Dynamics of Small Loops in DNA Molecules

Alexei A. Podtelezhnikov and Alexander V. Vologodskii*

Department of Chemistry, New York University, New York, New York 10003

Received October 25, 1999; Revised Manuscript Received January 24, 2000

ABSTRACT: The kinetics and thermodynamics of loop formation by short segments of double-stranded DNA was studied by computer simulation. The DNA molecule was modeled as a discrete wormlike chain. Brownian dynamics was used to simulate the dynamic properties of the chain. Since the average time of loop formation, τ_a, grows sharply when the loop size drops below DNA persistence length, we were unable to simulate the process directly for such small loops. Instead, we used the relationship between the equilibrium probability of loop formation, P, τ_a, and the average time of loop decay, τ_d. The values of P and τ_d were simulated directly. A new Monte Carlo algorithm was developed allowing efficient calculation of P for small DNA loops. The algorithm is also applicable to more complex models of a polymer chain, particularly to DNA models with intrinsic curvature. We also considered loop formation by a segment of a DNA molecule and found that the values of τ_d and τ_a are weakly affected by the total chain size. Our results showed that the formation of small loops is a very slow process: for loops less than 50 nm in size τ_a can be comparable to the lifetime of the cell.

I. Introduction

The formation of loops by DNA molecules is an important step in many key biological processes.[1,2] The thermodynamics of the process have been carefully studied both experimentally and theoretically. Shimada and Yamakawa obtained accurate theoretical results for the probability of loop formation by a wormlike chain, usually used to describe DNA conformational properties.[3] These results are in agreement with the experimental data of Baldwin and co-workers.[4,5] Because it is more difficult to obtain theoretical results for intrinsically curved DNA molecules (however, see ref 6), computational approaches based on Monte Carlo simulations were developed.[7,8]

On the other hand, the kinetics of loop formation has not been studied extensively. It is difficult to investigate the process experimentally, since the rate of cyclization of DNA molecules by joining their cohesive ends is not limited by diffusion. Thus, the cyclization kinetics gives no information about the kinetics of loop formation. Another experimental approach to the problem based on competition between two reactions of site-specific recombination was used recently.[9,10] It seems, however, that in this system the enzymatic kinetics rather than the kinetics of juxtaposition of the specific sites is the rate-limiting stage of the reactions.[10,11] Theoretical approaches gave important results for the rate of loop formation in long polymer chains,[12] but they cannot be applied to small loops since they are based on Gaussian statistics of the chain conformations. In this situation, Brownian dynamics (BD) simulations[13] provide a unique way to study the question. During the past decade, it has been shown that the BD approach is capable of describing the actual times of large-scale DNA motion.[14−17] The method was used recently to study the kinetics of DNA looping for linear DNA molecules[18] and for segment juxtaposition in supercoiled circular DNA.[19] The power of modern computers does not allow, however, the simulation of slow processes in DNA dynamics, like the formation of very small or very large DNA loops. Here we developed a special approach to estimate the rate of loop formation for such cases. We found that the formation of small DNA loops, 50−100 base pairs in length, is an extremely slow process with a characteristic time that can exceed the lifetime of the cell.

II. Methods of Computations

A. DNA Model. Our DNA model combines the features of the wormlike chain and bead−spring models.[14,20] A DNA molecule of the length L was modeled as a chain consisting of N straight segments of equal equilibrium length l_0.

The bending energy, E_b, was specified by the angular displacements θ_i of segment $i + 1$ relative to segment i:

$$E_b = \alpha RT \sum_{i=1}^{N-1} \theta_i^2 \qquad (1)$$

where R is the gas constant and T is the absolute temperature. The bending rigidity constant α is defined so that k straight segments of the model chain correspond to one Kuhn statistical length of DNA. The exact relationship between k and α is described in ref 20; for $k \gg 1$ there is an approximate equation $\alpha \cong k/4$.[21] The replacement of the continuous wormlike chain with a semiflexible chain consisting of hinged rigid segments is an approximation that improves as k increases.

