Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5307–5311, June 1993
Biophysics

Probability of DNA knotting and the effective diameter of the DNA double helix

(DNA topology/DNA excluded volume/polyelectrolytes)

VALENTIN V. RYBENKOV*, NICHOLAS R. COZZARELLI†, AND ALEXANDER V. VOLOGODSKII*

†Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; and *Institute of Molecular Genetics, Kurchatov sq., Moscow 123182, Russia

Contributed by Nicholas R. Cozzarelli, March 5, 1993

ABSTRACT **During the random cyclization of long polymer chains, knots of different types are formed. We investigated experimentally the distribution of knot types produced by random cyclization of phage P4 DNA via its long cohesive ends. The simplest knots (trefoils) predominated, but more complex knots were also detected. The fraction of knots greatly diminished with decreasing solution Na^+ concentration. By comparing these experimental results with computer simulations of knotting probability, we calculated the effective diameter of the DNA double helix. This important excluded-volume parameter is a measure of the electrostatic repulsion between segments of DNA molecules. The calculated effective DNA diameter is a sensitive function of electrolyte concentration and is several times larger than the geometric diameter in solutions of low monovalent cation concentration.**

A classic problem in polymer physics is the probability that the random cyclization of a polymer produces a knot of a particular topology. The issues were formulated more than 30 years ago by Frisch and Wasserman (1) and by Delbrück (2). The initial solution was provided by a Monte Carlo simulation in 1974 (3), and since then several computational analyses of the knotting probability of polymer chains have appeared (reviewed in refs. 4 and 5).

Experimental measures of knotting frequency have not been made, however, despite the considerable interest in these forms. Ever since the production of linked hydrocarbon rings, organic chemists have striven to synthesize knotted molecules. These attempts have recently come to fruition with the elegant and complete synthesis of a hydrocarbon knot by Dietrich-Buchecker and Sauvage (6). A distinct synthetic route based on the properties of nucleic acids has been used by Seeman's group to synthesize different types of knots (7).

Knots made of DNA, first identified in 1976 (8), are much easier to synthesize and to analyze than other knotted polymers. Knotted DNA has been identified in numerous *in vitro* and *in vivo* experiments (9, 10). The topology of the knots is diagnostic of the mechanism of the enzymes that produce them and of the structure of the DNA substrates (9). The experimental studies of DNA knots have been guided by the development of mathematical methods for analyzing DNA topology (11, 12).

DNA is thus the most suitable polymer for the experimental investigation of the probability of knotting. The measurement of the probability of knotting duplex DNA is of special interest because it provides an excellent means for determining the effective diameter of DNA. This important parameter characterizes the excluded volume of DNA. The effective diameter of DNA is the diameter of an uncharged polymer chain that mimics the conformational properties of actual electrically charged DNA. Because DNA is a highly charged polyelectrolyte, electrostatic repulsion can make the effective diameter much greater than the geometrical diameter (13–15). Computer simulations show that the probability of a knot in a randomly cyclized polymer strongly depends on the effective chain diameter. Therefore, we have determined the DNA effective diameter by comparing calculated and experimentally measured values of knotting probability.

Because excluded-volume effects in DNA depend strongly on ionic conditions, we investigated knotting probability over a wide range of conditions. We could not cyclize DNA enzymatically, because ligases are active only under a limited range of conditions. Instead, we cyclized P4 phage DNA via its "cohesive ends," the complementary 19-base single-stranded termini (16). DNA with such a length of cohesive ends is stable in a circular form at room temperature (17). Simulations show that the knotting probability should be in the range of a few percent for this 10-kb molecule (18), which is frequent enough for reliable quantitative measurements. At the same time, this length is short enough for good resolution of knots by agarose gel electrophoresis. We found that the probability of knotting was a sensitive function of electrolyte concentration and was therefore an efficient way to measure the effective diameter of DNA. The values of the effective DNA diameter as a function of Na^+ concentration fit very well the theoretical predictions from polyelectrolyte theory (13).

