table 2). These knots can be drawn without intersection on the surface of a torus and can be specified by two integer variables, p and q, the numbers of intersections of the torus meridian and longitude (34). Torus knots can be readily identified by their Jones polynomial (43), $V(t)$, since, for these knots,

$$V(t)(1 - t^2) = t^{(p-1)(q-1)/2}(1 - t^{p+1} - t^{q+1} + t^{p+q}). \quad [13]$$

We calculated $V(t)$ for the knots appearing in $P(K|\Delta Lk)$ and identified torus knots by trying to fit $V(t)$ to Eq. **13**. The values of p and q for the obtained torus knots are shown in Table 2. The majority of torus knots with less than 17 crossings are in the table. Only the knots with $p = -2$ and $q = 9$, 11, 13, or 15 are absent. Certainly, all torus knots are highly chiral, and this is why they appear so prominently in the distribution. It seems that, in the interval of ΔLk studied, there are few other knots with similar values of $\langle Wr \rangle$ to compete for appearance in $P(K|\Delta Lk)$ with the tiny group of torus knots. It is possible that the other knots presented in $P(K|\Delta Lk)$ can be obtained from corresponding torus knots by only a few stand-passages (44).

Discussion

We considered the equilibrium distribution of ΔLk and K that define the topological state of closed circular DNA. Using computer simulation, we calculated $P(\Delta Lk, K)$ and four derivative distributions: $P(\Delta Lk|K)$, $P(K|\Delta Lk)$, $P(K)$, and $P(\Delta Lk)$. We found that $P(\Delta Lk|K)$ is approximated well by a Gaussian distribution. This fact has a simple explanation. We can express the free energy of a knot K, $G_K(\Delta Lk)$, as the Taylor expansion:

$$G_K(\Delta Lk) = G_K(\Delta Lk_0) + a(\Delta Lk - \Delta Lk_0)^2 + \cdots, \quad [14]$$

where ΔLk_0 is the value of ΔLk when $G_K(\Delta Lk)$ has its minimum value. $P(\Delta Lk|K)$ will have a Gaussian distribution if the terms in Eq. **14** after the quadratic term are small in comparison with the quadratic one. The simulation results show that this is the case. The variance of the distribution diminishes as knot complexity grows (Table 1). For achiral knots, the distribution maximum is at a $\Delta Lk = 0$. For the two mirror image forms of a chiral knot, the distributions of $P(\Delta Lk|K)$ are symmetrical, and their maxima are well separated (Fig. 2). We found that, for DNA molecules a few kilobases in length, $P(K)$ decreases very rapidly with increasing knot complexity. Our results for $P(\Delta Lk|K)$ and $P(K)$ extend the data obtained in earlier studies (15, 16, 24).

The most interesting and most unexpected results were obtained for the conditional distribution $P(K|\Delta Lk)$. The simulations showed that only a few highly chiral knots make a major contribution to the distribution at any particular value of ΔLk. Only 12 of more than 1.7 million knots with fewer then 17 crossings in standard form appear in the distribution with probability larger than 0.1, but for these knots the probability approaches 1 at particular values of ΔLk (Fig. 4). The major feature of these knots is a high value of average Wr when torsionally unstressed. A large fraction of the knots belongs to the torus family, although not all torus knots contribute to the distribution. Increasing the DNA length from 2.4 to 4 kb does not change $P(K|\Delta Lk)$ substantially.