To mimic hydrodynamic properties of DNA, we attached virtual beads to each vertex of the model chain. We chose l_0 so that it equals to the diameter, d, of touching beads, $d = l_0 = 3.18$ nm.[22] Considering that DNA Kuhn statistical length is equal to 100 nm,[23] this choice of l_0 corresponds to $k = 31.45$ ($\alpha = 7.775$). This k guarantees good approximation of the continuous wormlike chain.[24] We used the Rotne−Prager approximation for the tensor of hydrodynamic interaction.[25]

We also introduced the stretching elasticity to the model chain to facilitate the dynamic simulations. The stretching energy, E_s, was computed as

$$E_s = \frac{\beta RT}{l_0^2} \sum_{i=1}^{N} (l_i - l_0)^2 \qquad (2)$$

where l_i is the length of the segment i. The stretching rigidity β was set equal to 50, and this allowed us to

10.1021/ma991781v CCC: $19.00
Published on Web 03/15/2000

choose a sufficiently large time step in BD simulation. Although this value of β is lower than actual stretching rigidity of the double helix, this does not affect dynamic properties of the model chain in the time scale of nanoseconds or larger.[14,26] In the MC simulations the segments retained the constant length l_0.

We did not account for the excluded-volume interaction between segments; the effect is insignificant for the short molecules considered here due to the fact that the probability of collisions of internal segments is extremely low for short chains.

B. Calculation of the Equilibrium Properties. The Metropolis Monte Carlo procedure used in this work is similar to the one used in ref 27. In this procedure, an equilibrium set of chain conformations is obtained by successive displacements from an initial conformation. Two types of displacements are repeatedly applied to randomly chosen subchains consisting of a random number of adjacent segments. The first type of displacement is a crankshaft rotation,[28] when the subchain is rotated by an angle, ϕ, around a line connecting the ends of the subchain. The value of ϕ is uniformly distributed over the range $(-\phi_0, \phi_0)$. This rotation never changes the end-to-end distance of the chain. The second type of displacement is a rotation of the subchain containing one of the chain ends around a random axis by an angle φ. The value of φ is uniformly distributed over the range $(-\varphi_0, \varphi_0)$. The trial rotations are accepted or rejected according to the standard rule based on the energy test.[29] The amplitudes of the rotations, ϕ_0 and φ_0, are automatically adjusted during each simulation run so that approximately half of the trial rotations are accepted. It is easy to see that these displacements satisfy the principle of the microscopic reversibility of the Metropolis procedure, both for unbiased sampling and for sampling with restricted end-to-end distance (see below). Comparison of the simulation results for the j-factor with theoretical calculations[3] shows that the displacements provide efficient sampling in the conformational space.

In this work we needed to calculate the probability of loop formation, $P(r_0)$, which is the probability of the conformations with end-to-end distance, r, less than a small value r_0. This probability is very small for very short DNA molecules, so the looped conformations may not be present even in a very large equilibrium set generated by usual MC procedure. To overcome this problem, we developed a new biased MC method of calculating small probabilities.

We chose a sequence of distances $r_0 < r_1 < ... < r_n$, where r_n is larger than or equal to the chain contour length. For each r_i we can define the probability, $P(r_i)$, of conformations with $r < r_i$ among all possible conformations. We can also define the conditional probability, $P(r_i|r_{i+1})$, of conformations with $r < r_i$ in the subset of conformations with $r < r_{i+1}$. Since $P(r_i) = P(r_i|r_{i+1})P(r_{i+1})$ and $P(r_n) = 1$, the probability of loop formation can be found as

$$P(r_0) = \prod_{i=0}^{n-1} P(r_i|r_{i+1}) \tag{3}$$

The sequence of distances $r_0 < r_1 < ... < r_n$ can be chosen so that all $P(r_i|r_{i+1})$ values are relatively large. This can always be achieved since $P(r_i|r_{i+1})$ approaches 1 when r_{i+1} approaches r_i. The large values of $P(r_i|r_{i+1})$ can be efficiently and accurately calculated using the MC procedure. We estimated that the standard error in the calculation of $P(r_0)$ is minimized when the values of $P(r_i|r_{i+1})$ are close to 0.2.