MATERIALS AND METHODS

DNA. P4vir1del22 phage (19) was grown on *Escherichia coli* strain C-117 (20) and purified as described (21). As additional purification, the phages were treated with DNase I and RNase A, extracted with $CHCl_3$, filtered through a Whatman DE-52 column, and centrifuged through a CsCl step gradient. P4 DNA was then extracted with phenol, phenol/chloroform, and chloroform, precipitated twice with ethanol, and redissolved in 10 mM Tris·HCl, pH 7.6/1 mM EDTA.

DNA Cyclization. To ensure that the P4 DNA was linear at the start of experiments, it was heated at 57°C in 10 mM NaCl for 5 hr and then rapidly (<10 sec) cooled to 0°C. To cyclize the DNA, 0.3 nM P4 DNA molecules were incubated at 24°C for 1–20 days, depending on the ionic conditions. Because the rate of cyclization is strongly dependent on these conditions (22), cyclization was incomplete at low ionic strengths. Further cyclization and knotting during electrophoresis was suppressed by the addition of the oligodeoxynucleotide 5′-GGCGAGGCGGGGAAAGCAC-3′ (1 μM), which is identical to one of the cohesive ends (16). This concentration is at least 300 times higher than the local concentration of the cohesive end from the same molecule (*j* factor).

Electrophoretic Analysis. DNA samples were electrophoresed for 60 hr at 1.2 V/cm at 24°C in a 0.7% agarose gel. The SDS-containing Tris/acetate electrophoresis buffer (23) was

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.

continually recirculated, and DNA was detected by Southern blotting. Open circular DNA molecules of the same size migrate according to the number of minimal nodes in the knot (24). Marker knots were prepared by incubation of ligated supercoiled P4 DNA with phage T2 topoisomerase, followed by nicking with DNase I (25).

Quantitative analysis of the bands corresponding to different knots was done with a Molecular Dynamics 400 PhosphorImager (26). We checked the validity of our measurement procedure by using serial dilutions of *Hin*dIII restriction enzyme-digested phage λ DNA. Over a 100-fold range of dilution, the measured relative amount of all fragments was the same.

Monte Carlo Simulations. The probabilities of different types of knots, P_t, were calculated by a Monte Carlo method (4, 5). We constructed a set of M closed model chains that corresponded to the equilibrium distribution of conformations. The probabilities P_t were estimated as the ratio m_t/M, where m_t was the number of chains with knot type t in the set.

For the simulations, a closed DNA molecule composed of n Kuhn segments was modeled as a closed chain of kn rigid impenetrable cylinders of equal length and diameter, d. The conformational sets were obtained by successive deformations of the starting conformation in accordance with the Metropolis–Monte Carlo procedure (27). The deformation was a rotation of a subchain containing an arbitrary number of adjacent segments by a randomly chosen angle, ϕ, around the straight line connecting the vertices bounding the subchain. The value of ϕ was uniformly distributed over an interval ($-\phi_0$ to ϕ_0), chosen so that about half of the moves were accepted. The trial conformation generated was accepted or not depending on its energy (27). The energy of the chain, E, was calculated as

$$E = RT\alpha \sum_{i=1}^{kn} \theta_i^2, \qquad [1]$$

where summation is done over all the joints between the elementary segments, R is the gas constant, T is the absolute temperature, θ_i is the angular displacement of the ith segment relative to segment $i - 1$, and α is the bending rigidity constant. The bending rigidity constant was chosen so that exactly k elementary segments corresponded to the Kuhn statistical segment length (28). For each set of values of n and d, we generated about 10^6 conformations.

The knot type of a particular chain was determined by calculating the Alexander polynomial $\Delta(t)$ at $t = -1$ and $t = -2$ (4). Whereas the Alexander polynomial distinguishes not all knots (enantiomers, for example), its values at $t = -1$ and $t = -2$ distinguish simple knot types (ones with less than seven nodes) from each other and from more complex knots.

RESULTS

P4 DNA Knotting Probability. The simplest knot, the trefoil (Fig. 1*B*), contains three nodes and is standardly designated 3_1 (29). Random DNA cyclization should produce equal amounts of the two enantiomeric trefoil forms, and we will not distinguish such stereoisomers. There is one type of four-noded knot (4_1), but two types of five-noded knots (Fig. 1*B*), designated 5_1 and 5_2 (29). We detected more complex knots at levels too low to be quantified.