In the simulations, we used only one particular value of DNA torsional rigidity, C, but a wide range of values for C have been reported (8, 15, 25, 45–47). The effect of the torsional rigidity on the distributions studied is rather simple, as long as the torsional and bending deformations in DNA are energetically independent. In this case, $P(\Delta Lk, K)$ can be expressed as a convolution of the twist distribution, $P(\Delta Tw)$ and the distribution of writhe and knot types, $P(Wr, K)$, similar to Eq. **4**. Since $P(Wr, K)$ gives the distribution for nicked circular DNA, it does not depend on C. Thus, the dependence of $P(\Delta Lk, K)$ on C depends only on $P(\Delta Tw)$. $P(\Delta Tw)$ is a Gaussian distribution centered at $\Delta Tw = 0$ that broadens as C decreases (see Eq. **5**). The conclusion of this analysis is that lowering C will not change the knots that appear in $P(K|\Delta Lk)$, but particular knots will be found at higher values of ΔLk (see Fig. 4) and the peaks will be broader. We confirmed this. In a special simulation in which $C = 1{\cdot}10^{-19}$ erg·cm, $P(K|\Delta Lk)$ was shifted by 2 in comparison with the results in Fig. 4*B*, which was calculated by using $C = 3{\cdot}10^{-19}$ erg·cm.

It is important to emphasize that the distributions $P(\Delta Lk|K)$ and $P(K)$ can be measured experimentally and that all other distributions, $P(\Delta Lk, K)$, $P(K|\Delta Lk)$, and $P(\Delta Lk)$, can be calculated from the measured ones. The distribution $P(K)$ can be generated by cyclization of linear DNA molecules via cohesive ends (17, 24). Separation of knots by gel electrophoresis and measurement of their relative amounts allows the evaluation of $P(K)$. The equilibrium distribution of ΔLk for a particular knot $P(\Delta Lk|K)$ can be obtained by ligation of nicks in these knots (17) and measurement of the distribution of ΔLk topoisomers by gel electrophoresis. Recombinases, topoisomerases, and DNA replication can also be used to obtain specific knot types for similar analyses (48).

The distribution $P(\Delta Lk|K)$ has been studied in great detail for unknotted molecules (4–10), and there is very good agreement between these measurements and the results of computer simula-

tion for all DNA lengths and ionic conditions studied (10, 15). $P(\Delta Lk|K)$ was also measured for trefoils (17) and is in good agreement with our simulation results (Table 1). Both the experimental data and simulation indicate that the distribution variance, σ_K^2, for trefoils is 1.3× smaller than for unknotted DNA of the same length. There is also very good agreement between measured and simulated results for $P(K)$ (17, 24). We conclude that the simulations reliably describe the equilibrium distributions of topological variables in circular DNA. The simulated results have the advantages over experimental measurements covering a much greater range of DNA size, ΔLk, and topological complexity, and some of the distributions cannot as yet be measured at all.

It has been observed that, under certain conditions, supercoiling strongly promotes knot formation by type II topoisomerases (49, 50). All these experiments, however, used very high concentrations of enzyme, so that DNA-bound enzyme molecules promote knotting. Does supercoiling promote knot formation at low enzyme concentrations, where DNA conformations are minimally disturbed by enzyme binding? This question is important for understanding knotting inside the cell, and it is interesting to analyze it in terms of the current study.

Let us consider the free energy of circular DNA, which is a function of its topological state, $G(\Delta Lk, K)$. This energy can be calculated as

$$G(\Delta Lk, K) = -RT \ln P(\Delta Lk, K). \quad [15]$$

Since $P(\Delta Lk, K)$ has a global maximum at unknotted DNA with $\Delta Lk = 0$ (this is the most probable state for DNA molecules <100 kb in length), $G(\Delta Lk, K)$ reaches the global minimum for DNA in this topological state. Therefore, a circular DNA must relax to this or close to this state when at equilibrium. Very often, however, topological equilibrium cannot be reached. If DNA is in closed form and no topoisomerases are present, its topology cannot be changed at all.

If a type I topoisomerase is added, ΔLk but not knot type can change. The result will be an equilibrium distribution over ΔLk for each knot present. The molecules relax to the states that correspond to the minimum $G(\Delta Lk, K)$ under the condition that K is constant. This minimum is not, in general, a local minimum of $G(\Delta Lk, K)$. Relaxation is restricted by a particular value of K due to the topological constraint.

