Conformation of local denaturation in double-stranded DNA

Wokyung Sung and Jae-Hyung Jeon
Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Korea
(Received 7 August 2003; published 12 March 2004)

Double-stranded DNA (dsDNA) undergoes a denaturing transition above which the strands unbind completely. At temperatures (including the physiological temperature) below the transition the base pairs tend to unbind locally, giving way to loops, i.e., locally denatured states. In the flexible-chain model, the imaginary time Schrödinger equation describes the interstrand distance distribution of dsDNA with the time variable replaced by the sequence number. We transform the equation to the Fokker-Planck equation (FPE), which provides a convenient and powerful analytical method and, via the equivalent Langevin equation, a simulation scheme. The temperature-dependent potential that emerges in the FPE manifests how the DNA conformation changes dramatically near the transition temperature. We present several simulation plots along with analytical results illustrating the order parameter (concentration of bound base pairs), base pair distance correlation function, and loop size distribution at different temperatures.

DOI: 10.1103/PhysRevE.69.031902 PACS number(s): 87.14.Gg, 87.15.Aa, 05.40.−a

I. INTRODUCTION

DNA is a biomolecule that exhibits complex hierarchical structures in the nucleus to safely and effectively store the genetic information needed in living organisms. In nature, DNA exists predominantly in the *B*-form structure, a right-handed helix consisting of two strands, to protect the base pairs against background chemical contamination. However, when replication/transcription processes begin, the duplex structure is partially broken by external enzymes, giving way to loop (or bubble) formations (Fig. 1) for reading the sequence [1]. To obtain some information relevant to the process, forced-unzipping experiments on DNA have recently been extensively performed using several single-molecule tools. For example, with microapparatus such as optical tweezers, atomic force microscopy cantilevers, and magnetic tweezers, one can now measure the forces needed to unzip (or unwind) a single DNA strand with piconewton resolution and the unzipped displacements on the nanometer scale [2].

Double-stranded (ds) DNA can be globally denaturated by intrinsic thermal fluctuation as well as the external mechanical force mentioned above. The DNA undergoes a denaturation (or melting) transition at a critical temperature T_c (350−400 K), above which a dsDNA molecule separates into two single-stranded DNA molecules [3]. This phenomenon, regarded as a rare and novel example showing a one-dimensional phase transition, has been a hot issue since the 1960s. It was found by several authors [4] that the denaturation transition is a kind of unbinding transition, driven by the entropic gain which is dominant over the internal energy loss. But the discontinuity and multistep nature of the melting curve were not sufficiently explained. Recently more advanced studies including aspects of real DNA are progressing; e.g., sequence heterogeneity [5,6], stiffness and excluded volume effect of the strands [7–9], and helical structures [10].

Also, the local denaturation phenomenon or loop formation along DNA is possible due to the thermal fluctuation even far below T_c [3,11]. Since loop formation can be regarded as a precursor to denaturation, it has been partially studied in the course of denaturation studies. From the standard Peyrard-Bishop model [12,13] a Schrödinger type equation for the interstrand distance distribution of dsDNA results, allowing some information on static loop conformation. Peyrard *et al.* [12] also showed, using molecular dynamics simulation on the model, that a large loop can be excited by thermal activation and localized as a breather mode with a long lifetime. To investigate the nature of denaturation transitions, Carlon *et al.* [9] and Baiesi *et al.* [20] recently studied the interstrand distance and loop size distributions with the Poland-Scheraga model including the excluded volume effect and chain stiffness. They numerically found the critical exponents in the scaling forms of those distributions near T_c and showed that the first order nature of the transition is due to the excluded volume effect. Yet many aspects of DNA loop conformation for temperatures either physiological or approaching T_c, even within the flexible-chain model, remain to be studied analytically or simulationally. Here, we introduce an alternative methdology, i.e., a stochastic approach which enables us to analytically evaluate a number of statistical quantities for locally denaturated loops, e.g., interstrand correlation function and loop distribution.

Consider an experiment where a base pair (bp) of dsDNA is unzipped with a given separation and ask what happens to

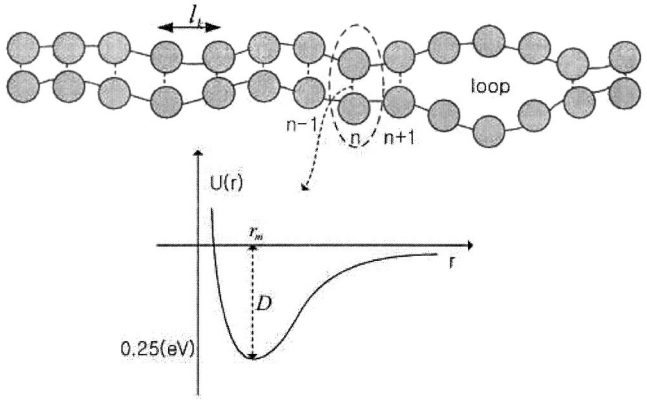

FIG. 1. Schematic figure of the model dsDNA with a loop and the Morse potential for the base pair interaction.