The distribution of knots was analyzed by agarose gel electrophoresis. A typical autoradiograph of a Southern blot is shown in Fig. 1*A*. Knots containing three to five nodes were well separated from each other and from unknotted circular molecules. The identification of the knot types is based on comparison of their electrophoretic mobility to that of reference knots made by the phage T2 topoisomerase (Fig. 1*A*, lane 1), whose topology has been determined by electron microscopy (25). The scan of the gel lane (Fig. 1*C*) shows that it is possible to measure the relative amounts of the three-, four-, and five-noded knots, at least for those formed in high salt concentration, and that the trefoil is the predominant knot species.

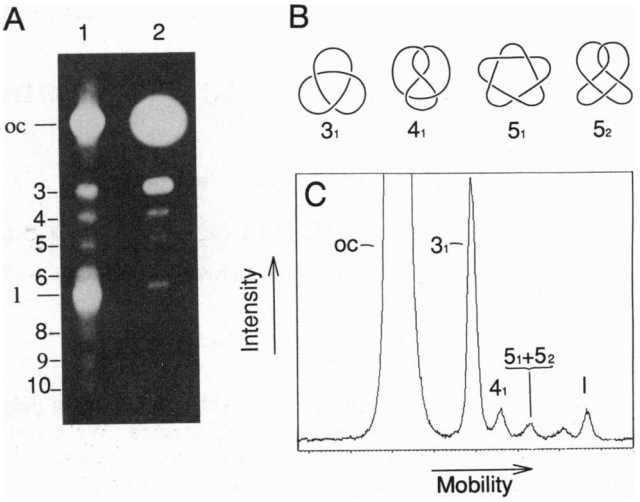

FIG. 1. Electrophoretic separation of DNA knots. P4 DNA knots were resolved by electrophoresis through a 0.7% agarose gel. (*A*) Autoradiogram. Lane 1, reference ladder of 20 ng of phage T2 topoisomerase-generated P4 DNA knots which were nicked prior to loading; lane 2, 20 ng of P4 DNA cyclized in 1 M NaCl. On the left are indicated the minimal node numbers of the knots; oc, open circular; l, linear. (*B*) Minimal representations of simple knots with the standard designations (29). (*C*) Typical scan of sample in lane 2 from *A*.

Fig. 2 shows the probability of a trefoil knot among randomly cyclized P4 DNA as a function of the NaCl concentration during cyclization. The proportion of trefoils sharply increases with increasing salt concentration. The sharp drop in the rate of DNA cyclization with decreasing ionic strength (22) precluded measurements at NaCl concentrations below 0.01 M. In 1 M NaCl, the fraction of three-,

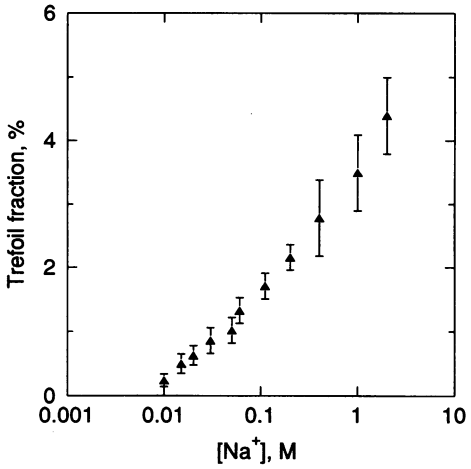

FIG. 2. Fraction of trefoil knots in P4 DNA cyclized in solutions of different NaCl concentration. Each sample contained NaCl and up to 10 mM Tris·HCl (pH 7.5) and 1 mM EDTA (pH 7.5); the amount of buffer and chelator was adjusted to <10% of the total salt concentration. Each point on the graph is the mean of 6–20 determinations. The error was calculated in two ways and the larger value is shown. For samples containing 0.4 or 1.0 M NaCl, the standard deviation in the measured probabilities was the major input to the error. In all other cases, the error bar was determined by the uncertainty in the position of the baseline in the scan of the gel.