Each $P(r_i|r_{i+1})$ was calculated as the fraction of the conformations with $r < r_i$ in the subset of equilibrium conformations with $r < r_{i+1}$. These subsets were generated using our MC procedure by rejecting any trial conformation with $r > r_{i+1}$. $P(r_i|r_{i+1})$ values were calculated sequentially from $P(r_0|r_1)$ to $P(r_{n-1}|r_n)$. The starting conformation for each subset was the last conformation from the previous subset. The calculation of $P(r_0|r_1)$ was started from a conformation with $r = 0$. The statistical error was estimated by performing a few independent computations for each chain length.

Using the described biased MC approach, we were able to speed up the computations by many orders of magnitude. For example, it takes 35 min of Pentium-II 266 MHz processor time to calculate the probability of loop formation for $L = 19$ nm and $r_0 = 5$ nm with the standard error of 5.7%. It would require a few years to reach the same accuracy if we used the standard Metropolis procedure. Our approach is similar to the umbrella sampling method,[30] but the selection of the r_i sequence is more intuitive than that of the potential-modifying function in the umbrella sampling procedure.

A similar procedure was used for calculations of j-factor of the model chain, $j(L)$. By definition[31]

$$j(L) = \lim_{\substack{r_0 \to 0 \\ \alpha_0 \to 0}} \left(\frac{1}{N_A} \frac{3P(r_0)}{4\pi {r_0}^3} \frac{2P(\alpha_0)}{1 - \cos \alpha_0} \right) \tag{4}$$

where N_A is Avogadro's number and $P(\alpha_0)$ is the probability that the angle between the end segments of the chain is less than α_0 under condition $r < r_0$. The probability of loop formation, $P(r_0)$, was calculated as described above. To calculate $P(\alpha_0)$, we chose a sequence of angles $\alpha_0 < \alpha_1 < ... < \pi$ and calculated the product of the conditional probabilities $P(\alpha_i|\alpha_{i+1})$. During this procedure, the value of r was less than r_0. Note that to obtain $j(L)$ in mol/L the distances have to be measured in decimeters.

C. Brownian Dynamics Simulations. We used the second-order Brownian dynamics algorithm[32] with some modifications.[26] For the chosen model parameters, the integration time step was equal to $\Delta t = 0.21$ ns.

The time of loop dissociation was calculated directly counting the number of steps before the end-to-end distance of the initially looped chain exceeded r_0. r_0 was equal to either 5 or 10 nm. The equilibrium set of initial conformations with $r < r_0$ was prepared by the Metropolis Monte Carlo procedure. The average time of loop dissociation was calculated over 1000 events.

III. Results

A. Computational Strategy. Thermal motion of a DNA molecule can result in the juxtaposition of its ends. This motion can be accurately simulated by Brownian dynamics (BD).[13] The computations are time-consuming, however, and only rather fast processes can be simulated directly. The rate of loop formation decreases dramatically if the loop size diminishes below one persistence length of the chain (50 nm for double-stranded DNA). We estimated, using BD simulation, that the average time preceding the first juxtaposition, τ_a, exceeds 10 ms for such loops, making it impossible to compute this value directly. To evaluate τ_a for short

