

## Approach to Monte Carlo calculation of the buckling of supercoiled DNA loops

Yang Zhang

*Institute of Theoretical Physics, Academia Sinica, P.O. Box 2735, Beijing 100080, China*

(Received 31 January 2000)

The short supercoiled circular DNA molecules are shown to be glassy systems and canonical Metropolis Monte Carlo simulations of the systems tend to get stuck in local metastable energy basins. A Monte Carlo algorithm is developed to alleviate the problem of ‘‘ergodicity breaking’’ of the glassy systems, in which the Markov process is driven by an explicitly analytic weight factor with enhanced probability in both low- and high-energy regions. To characterize the degree of puckering of the supercoiled DNA loops, a different quantity of aplanarity is introduced as the shortest principal axis of configurational ellipsoid of DNA. With the suggested Monte Carlo method, the quantitative correlation between supercoiling degree and buckling of DNA is attained. With supercoiling stress increasing, the conformational transition from a circle to mono-, diplo-, or triple interwound superhelical structure will take place in a successive but decreasingly abrupt mode.

PACS number(s): 87.15.Aa, 36.20.Ey, 87.14.Gg

The onset of supercoiling and supercoiling transitions of closed circular deoxyribonucleic acid (DNA) are intriguing phenomena from both a theoretical and a biological viewpoint. It is generally believed that proper supercoiling strain is crucial for the DNA molecule to pack and fold into chromosomes or to interact with specific protein [1]. In the simplest mechanical description, in which one neglects all molecular inhomogeneities, a double-helical DNA molecule can be viewed as an elastic rod, the elastic energy of which can be decomposed into bending and torsional parts [2]. These two energy terms are usually decoupled since the bending energy depends on local curvature of the center axis, while torsional energy is the function of displacements of twist of the DNA base pairs from the equilibrium positions. For a torsionally relaxed small loop, (e.g., nicked DNA polymer with one broken strand), the torsional energy of DNA vanishes, and DNA molecule has the minimum bending energy of a flat circle. The effective coupling between bending and torsional energies rises for covalently closed DNA polymers, since the number of times two strands of the DNA duplex are interwound, i.e., the linking number  $Lk$ , is of topological invariance. The planar circle would no longer stand as a stable minimum of the total elastic energy when the link number deficit or excess is beyond some threshold which is a function of bending and torsional stiffness of DNA polymer [3]. In fact, the spatial configuration of circular DNA depends upon the competition between bending and twist energies. Planar DNA microcircles will pucker if the decrease in torsional energy through changing the twist of the base pairs exceeds the increase of bending energy caused by writhing of the polymer axis.

The buckling of short supercoiled DNA loops of 1000 base pairs ( $bp$ ) was computationally studied by Schlick *et al.* with deterministic techniques of energy minimization and molecular dynamics (MD) [4]. A catastrophic buckling transition from the circle to the confirmation of figure-8 was presented with increasing supercoiling strain of DNA, coinciding with earlier analytical calculations [3]. The Metropolis Monte Carlo (MC) approach to this buckling transition of DNA circle of 468 $bp$  was performed by Gebe and Schurr, through tracing the writhing of the molecule [5].

In this Rapid Communication, we study the buckling dynamics of DNA loops of 168 $bp$  through a different defined puckering degree (see below), by Monte Carlo simulation of a discrete wormlike chain model [6]. The deformation energy of the DNA rod is then approximated by the harmonic bending and twisting components. In each update, an interval subchain containing arbitrary amount of links is rotated around the straight line connecting the vertices bounding the subchain. The rotating angle is randomly taken in a interval chosen so that about half of the proposed moves are accepted. The excluded-volume effects and electrostatic repulsive interaction are incorporated by a hard-cylinder potential with an effective diameter ( $d > 2$  nm) in the way the free energy of DNA will be infinite when the distance of any nonadjacent parts of DNA is less than  $d$ . However, as shown in Fig. 1(a), the conventional Metropolis simulation [7] of short DNA polymer of 168 $bp$  with Boltzmann weight factor tends to get stuck in some configurations with local metastable minimum energies.