In our computation of $P(K|\Delta Lk)$, we restricted topological relaxation in a reciprocal fashion and kept the value of ΔLk constant but allowed K to change. The most probable states found in this computer experiment correspond to the minima of $G(\Delta Lk, K)$, under the condition that ΔLk is constant. Again, these minima are not, in general, the local minima of $G(\Delta Lk, K)$.

What would happen if we added a type II topoisomerase to closed circular DNA? These enzymes can change both topological variables, ΔLk and K, by catalyzing the passages of one double-stranded segment through another. We might expect that these enzymes will yield unknotted molecules with $\Delta Lk \approx 0$, corresponding to the global minimum of $G(\Delta Lk, K)$. The situation is more complex, however. DNA gyrase introduces negative supercoils in circular DNA (51), and other type II topoisomerases untie knots in DNA molecules below equilibrium level (52). This is possible because the strand passage reactions catalyzed by the enzymes are coupled to ATP hydrolysis, which serves as a source of energy. Thus, it is difficult to predict the distribution of topological states of circular DNA in the presence of type II topoisomerases. The presence of type I topoisomerases inside prokaryotic cells, which relax negative supercoils, makes this dynamic picture even more complex.

We know that DNA molecules inside of cells adopt plectonemically supercoiled unknotted conformations [see review (41) and references therein]. This is certainly the desired result for the cell because plectonemic (−) supercoils perform essential work in promoting double-helix opening and DNA compaction. The equilibrium distribution of topological states studied in this work shows that this result is far from obvious, because for DNA with even a modest ΔLk the free energy of highly chiral knots is lower than that of an unknotted plectonemic superhelix. It is possible that changing the amounts and/or activity of topoisomerases or other DNA ligands could result in knotting of circular DNA *in vivo*.

We thank Vaughan Jones for helpful discussions and Christine Hardy for the initial identification of many of the knots. The work was supported by National Institutes of Health Grants GM 54215 to A.V. and GM31657 to N.R.C.