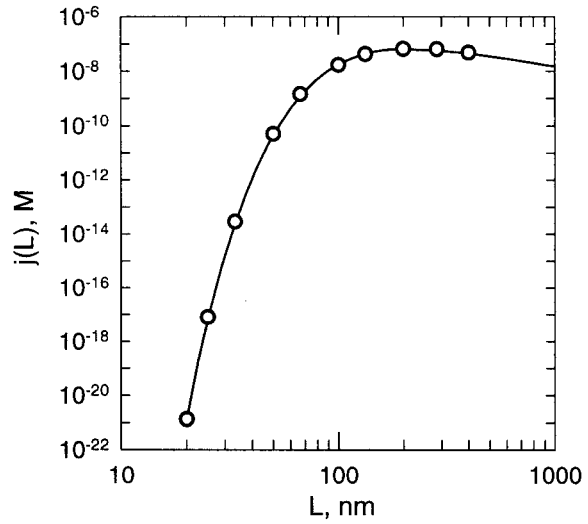

Figure 1. *j*-factor, $j(L)$, for short wormlike chains as a function of their length, L. Theoretical results[3] (line) for $j(L)$, eq 6, are shown together with the results of our computations (○) based on eq 4. It was tested that $r_0 = 1$ nm and $\alpha_0 = 0.1$ used in the computation are sufficiently small to provide a correct estimate for $j(L)$.

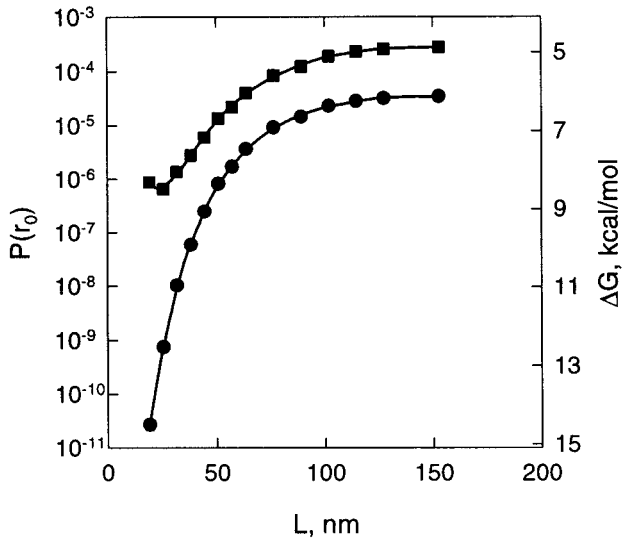

Figure 2. Probability of cyclization, $P(r_0)$, of short DNA molecules as a function of their length, L. The value of $P(r_0)$ is defined as the probability that the distance between the molecule ends is less than r_0. Results of Monte Carlo simulations for r_0 equals 5 nm (●) and 10 nm (■) are shown.

DNA fragments, we used the relation between the equilibrium probability of loop formation, $P(r_0)$, and the average times of association and dissociation of DNA ends, τ_a and τ_d:

$$P(r_0) = \tau_a/(\tau_a + \tau_d) \tag{5}$$

The values in eq 5 are related to a specific distance between molecule ends, r_0: we assume that the loop is formed when the end-to-end distance, r, is less than r_0. Decay of the loops is a fast process, and τ_d can be found directly by BD simulation. Calculating τ_d, we have to start simulations from conformations with $r < r_0$ and take the average over different initial conformations under this condition. The set of initial conformations has to correspond to the equilibrium distribution of DNA conformations under the conditions where the ends are juxtaposed. Thus, we need to simulate both the value of $P(r_0)$ and the conformational distribution under the condition $r < r_0$. The value of $P(r_0)$ is also very small for short fragments, and regular Monte Carlo sampling of the fragment conformations would give very few conformations with $r < r_0$. We developed a new biased Monte Carlo approach which allows fast calculation of $P(r_0)$ as well as the equilibrium sampling of fragment conformations under the condition $r < r_0$ (see section II.B). The method can be useful for other similar problems, such as the time-consuming computations of *j*-factors of intrinsically curved DNA fragments.

It should be noted that eq 5 has to be considered as an approximation only. It would be precise if τ_a and τ_d were equal to the average lifetimes of the corresponding state. The average lifetimes and the average times of the transitions are equal only in special cases such as Poisson's random processes.[33] This is only approximately true for the process of loop formation. We estimated the accuracy of eq 5 for our system by direct BD simulation of loop formation for large values of r_0 when we were able to estimate $P(r_0)$, τ_a, and τ_d directly. The results showed that the values of τ_a obtained directly do not differ from those obtained from eq 5 by more than a factor of 3 (data not shown).