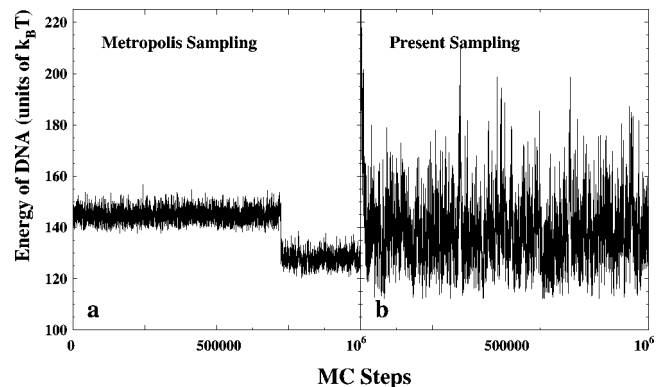


FIG. 1. The energy of a DNA loop of 168 base pairs with linking number deficit  $\Delta Lk = 5$ , as a function of computer time: (a) for the canonical Metropolis method, the simulation is getting stuck in a local energy basin at the first 700 000 steps and in another local energy minimum afterwards; (b), for the present sampling with new weight factor of Eq. (1), much larger fluctuations are implemented, which keeps the simulation from getting trapped in local energy basins.

In fact, the energy landscape of supercoiled DNA chain is characterized by numerous local minima separated by energy barriers. At length scales comparable to the double-helix repeat of 3.4 nm ( $\approx 10.5$  base pairs) or the diameter of 2.0 nm [9], the pairing and stacking enthalpy of the bases make the polymer remarkably rigid and the energy barriers significantly high compared with Boltzmann weight factor. Since the probability of a canonical Metropolis procedure to cross the energy barrier of height  $\Delta E$  is proportional to  $\exp(-\Delta E/k_B T)$ , the simulations therefore tend to get trapped in some local energy basins [Fig. 1(a)], although Metropolis Monte Carlo sampling is usually more effective than molecule dynamics simulation for conformational changes which jumps to different area of phase space [8]. For the glassy systems such as short DNA loops, different MC and MD simulations within finite CPU time and sweeps may get stuck in different energy basins, which renders the calculations of physical quantities unreliable. One of the main aims of present work is at alleviating this problem of ‘ergodicity breaking (EB)’ in Monte Carlo simulations of glassy systems.

Technically, the above-mentioned EB problem is due to the fact that the Boltzmann weight factor drops too quickly (exponentially) with system energy  $E$ . Therefore, the probability of the Markov process to jump over the high energy barriers is exponentially damped. In the following, we consider a new weight factor

$$w(E) = e^{-E + (\sqrt{2}/\sigma)|E - \langle E \rangle|}, \quad (1)$$

where  $\langle E \rangle$  is the averaged energy of system,  $\sigma^2 (= n_F/2)$  the mean squared deviation of energy of the canonical thermodynamic system, and  $n_F$  the number of degrees of freedom. For the discrete closed wormlike chain with  $N$  links, the number of degrees of freedom is  $n_F = 2N - 6$ . Here and after the energy and rigidity parameters are scaled by  $k_B T$ , so the Boltzmann inverse temperature parameter  $\beta = 1/k_B T$  is omitted in our equations.

With the introduction of the factor of  $\exp[(\sqrt{2}/\sigma)|E - \langle E \rangle|]$ , the sharp peak of energy distribution of the canonical ensemble is damped and the important sampling in the low-energy region is reinforced. Especially within the new weight factor of Eq. (1), the probability in the higher-energy region is enhanced exponentially, which makes the simulation escape from local energy basins easily. In Fig. 1(b) we present the MC time series of energy of the same DNA system but with the new weight factor of Eq. (1). Indeed, the simulation of the new artificial ensemble covers a much wider energy range than that of the canonical run, which efficiently keeps the simulation from getting stuck in local minima.

In the artificial ensemble of Eq. (1), each configuration of energy  $E$  of DNA loop, in fact, represents  $n(E)$  ( $= \exp[-(\sqrt{2}/\sigma)|E - \langle E \rangle|]$ ) ones of the real thermodynamic system. Through reweighting the artificial sample [10], i.e.,

$$\langle A \rangle = \frac{\sum_{i=1}^{N_{\text{sweep}}} A(E_i) e^{-(\sqrt{2}/\sigma)|E_i - \langle E \rangle|}}{\sum_{i=1}^{N_{\text{sweep}}} e^{-(\sqrt{2}/\sigma)|E_i - \langle E \rangle|}}, \quad (2)$$