the bp at a distance away along the strands. The response can be related to the correlation function of the interstrand distance in accordance with the linear response theory. The correlation function also gives valuable information on the stability of the duplex structure and cooperativity of the strand characteristics of long chains. In order to theoretically investigate this and related aspects of dsDNA conformation within the standard model of the flexible chain, which remain to be clarified, we formulate the problem within the well-known Edwards (imaginary time Schrödinger) equation descriptive of the interstrand distance distribution in Sec. II. We approximate the Green's function as the sum of two contributions, one from the bound base pairs and the other from the unbound bp conformation regarded as free in the half space. To facilitate the analysis further, we reformulate the problem in stochastic dynamics language by transforming the Edwards equation to the equivalent Fokker-Planck equation (FPE) in Sec. III. The transition probability and stationary probability distributions that result facilitate analytical evaluation of the various statistical quantities in question, the average interstrand distance, correlation function, and average loop size. These quantities are also computed numerically using the equivalent Langevin equation, giving results that are in close agreement with the analytical results.

II. THE POLYMER GREEN'S FUNCTION FORMALISM

For a description of dsDNA conformation, we adopt the standard model of flexible chains used by many authors [5,8,12], as recapitulated below. In the continuum limit (where we treat chains as continuous strings), the effective Hamiltonian descriptive of dsDNA of N base pairs is [14]

$$H=\int_0^N dn\left\{\frac{3k_BT}{2l_k^2}\left[\left(\frac{\partial\mathbf{r}_1}{\partial n}\right)^2+\left(\frac{\partial\mathbf{r}_2}{\partial n}\right)^2\right]+U(|\mathbf{r}_1-\mathbf{r}_2|)\right\}. \tag{2.1}$$

The first term of the intergrand is the strand elastic energy of entropic nature due to the chain connectivity, l_k is the segmental length, $\mathbf{r}_1(n),\mathbf{r}_2(n)$ denote the positions of the two strands at the nth base pair, and $U(r)$ is the Morse potential of hydrogen bonding between a base pair at a distance r (Fig. 1):

$$U(r)=De^{-a(r-r_m)}(e^{-a(r-r_m)}-2). \tag{2.2}$$

D is the depth of the potential well, and r_m, a^{-1} are the characteristic lengths of the potential, each representing the potential minimum and width. Throughout the paper, we put $k_B=1$ so that $\beta=1/T$. Other interactions due to mismatched pairing, excluded volume effect, and twist energy will be neglected. In terms of the center of mass $\mathbf{R}=(\mathbf{r}_1+\mathbf{r}_2)/2$ and relative coordinates $\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2$, the Hamiltonian can be rewritten as the sum of two terms,

$$H_R=\int_0^N dn\,\frac{3T}{l_k^2}\left(\frac{\partial\mathbf{R}}{\partial n}\right)^2, \tag{2.3}$$

$$H_r=\int_0^N dn\left[\frac{3T}{4l_k^2}\left(\frac{\partial\mathbf{r}}{\partial n}\right)^2+U(\mathbf{r}(n))\right]. \tag{2.4}$$

Of interest to us is the information concerning the relative coordinate; from Eq. (2.4), the Edwards equation is derived [15]:

$$-\frac{\partial}{\partial n}G(\mathbf{r},\mathbf{r}_0;n)=\left[-\frac{l^2}{6}\nabla_{\vec{r}}^2+\beta U(r)\right]G(\mathbf{r},\mathbf{r}_0;n). \tag{2.5}$$

Here $l=\sqrt{2}l_k$, and $G(\mathbf{r},\mathbf{r}_0;n)$ is the polymer Green's function descriptive of the relative probability of finding the distance of the nth bp at $\mathbf{r}$, given the initial ($n=0$) distance at $\mathbf{r}_0$. Because of the spherical symmetry of the potential, we are allowed to focus on the radial part,

$$-\frac{\partial}{\partial n}G(r,r_0;n)=\left[-\frac{l^2}{6}\frac{\partial^2}{\partial r^2}+\beta U(r)\right]G(r,r_0;n). \tag{2.6}$$

The radial Green's function can be expanded:

$$G(r,r_0;n)=\sum_k e^{-n\epsilon_k}u_k(r)u_k(r_0), \tag{2.7}$$

where u_k and ϵ_k are the normalized eigenfunction and eigenvalue satisfying

$$\left[-\frac{l^2}{6}\frac{\partial^2}{\partial r^2}+\beta U(r)\right]u_k(r)=\epsilon_k u_k(r). \tag{2.8}$$

It should be noted that the index k above includes not only discrete bound states but also continuous unbound states. Also, we note that the one-dimensional equations (2.6)–(2.8) are those encountered in the theory of polymer adsorption on a flat wall, with the r denoting the vertical distance from it. This analogy often gives valuable insight.