B. Probability of Cyclization. The probability of cyclization is usually specified by the *j*-factor that defines the local concentration of one end of the chain molecule in the vicinity of the other end.[4] To test our MC approach, we compared our simulation data with theoretical results for *j*-factors obtained by Shimada and Yamakawa.[3] These authors found the following equation for *j*-factors of short wormlike chains:

$$j(L) = 4\pi^3 \frac{b^3}{N_A L^6} \exp\left(-\frac{\pi^6 b}{L} + 0.514\frac{L}{b}\right) \tag{6}$$

Here L is the length of the model chain, and b is the Kuhn statistical length which is close to 100 nm for DNA at physiological ionic conditions;[23] N_A is Avogadro's number. Equation 6 as well as eq 4 used in our computations accounts for proper angular orientation of the chain ends. Figure 1 shows our simulation results for $j(L)$ together with the values given by eq 6. We see agreement between the computed and theoretical results over this range of small DNA lengths, while the value of $j(L)$ varies by a few orders of magnitude.

The probability of loop formation $P(r_0)$, as defined in section III.A, is related to the specific distance between the ends, r_0, but is free from the angular restrictions. The calculated dependence of $P(r_0)$ on DNA length is shown in Figure 2 for two values of r_0, 5 and 10 nm. We see that $P(r_0)$ for $r_0 = 5$ nm drop dramatically when the DNA length decreases below one persistence length.

C. Rate of Cyclization. We calculated the average time of loop dissociation, τ_d, by BD simulation (Figure 3). Although the values of τ_d diminish with the reduction of DNA length below 50 nm, the effect is small in comparison with the decrease of $P(r_0)$. The data shown in Figures 2 and 3 and eq 5 allowed us to evaluate the average time before the first juxtaposition of DNA ends, τ_a. Results for the two values of r_0 are shown in Figure 4. We see that τ_a becomes extremely large for DNA molecules shorter than 50 nm and $r_0 = 5$ nm.

All the results shown so far were related to loop formation between the ends of small DNA molecules. It is more important, from a biological point of view, to

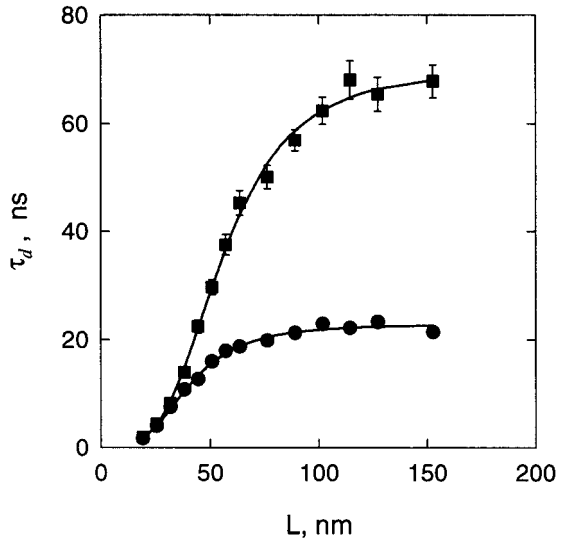

Figure 3. Average time of dissociation of DNA loops. A loop was considered to be formed if the distance between the molecule ends, r, was less than r_0. Equilibrium sets of initial conformations with $r < r_0$ was prepared by MC simulation. The sets were used in BD simulations of loop dissociation for r_0 equals 5 (●) and 10 nm (■). Error bars indicate 1 standard deviation.