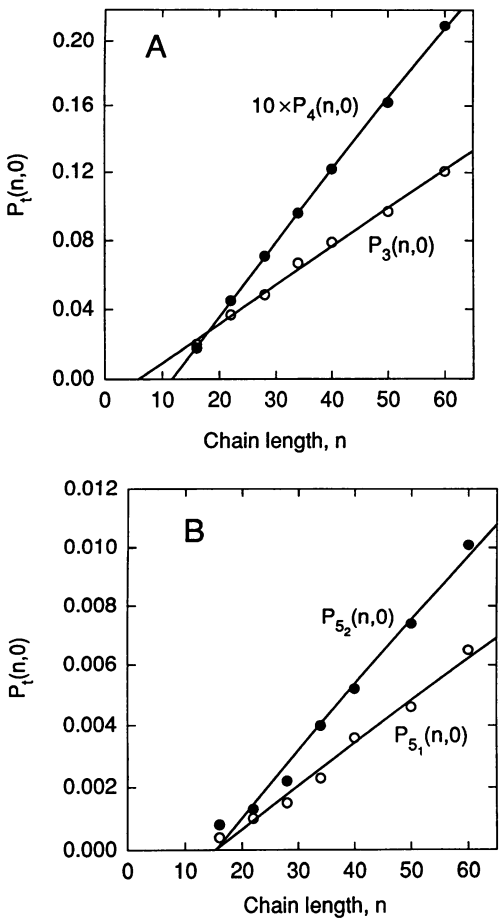

FIG. 3. Monte Carlo calculations of the probability of formation of specific knot types as a function of chain length. $P_t(n, 0)$ is the probability of a knot of type t in a cyclized chain of n Kuhn lengths and zero effective diameter. $P_t(n, 0)$ of 3_1 (○) and 4_1 (●) knots in *A* and 5_1 (○) and 5_2 (●) knots in *B* were computed for closed polymer chains of zero diameter and varying chain lengths. The probability of forming the 4_1 knot, $P_4(n, 0)$, has been multiplied by 10 to be in the same range as the probability of the 3_1 knot, $P_3(n, 0)$.

four-, and five-noded knots was 3.5×10^{-2}, 4.4×10^{-3}, and 2.5×10^{-3}, respectively. Because the amount of four- and five-noded knots was small, we could not measure them reliably at lower NaCl concentrations. Control measurements showed that the knotting probability was independent of the temperature of DNA cyclization in the range from 4°C to 50°C. It also remained the same with a 5-fold decrease or increase in DNA concentration during cyclization.

We investigated the influence of other ions on the probability of forming a trefoil knot. The replacement of Na^+ by other monovalent cations—K^+, Li^+, and Cs^+—had no effect. A more complex pattern was observed for solutions containing $MgCl_2$. In this case, the results depended strongly on the presence of low amounts (1–10 mM) of monovalent ions. Moreover, the fraction of trefoils changed with the temperature of cyclization, which implies that the set of DNA conformations was not at equilibrium.

Monte Carlo Simulations of Knotting Probability. To determine the DNA effective diameter from the data obtained, we simulated random knotting. In contrast to previous work by us and others, we calculated the probability of the individual knot types formed.

We first checked that the calculated knotting probabilities did not depend on the local properties of the model chain. Previous simulations used the simple free-jointed model of a polymer chain, whereas DNA is more aptly modeled by a wormlike chain. We found that the knotting probabilities did not depend on the number, k, of cylindrical segments per Kuhn statistical length when $k = 1$, 5, or 10. Therefore, to a very good approximation, the knotting probabilities have the same values for free-jointed and wormlike models of polymers. The probability of a particular knot type depends only on the number of Kuhn statistical lengths in the chain, n, and the ratio of the chain diameter, d, to the Kuhn statistical length, b.