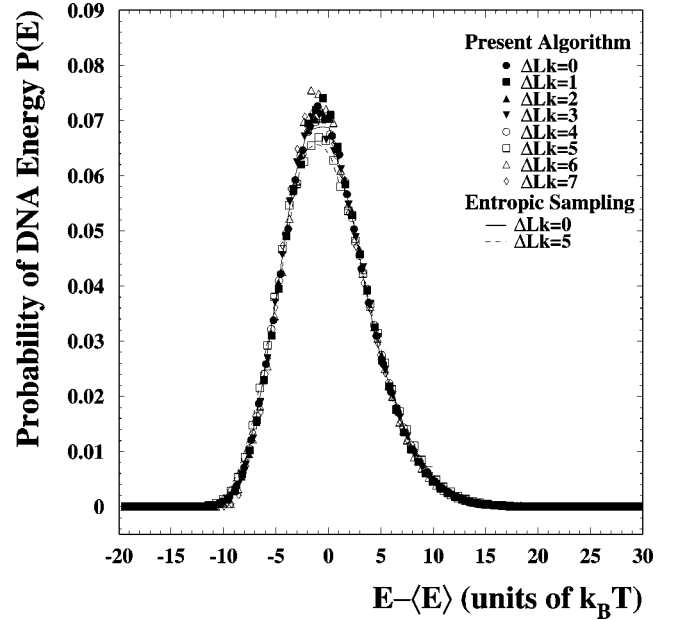


FIG. 2. The probability distributions of energies of DNA loops, which is calculated by reweighting the artificial ensemble with Eq. (2). After translating the energy to the averaged value  $\langle E \rangle$ , the distributions of different linking number deficit  $\Delta Lk$  fall on the same curve. The results of the simulation by entropic sampling method [11–14] coincide with that by present approach, but its CPU time is twice more than the latter.

where  $N_{\text{sweep}}$  is the number of MC sweeps of the artificial sample, we can obtain the expectation value or probability distribution of the considered quantity  $A$  in the real physical system. As an illustrated example, we show in Fig. 2 the probability distributions of DNA energies, which are calculated from the artificial sample of 2 000 000 MC runs. As expected, after the translation of the DNA energies to their averaged values  $\langle E \rangle$ , all the distributions with different supercoiling strains precisely fall onto the same curve. In fact, this scaling behavior of the energy distribution of canonical ensemble is the foundation that we could use the same compensative factor in Eq. (1) to alleviate the EB problem in different energy systems.

It should be mentioned that the EB difficulty has been long standing in the Monte Carlo simulations of glassy systems. The most often-used techniques of dealing with this problem in first-order phase transitions [11], protein folding problem [12], and other glassy systems [13], has been the entropic sampling method [14,15], in which the spectral density of the considered system is calculated numerically by an iterative procedure so that the simulation can be performed as a random walk in a desired range of phase space. Depending on the rate of convergence of iteration and the size of simulated energy range, the determination of the weight factor is usually tedious and very time-consuming [11–14]. To get an approximate flat distribution of the energy spectral in the simulation of our short DNA loops, for example, we need to spend more than 50% of the whole CPU time to achieve this task. In Fig. 2 we also show the probability distribution of DNA energies of the canonical ensemble obtained from the entropic sampling technique for  $\Delta Lk = 0$  and 5, respectively. The results coincide with that by the present method.

The only parameter to be determined in Eq. (1) is the

TABLE I. The averaged energies of 168-basepair-DNA loops for different supercoiling restrains, which are obtained by three iterations as demonstrated in the text.

$\Delta Lk$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.5	3.0	3.5	4.0
$\langle E \rangle$ (in $k_B T$ )	35.09	36.03	39.11	44.15	51.11	60.00	69.78	73.80	78.84	85.10	91.58	107.84	127.47	150.38	177.78

averaged energy of the system. In principle, such estimates can be found in a simple iterative way [16]. One first sets an initial guess of the averaged energy and performs a simulation with small number of Monte Carlo sweeps, and gets a new value of the averaged energy which is a better estimator for the real energy. One can run the simulation using the new estimator of  $\langle E \rangle$  to get a newer one and iterate this process until  $\langle E \rangle \approx \sqrt{\langle E^2 \rangle - n_F/2}$ . According to our calculations, the averaged energy converges to the stable value very quickly, and just a few times of iterations are enough to get a precise estimator of the averaged energy. The averaged energies of the DNA loops of some different supercoiling restrains, which are obtained by three iterations, are listed in Table I. The estimation of averaged energy can be further facilitated by information of its ground state value  $E_0$ . For example, the ground state of the DNA loop for  $\Delta Lk=0$  corresponds to the flat circle, whose energy is  $18.14k_B T$  in our system. So we can directly calculate the averaged energy by  $\langle E \rangle = E_0 + n_F/2 = 35.14k_B T$ .