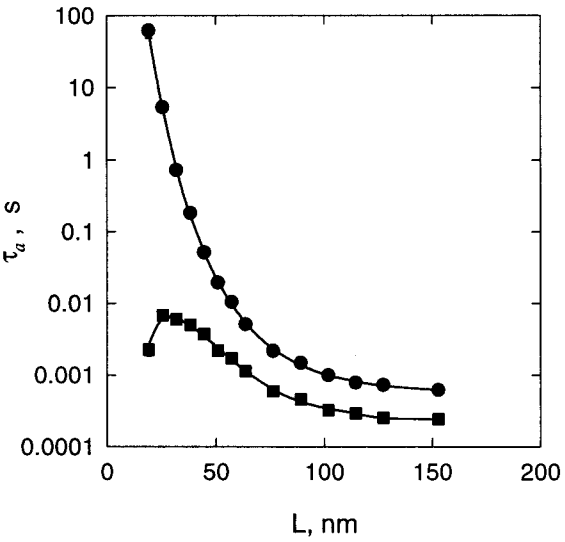

Figure 4. Average time of loop formation by DNA molecules. The results were obtained by using eq 5 and the data of Figures 2 and 3. The data correspond to the values of $r_0 = 5$ nm (●) and $r_0 = 10$ nm (■).

estimate τ_a for internal loops in large DNA molecules. Again, we can consider the issue in terms of eq 5. The excluded-volume effect associated with the interaction of the loop and the rest of DNA molecules is the only factor that can change $P(r_0)$, but for small internal loops this effect is small. Using the BD simulation, we estimated τ_d for loops that are formed by internal segments of larger molecules. We found that τ_d for the internal loops is larger than τ_d for the loops formed between the ends of small molecules by no more than a factor of 2 (data not shown). Thus, the increase in τ_a for internal loops will also be close to a factor of 2, which is not essential for the current analysis.

IV. Discussion

We evaluated the average time of loop formation, τ_a, for short double-helical DNA molecules. Our results show that the times of loop formation by DNAs shorter than 50 nm can approach the cell lifetime. This means that formation of protein bridges between the ends of the small loops requires additional assistance. Binding by another protein, which bends the double helix at the apex of the future loop, can greatly speed up the process. Our results also show that τ_a depends strongly on the distance between the DNA ends in the loop. Therefore, increasing the length of the bridges may serve as an alternative solution for loops smaller than 50 nm.

Similar BD simulations of the DNA looping were performed recently by Langowski and co-workers.[18] They used a direct approach to evaluate τ_a which does not allow one to estimate the value for the very short loops considered here. Their results for loops of 80 and 160 nm in length are in reasonable agreement with our calculation, although accurate comparison is difficult because of the approximate character of eq 5.

The very low probability of loop formation, $P(r_0)$, for short DNA fragments means that the process is associated with a large positive free energy, ΔG. We can estimate ΔG as

$$\Delta G = -RT \ln P(r_0) \qquad (7)$$

where R is the gas constant and T is the absolute temperature. If we substitute the value of $P(r_0)$ for $L =$ 50 nm and $r_0 = 5$ nm, the value of ΔG will be close to 8 kcal/mol and will increase sharply if r_0 is reduced. It is interesting that the values of ΔG obtained from eq 7 are close to the elastic energy of bending for the corresponding DNA conformations.[34] This elastic energy was often associated with the free energy of small loops (see ref 34, for example). However, the value of $P(r_0)$ depends strongly on r_0 (see Figure 2), whereas the energy of elastic deformation is nearly independent of it if r_0 is sufficiently small. Thus, under no circumstances can one neglect the entropic part of the looping free energy.

Acknowledgment. We thank Dr. S. M. Mirkin for helpful discussions. This work was supported by NIH Grant GM54215 to A.V.V. We thank Dr. S. Sarma for the editorial assistance.