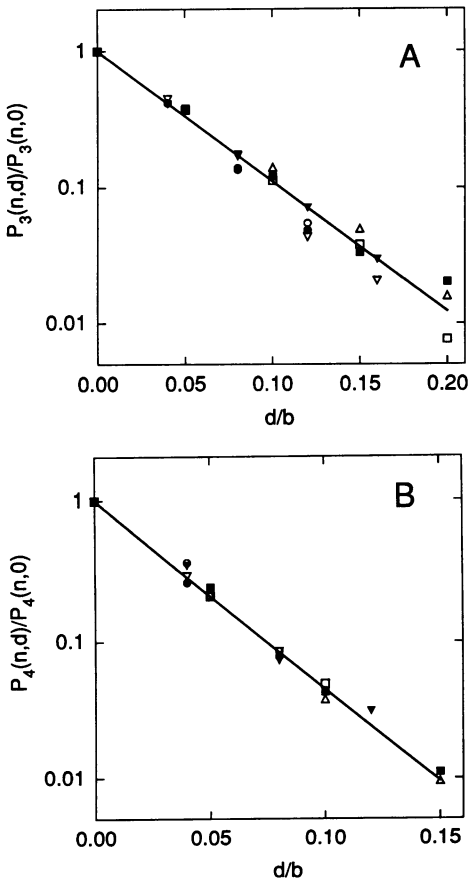

FIG. 4. Monte Carlo calculations of the probability of formation of specific knot types as a function of chain diameter. The probability of knots as a function of chain length and diameter was computed for $n = 16$ (○), 22 (●), 28 (▽), 34 (▼), 40 (□), 50 (■), and 60 (△). The ratio of knotting probability for a given value of d to that at $d = 0$ is plotted as a function of d for 3_1 (in *A*) and 4_1 (in *B*) knots. The straight lines correspond to Eq. 2 with $r = 22$ (*A*) and $r = 31$ (*B*).

We computed the chain length and diameter dependence of the probabilities of formation of knots 3_1, 4_1, 5_1, and 5_2. The results of these calculations for chains with zero diameter are presented in Fig. 3. Beyond a minimal value, knotting probabilities are proportional to n over the range of chain lengths investigated ($16 \leq n \leq 60$). The probability of knotting fell off exponentially with the relative chain diameter, d/b, according to the following empirical equation:

$$P_t(n, d) = P_t(n, 0)\exp(-r_t d/b), \quad [2]$$

where $P_t(n, d)$ is the probability of forming a knot of type t in a chain with diameter d consisting of n Kuhn statistical lengths, r_t is a proportionality constant that was determined by simulations to be equal to 22, 31, 42, and 42 for 3_1, 4_1, 5_1, and 5_2 knots, respectively. Fig. 4 demonstrates that Monte Carlo data for 3_1 and 4_1 knots are well described by this equation. It shows that the ratio of $P_t(n, d)$ to $P_t(n, 0)$ depends on d but not n. The same quantitative agreement was found for knots 5_1 and 5_2. Eq. 2 has been validated for knotting

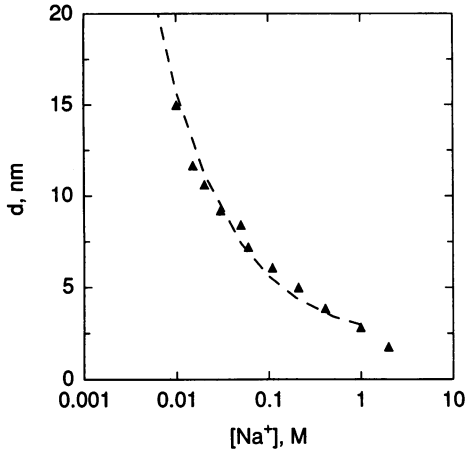

FIG. 5. DNA effective diameter as a function of solution NaCl concentration. The probabilities of trefoil knot formation (Fig. 2) were substituted in Eq. 2, and the value of $P_3(n, 0)$ was found from Fig. 3*A*. The resulting values of d_{eff} (▲) as a function of NaCl concentration are compared with those calculated by Stigter (– – –) on the basis of polyelectrolyte theory (13).

probabilities $\geq 5 \times 10^{-4}$. The probability of knots containing more than five crossings was less than $4 \times 10^{-4} \times n$ and decreased rapidly with increasing values of d/b.