In previous literatures, e.g., Refs. [5,6], the writhing number ( $Wr$ ) was often used to characterize the tertiary structure and handedness of circular DNA, which is defined as the difference between the linking number  $Lk$  of two strands and the twisting number  $Tw$  of base pairs [17]. However,  $Wr$  has some features which obviously hinder as the most proper definition of the puckering degree of DNA loops. For example, for a circular DNA wrapping around a sphere, it may be considerably puckered, however,  $Wr$  definitely equals to zero; when DNA takes a small displacement of passage through itself, the puckering degree and aplanarity should stay almost unchanged; however,  $Wr$  discontinuously jumps.

To give a proper definition of the degree of puckering of DNA loops, let us first define a  $3 \times 3$  symmetric coordinate tensor

$$T_{ab} = \frac{3 \int [r_a(s) - r_{0a}][r_b(s) - r_{0b}] ds}{\int [r(s) - r_0]^2 ds}, \quad a, b = 1, 2, 3, \quad (3)$$

where  $\mathbf{r}(s)$  denotes the axis vector of the DNA polymer along with its arc-length  $s$ ,  $\mathbf{r}_0 = (1/L) \int \mathbf{r}(s) ds$  the center of mass of the DNA polymer with total arc-length of  $L$ ,  $r_{a(b)}(s)$  the projection of the axis vector on the  $a(b)$  axis of a 3D Cartesian coordinate system. The positive definite tensor  $T_{ab}$  has three eigenvalues  $T_i$ 's with  $0 \leq T_1 \leq T_2 \leq T_3$  and  $T_1 + T_2 + T_3 = 3$ . In fact,  $T_{ab}$  represents the configurational ellipsoid of DNA polymer and  $T_i$  its  $i$ th major axis. So the smallest eigenvalue  $T_1$  signifies the degree of puckering of the DNA loop from a planar circle and we call it "aplanarity of DNA loop."

In Fig. 3 is shown the major axes of DNA loop of 168bp with different supercoiling stresses, which are calculated

through the reweighting equation (2) after 2 000 000 MC runs. To access the handedness of the buckling, we also present the values of  $Wr$  and  $\Delta Tw$  versus  $\Delta Lk$  in Fig. 3(d). With supercoiling strain increasing, the planar circle will become unstable and a conformational transition from circle to figure-8 takes place at  $\Delta Lk_c \sim -1.2$ , which manifests itself as an abrupt jump in all data of major axes, writhing and twisting number, as well as bending and twisting energies (data not shown). This critical value of  $\Delta Lk_c$  is in good agreement with the analytical results [3], i.e.,  $\Delta Lk_c = \sqrt{3}A/C$ , where bending persistent length  $A$  and twisting persistent length  $C$  are taken as 53 and 72.5 nm, respectively in our simulations according to corresponding experimental data [18]. When  $\Delta Lk$  continues to increase, the figure-8 configuration will become unstable again. The DNA loop will take diplo- or triple interwound superhelical confirmation after  $\Delta Lk$  is beyond about  $-2.2$  or  $-3.4$ , which can be most clearly identified in the data of aplanarity [see Fig. 3(a)]. However, the latter transitions are less abrupt than that of a circle to figure-8, as denoted by the width of peaks in Fig. 3(a).

In summary, it was shown that a short supercoiled DNA polymer is a glassy system and a canonical Metropolis simulation tends to get stuck in some local metastable energy basins. A new Monte Carlo algorithm with an explicitly analytic weight factor is introduced to solve the problem of ergodicity breaking, in which the thermalization is reinforced in both low- and high-energy regions. Compared with an earlier approach of entropic sampling, the probability weight