References and Notes

(1) Rippe, K.; von Hippel, P. H.; Langowski, J. *Trends Biochem. Sci.* **1995**, *20*, 500.
(2) Kanaar, R.; Cozzarelli, N. R. *Curr. Opion. Struct. Biol.* **1992**, *2*, 369.
(3) Shimada, J.; Yamakawa, H. *Macromolecules* **1984**, *17*, 689.
(4) Shore, D.; Langowski, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4833.
(5) Shore, D.; Baldwin, R. L. *J. Mol. Biol.* **1983**, *170*, 957.
(6) Livshits, M. A. *Mol. Biol.* **1996**, *30*, 85.
(7) Levene, S. D.; Crothers, D. M. *J. Mol. Biol.* **1986**, *189*, 61.
(8) Hagerman, P. J.; Ramadevi, V. A. *J. Mol. Biol.* **1990**, *212*, 351.
(9) Parker, C. N.; Halford, S. E. *Cell* **1991**, *66*, 781.
(10) Oram, M.; Marko, J. F.; Halford, S. E. *J. Mol. Biol.* **1997**, *270*, 396.
(11) Sessions, R. B.; Oram, M.; Szczelkun, M. D.; Halford, S. E. *J. Mol. Biol.* **1997**, *270*, 413.
(12) Friedman, B.; O'Shaughnessy, B. *Int. J. Mod. Phys. B* **1994**, *8*, 2555.
(13) Ermak, D. L.; McCammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352.
(14) Allison, S. A. *Macromolecules* **1986**, *19*, 118.
(15) Allison, S.; Austin, R.; Hogan, M. *J. Chem. Phys.* **1989**, *90*, 3843.
(16) Allison, S. A.; Sorlie, S. S.; Pecora, R. *Macromolecules* **1990**, *23*, 1110.

(17) Chirico, G.; Langowski, J. *Macromolecules* **1992**, *25*, 769.
(18) Merlitz, H.; Rippe, K.; Klenin, K. V.; Langowski, J. *Biophys. J.* **1998**, *74*, 773.
(19) Jian, H.; Schlick, T.; Vologodskii, A. *J. Mol. Biol.* **1998**, *284*, 287.
(20) Frank-Kamenetskii, M. D.; Lukashin, A. V.; Anshelevich, V. V.; Vologodskii, A. V. *J. Biomol. Struct. Dyn.* **1985**, *2*, 1005.
(21) Landau, L.; Lifshitz, E. *Statistical Physics*; Pergamon Press: London, 1958; pp 478−482.
(22) Hagerman, P. J.; Zimm, B. H. *Biopolymers* **1981**, *20*, 1481.
(23) Hagerman, P. J. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 265.
(24) Vologodskii, A. V.; Levene, S. D.; Klenin, K. V.; Frank-Kamenetskii, M. D.; Cozzarelli, N. R. *J. Mol. Biol.* **1992**, *227*, 1224.
(25) Rotne, J.; Prager, S. *J. Chem. Phys.* **1969**, *50*, 4831.
(26) Jian, H.; Vologodskii, A.; Schlick, T. *J. Comput. Phys.* **1997**, *73*, 123.
(27) Vologodskii, A. V. *Macromolecules* **1994**, *27*, 5623.
(28) Klenin, K. V.; Vologodskii, A. V.; Anshelevich, V. V.; Dykhne, A. M.; Frank-Kamenetskii, M. D. *J. Mol. Biol.* **1991**, *217*, 413.
(29) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(30) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, UK, 1987.
(31) Crothers, D. M.; Drak, J.; Kahn, J. D.; Levene, S. D. *Methods Enzymol.* **1992**, *212*, 3.
(32) Iniesta, A.; Garcia de la Torre, J. *J. Chem. Phys.* **1990**, *92*, 2015.
(33) Korn, G. A.; Korn, T. M. *Mathematical Handbook for Scientists and Engineers*; McGraw-Hill: New York, 1961.
(34) Hochschild, A. *Protein−Protein Interactions and DNA Loop Formation*; Cozzarelli, N. R., Wang, J. C., Eds.; Cold Spring Habor Laboratory Press: Cold Spring Harbor, NY, 1990; pp 107−138.

MA991781V