DNA Effective Diameter. The availability of both experimental and numerical results allowed us to calculate the DNA effective diameter as a function of electrolyte concentration. The only additional information we needed was the number of Kuhn statistical segments in P4 DNA. For the ionic conditions in our experiments, the Kuhn statistical length is known with high accuracy to be 100 ± 5 nm (30). Therefore, the length of P4 DNA is 34 Kuhn statistical lengths.

The values of the double-helix effective diameter, d_{eff}, calculated from the probability of trefoil formation are plotted as a function of solution NaCl concentration in Fig. 5. Even in solutions of moderate salt concentration, the value of d_{eff} is several times that of the geometric diameter of a double helix, 2.0 nm. For example, in solutions containing the physiological concentration of NaCl, 0.15 M, d_{eff} is equal to 5 nm.

The data in Fig. 5 are based on the probability of trefoil formation only. Therefore the data on the formation of four- and five-noded knots provided a check on the self-consistency of our results. In 1 M NaCl, d_{eff} equals 2.7 nm (Fig. 5). For this value of d, the simulated probabilities of four- and five-noded knots equal 4.7×10^{-3} and 1.9×10^{-3}, respectively. The measured values are 4.4×10^{-3} and 2.5×10^{-3}. When one considers the experimental error in measurements of such low frequencies of knots (≈30%), the agreement between the theoretical and experimental results is acceptable.

DISCUSSION

We have measured the knot type distribution for randomly cyclized DNA. Comparing these data with a Monte Carlo simulation of knotting probability, we calculated the double-helix effective diameter in different ionic conditions. This analysis is based on the assumption that cyclization produced the equilibrium set of DNA conformations. There are strong arguments favoring this supposition. First, it was shown by Wang and Davidson (17) that diffusion of the cohesive ends of phage λ DNA is not the limiting step in DNA cyclization. This means that the time required for linear DNA to reach conformational equilibrium is, in general, much less than the time for cyclization. Second, the same conclusion can be deduced from the cyclization-temperature independence of the knot distributions. The probability of knotting was the same at 4°C and 50°C even though the rates of cyclization at these two temperatures differed by almost 2 orders of magnitude. There was, therefore, 2 orders of magnitude more time at 4°C to reach conformational equilibrium. Over this temperature interval the diffusion coefficient of DNA changes <3-fold and, thus, cannot compensate for the difference in cyclization time.

Our data for DNA effective diameter in NaCl solutions are in remarkably good agreement with Stigter's predictions based on polyelectrolyte theory (13) shown in Fig. 5. Stigter (13) showed that long-distance interactions were the chief contributions to the value of DNA effective diameter. The striking agreement between his theoretical results and our data shows that polyelectrolyte theory describes well these long-distance interactions between the DNA segments.

Two other estimations of the dependence of d_{eff} on NaCl concentration have been made based on the osmotic pressure of concentrated DNA solutions (14, 15). The agreement is excellent for the results of one study (14) and fair for those of the other (15). An independent measure of the Na^+ dependence of knotting probability by Shaw and Wang (33) is also very similar to ours. Thus, there is convincing evidence that the marked dependence of DNA effective diameter on NaCl concentration is known with accuracy.

The finding that the effective diameter of double-stranded DNA can be several times greater than the geometric one has significant consequences for the conformation of DNA. These consequences are especially marked for supercoiled DNA, where there are close contacts between portions of the DNA chain. Computer simulations show that properties of supercoiled DNA such as average writhe, supercoiling free energy, average superhelix diameter, and, especially, the average number of branch points in the interwound superhelix depend strongly on the DNA effective diameter (31, 32). Any quantitative analysis of supercoiled DNA properties must take into account their dependence on ionic conditions, and our results provide the means of doing so.

This work was supported by National Institutes of Health Grant GM31657, National Science Foundation Grant DMS-8820208, and National Institute of Environmental Health Sciences Center Grant ESO1896 to N.R.C.