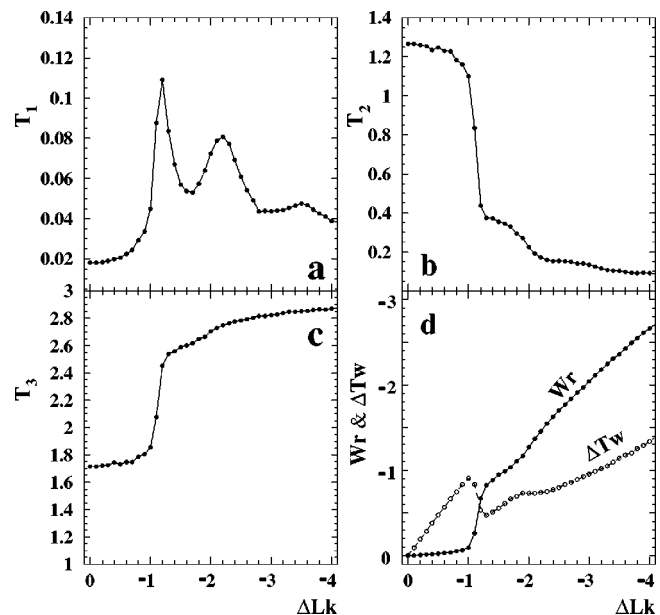


FIG. 3. Averaged major axes as well as writhing and twisting number as functions of the linking differences of the 168-base pair-DNA loop, with statistic error bars of MC calculations inside the symbols. The lines serve to guide the eye.

factor of the present algorithm is clearly easier to be determined and the implementation is about twofold CPU time-saving. As a general Monte Carlo method, the present approach can be also used in other thermodynamic systems with frustration.

To characterize quantitatively the degree of puckering of the circular DNA, a new quantity of aplanarity has been introduced as the shortest major axis of configurational tensor of DNA. With the developed Monte Carlo method, a quantitative correlation between the buckling of DNA loop of 168 base pairs and supercoiling degree is presented.

Abrupt configurational transition from circle to figure-8 takes place at the critical supercoiling stress of  $\Delta Lk_c$  ( $\sim 1.2$ ), which is in good agreement with previous analytical results. With further increasing linking difference, DNA loops will take successively diplo- and triple superhelical confirmation through the decreasingly abrupt transitions.

I am grateful to B. L. Hao, Z. C. Ou-yang, and H. J. Zhou for helpful discussions, and U. H. E. Hansmann, S. Lindsay, and J. M. Schurr for correspondence. The project is partly sponsored by SRF for ROCS, SEM.

- 
- [1] L. Stryer, *Biochemistry*, 4th ed. (Freeman and Company, New York, 1995).
- [2] L. D. Landou and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Pergamon Press, New York, 1986).
- [3] M. LeBret, *Biopolymers* **18**, 1709 (1979); C. Benham, *Phys. Rev. A* **39**, 2582 (1989).
- [4] T. Schlick and W. K. Olson, *J. Mol. Biol.* **223**, 1089 (1992); T. Schlick *et al.*, *Biopolymers* **34**, 565 (1994).
- [5] J. A. Gebe and J. M. Schurr, *Biopolymers* **38**, 493 (1996).
- [6] See, e.g., A. V. Vologodskii *et al.*, *J. Mol. Biol.* **227**, 1224 (1992); J. A. Gebe, S. A. Allison, J. B. Clendenning, and J. M. Schurr, *Biophys. J.* **68**, 619 (1995); Y. Zhang, H. J. Zhou, and Z. C. Ou-yang, *ibid.* **78**, 1979 (2000).
- [7] N. Metropolis *et al.*, *J. Chem. Phys.* **21**, 1087 (1953).
- [8] A. R. Leach, *Molecular Modelling: Principles and Applications* (Addison Wesley, Longman, 1996).
- [9] H. J. Zhou, Y. Zhang, and Z. C. Ou-yang, *Phys. Rev. Lett.* **82**, 4560 (1999); *Phys. Rev. E* **62**, 1045 (2000).
- [10] L. D. Fosdick, *Methods Comput. Phys.* **1**, 245 (1963); K. Binder, *J. Comput. Phys.* **59**, 1 (1985); A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [11] B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- [12] U. H. E. Hansmann and Y. Okamoto, *J. Comput. Chem.* **14**, 1333 (1993).
- [13] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, New York, 1999).
- [14] J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993).
- [15] As a variation, this general method has been given variant names in different occasions, e.g., “multicanonical ensemble,” “umbrella sampling.” For a review, see Ref. [13].
- [16] U. H. E. Hansmann and Y. Okamoto, *Phys. Rev. E* **56**, 2228 (1997).
- [17] J. H. White, *Am. J. Math.* **91**, 693 (1969); F. B. Fuller, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 815 (1971).
- [18] D. S. Horowitz and J. C. Wang, *J. Mol. Biol.* **173**, 75 (1984); C. Bustamante *et al.*, *Science* **265**, 1599 (1994).