1. Delius, H. & Worcel, A. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 53–58.
2. Kavenoff, R. & Ryder, O. A. (1976) *Chromosoma* **55**, 13–25.
3. Bauer, W. R. (1978) *Annu. Rev. Biophys. Bioeng.* **7**, 287–313.
4. Depew, R. E. & Wang, J. C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4275–4279.
5. Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. & Vosberg, H. P. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4280–4284.
6. Lee, C. H., Mizusawa, H. & Kakefuda, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2838–2842.
7. Shore, D. & Baldwin, R. L. (1983) *J. Mol. Biol.* **170**, 983–1007.
8. Horowitz, D. S. & Wang, J. C. (1984) *J. Mol. Biol.* **173**, 75–91.
9. Duguet, M. (1993) *Nucleic Acids Res.* **21**, 463–468.
10. Rybenkov, V. V., Vologodskii, A. V. & Cozzarelli, N. R. (1997) *Nucleic Acids Res.* **25**, 1412–1418.
11. Vologodskii, A. V., Anshelevich, V. V., Lukashin, A. V. & Frank-Kamenetskii, M. D. (1979) *Nature (London)* **280**, 294–298.
12. Frank-Kamenetskii, M. D., Lukashin, A. V., Anshelevich, V. V. & Vologodskii, A. V. (1985) *J. Biomol. Struct. Dyn.* **2**, 1005–1012.
13. Levene, S. D. & Crothers, D. M. (1986) *J. Mol. Biol.* **189**, 73–83.
14. Shimada, J. & Yamakawa, H. (1988) *Biopolymers* **27**, 657–673.
15. Klenin, K. V., Vologodskii, A. V., Anshelevich, V. V., Klisko, V. Y., Dykhne, A. M. & Frank-Kamenetskii, M. D. (1989) *J. Biomol. Struct. Dyn.* **6**, 707–714.
16. Le Bret, M. (1980) *Biopolymers* **19**, 619–637.
17. Shaw, S. Y. & Wang, J. C. (1993) *Science* **260**, 533–536.
18. Vologodskii, A. V., Lukashin, A. V., Frank-Kamenetskii, M. D. & Anshelevich, V. V. (1974) *Sov. Phys. JETP* **39**, 1059–1063.
19. Frank-Kamenetskii, M. D., Lukashin, A. V. & Vologodskii, M. D. (1975) *Nature (London)* **258**, 398–402.
20. Michels, J. P. J. & Wiegel, F. W. (1986) *Proc. R. Soc. London Ser. A* **403**, 269–284.
21. Des Cloizeaux, J. & Metha, M. L. (1979) *J. Phys.* **40**, 665–670.
22. Koniaris, K. & Muthukumar, M. (1991) *J. Chem. Phys.* **95**, 2873–2881.
23. Deguchi, T. & Tsurusaki, K. (1997) *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **55**, 6245–6248.
24. Rybenkov, V. V., Cozzarelli, N. R. & Vologodskii, A. V. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5307–5311.
25. Vologodskii, A. V. & Marko, J. F. (1997) *Biophys. J.* **73**, 123–132.
26. Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank-Kamenetskii, M. D. & Cozzarelli, N. R. (1992) *J. Mol. Biol.* **227**, 1224–1243.
27. Hagerman, P. J. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265–286.
28. Calugareanu, G. (1961) *Czech. Math. J.* **11**, 588–625.
29. White, J. H. (1969) *Am. J. Math.* **91**, 693–728.
30. Fuller, F. B. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 815–819.
31. Stigter, D. (1977) *Biopolymers* **16**, 1435–1448.
32. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21**, 1087–1092.
33. Klenin, K. V., Vologodskii, A. V., Anshelevich, V. V., Dykhne, A. M. & Frank-Kamenetskii, M. D. (1991) *J. Mol. Biol.* **217**, 413–419.
34. Murasugi, K. (1996) *Knot Theory* (Birkhauser, Boston).
35. Jenkins, R. (1989) Master's thesis (Carnegie Mellon Univ., Pittsburgh).
36. Vologodskii, A. V. & Cozzarelli, N. R. (1993) *J. Mol. Biol.* **232**, 1130–1140.
37. Benham, C. J. (1978) *J. Mol. Biol.* **123**, 361–70.
38. White, J. H. (1989) in *Mathematical Methods for DNA Sequences*, ed. Waterman, M. S. (CRC, Boca Raton, FL), pp. 225–253.
39. Hoste, J., Thistlethwaite, M. & Weeks, J. (1998) *Math. Intell.* **20**, 33–48.
40. Katritch, V., Michoud, D., Scharein, R., Dubochet, J. & Stasiak, A. (1996) *Nature (London)* **384**, 142–145.
41. Vologodskii, A. V. & Cozzarelli, N. R. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609–643.
42. Rolfsen, D. (1976) *Knots and Links* (Publish or Perish, Berkeley, CA).
43. Jones, V. F. R. (1987) *Ann. Math.* **126**, 335–388.
44. Darcy, I. K. & Sumners, D. W. (1998) in *Knot Theory* (Banach Center Publications), Vol. 42, pp. 65–75.
45. Shibata, J. M., Fujimoto, B. S. & Schurr, J. M. (1985) *Biopolymers* **24**, 1909–1930.
46. Shore, D. & Baldwin, R. L. (1983) *J. Mol. Biol.* **170**, 957–981.
47. Moroz, J. D. & Nelson, P. (1998) *Macromolecules* **31**, 6333–6347.
48. Wasserman, S. A. & Cozzarelli, N. R. (1986) *Science* **232**, 951–960.
49. Liu, L. F., Liu, C.-C. & Alberts, B. M. (1980) *Cell* **19**, 697–707.
50. Wasserman, S. A. & Cozzarelli, N. R. (1991) *J. Biol. Chem.* **266**, 20567–20573.
51. Gellert, M., Mizuuchi, K., O'Dea, M. H. & Nash, H. A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3872–3876.
52. Rybenkov, V. V., Ullsperger, C., Vologodskii, A. V. & Cozzarelli, N. R. (1997) *Science* **277**, 690–693.