1. Frisch, H. L. & Wasserman, E. (1961) *J. Am. Chem. Soc.* **83,** 3789–3795.
2. Delbrück, M. (1962) in *Mathematical Problems in the Biological Sciences*, ed. Bellman, R. E. (Am. Math. Soc., Providence, RI), pp. 55–62.
3. Vologodskii, A. V., Lukashin, A. V., Frank-Kamenetskii, M. D. & Anshelevich, V. V. (1974) *Sov. Phys. JETP (Engl. Transl.)* **39,** 1059–1063.
4. Frank-Kamenetskii, M. D. & Vologodskii, A. V. (1981) *Sov. Phys.-Usp. (Engl. Transl.)* **24,** 679–696.
5. Michels, J. P. J. & Wiegel, F. W. (1986) *Proc. R. Soc. London A* **403,** 269–284.
6. Dietrich-Buchecker, C. O. & Sauvage, J.-P. (1989) *Angew. Chem. Int. Ed. Engl.* **28,** 189–192.
7. Du, S. M. & Seeman, N. C. (1992) *J. Am. Chem. Soc.* **114,** 9652–9655.
8. Liu, L. F., Depew, R. E. & Wang, J. C. (1976) *J. Mol. Biol.* **106,** 439–452.
9. Wasserman, S. A. & Cozzarelli, N. R. (1986) *Science* **232,** 951–960.
10. Shishido, K., Komiyama, N. & Ikawa, S. (1987) *J. Mol. Biol.* **195,** 215–218.
11. White, J. H., Millett, K. C. & Cozzarelli, N. R. (1987) *J. Mol. Biol.* **197,** 585–603.
12. Sumners, D. W. (1990) *Math. Intelligencer* **12,** 71–80.
13. Stigter, D. (1977) *Biopolymers* **16,** 1435–1448.
14. Brian, A. A., Frisch, H. L. & Lerman, L. S. (1981) *Biopolymers* **20,** 1305–1328.
15. Yarmola, E. G., Zarudnaya, M. I. & Lazurkin, Y. S. (1985) *J. Biomol. Struct. Dyn.* **2,** 981–993.

16. Murray, K. & Murray, N. E. (1973) *Nature New Biol.* **243,** 134–139.
17. Wang, J. C. & Davidson, N. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 409–415.
18. Klenin, K. V., Vologodskii, A. V., Anshelevich, V. V., Dykhne, A. M. & Frank-Kamenetskii, M. D. (1988) *J. Biomol. Struct. Dyn.* **5,** 1173–1185.
19. Raimondi, A., Donghi, R., Montaguti, A., Pessina, A. & Deho, G. (1985) *J. Virol.* **54,** 233–235.
20. Bertani, L. E. (1964) *Biochim. Biophys. Acta* **87,** 631–640.
21. Liu, L. F., Davis, J. L. & Calender, R. (1981) *Nucleic Acids Res.* **9,** 3979–3989.
22. Wang, J. C. & Davidson, N. (1966) *J. Mol. Biol.* **15,** 111–123.
23. Sundin, O. & Varshavsky, A. (1980) *Cell* **21,** 103–114.
24. Dröge, P. & Cozzarelli, N. R. (1992) *Methods Enzymol.* **212,** 120–130.
25. Wasserman, S. A. & Cozzarelli, N. R. (1991) *J. Biol. Chem.* **266,** 20567–20573.
26. Johnston, R. F., Pickett, S. C. & Barker, D. L. (1990) *Electrophoresis* **11,** 355–360.
27. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21,** 1087–1092.
28. Frank-Kamenetskii, M. D., Lukashin, A. V., Anshelevich, V. V. & Vologodskii, A. V. (1985) *J. Biomol. Struct. Dyn.* **2,** 1005–1012.
29. Rolfsen, D. (1976) *Knots and Links* (Publish or Perish, Berkeley, CA).
30. Hagerman, P. J. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17,** 265–286.
31. Klenin, K. V., Vologodskii, A. V., Anshelevich, V. V., Dykhne, A. M. & Frank-Kamenetskii, M. D. (1991) *J. Mol. Biol.* **217,** 413–419.
32. Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank-Kamenetskii, M. D. & Cozzarelli, N. R. (1992) *J. Mol. Biol.* **227,** 1224–1243.
33. Shaw, S. Y. & Wang, J. C. (1993) *Science,* **260,** 533–536.