The eigenfunctions and eigenvalues for bound states with the Morse potential are well known from the related problem in quantum mechanics [16]:

$$u_n(r)=e^{-Ky}(2Ky)^{b/2}L_n^b(2Ky), \tag{2.9}$$

$$\epsilon_n=-\frac{a^2l^2}{24}(-2n-1+2K)^2 \tag{2.10}$$

for the integer n satisfying $0\leqslant n\leqslant(2K-1)/2$, where $y(r)=\exp[-a(r-r_m)]$, $K=(6\beta D/l^2a^2)^2$, and $b=4/K-2n-1$. L_n^b is an associated Laguerre polynomial. Bound states, at least one, can exist provided $\epsilon_0<0$ or $T<T_c$, where

$$T_c=\frac{24D}{a^2l^2} \tag{2.11}$$

is the critical temperature of the unbinding (denaturation) transition, above which the two strands separate globally. Using typical parameter values $D=0.25$ eV, $l_k=3.4$ Å, and $a=2.8$ Å^{-1}, T_c is estimated to be approximately 400 K. This temperature is much lower than the bound strength D

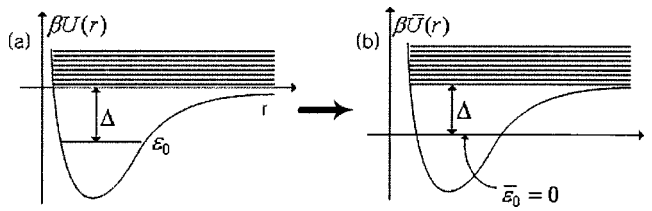

FIG. 2. The spectrum of eigenvalues of Eq. (2.8) for the Morse potential $U(r)$ and the displaced one $\bar{U}(r)=U(r)+\Delta$ where Δ is the gap between the bound and unbound states.

$\cong$3000 K because it is determined by competition between the internal energy of the bound pair and the entropy of broken pairs that can constitute loops. Using the same parameters, we note that for two bound states to exist the temperature should be lower than $T_c/9\approx 44$ K. Thus, at physiological temperatures and above, which are of our concern, it suffices to consider only the ground state of eigenvalue ϵ_0. The ground state can be rewritten as

$$u_0(r)=\exp\left(-\tfrac{1}{2}\sqrt{T_c/T}e^{-a(r-r_m)}\right.$$
$$\left.-\tfrac{1}{2}a(\sqrt{T_c/T}-1)(r-r_m)\right). \tag{2.12}$$

The other states in the summation of Eq. (2.7) form a continuum band of states (denoted by "c" below) separated by the gap $\Delta=-\epsilon_0=(a^2l^2/24)[(T_c/T)^{1/2}-1]^2$ from the ground state (Fig. 2), so that we can rewrite

$$G(r,r_0;n)=e^{n\Delta}u_0(r)u_0(r_0)+\sum_c e^{-n\epsilon_c}u_c(r)u_c(r_0). \tag{2.13}$$

As long as we consider a very long chain and large n value, the first term of the above (the bound state term) dominates; otherwise the other term, the unbound continuum contribution, albeit relatively small, should be considered for a consistent description. Here we propose that an unbound chain outside the potential well is closely approximated by a free chain in the half space $r>0$, i.e., the Green's function [17]

$$G_0(r,r_0;n)=\left(\frac{3}{2\pi nl^2}\right)^{1/2}[e^{-(3/2nl^2)(r-r_0)^2}-e^{-(3/2nl^2)(r+r_0)^2}] \tag{2.14}$$

with the hard-core boundary condition $G_0(r\to 0)=0$. Thus we consider the ansatz

$$G(r,r_0;n)=e^{n\Delta}u_0(r)u_0(r_0)+G_0(r,r_0;n), \tag{2.15}$$

which, upon substitution into Eq. (2.6) with $\partial G_0/\partial n=(l^2/6)\partial^2 G_0/\partial^2 r$, is found to be valid provided $\beta U(r)G_0(r,r_0;n)$ is negligible. We confirm the validity of the ansatz for large n: $U(r)$ is nonvanishing only within the molecular distance $r\lesssim r_m$, where G_0, varying with a significant value over the length scale $r\sim(nl^2)^{1/2}$, is quite small. Equation (2.15) is our central approximation; it yields not only a simple and intuitive picture of the bp conformation but also analytical tractability with well-defined validity, as we will show next. We note here that a shift of the potential level as shown in Fig. 2 does not affect the physics as far as the bp conformation is concerned.

With the Green's function, one can calculate a number of polymer properties. One is the average distance of the nth base pair [15]

$$\langle r(n)\rangle=\frac{\int\int\int dr_0 dr_N dr\, rG(r_N,r;N-n)G(r,r_0;n)}{\int\int dr_0 dr_N G(r_N,r_0;N)}. \tag{2.16}$$

In the long-chain limit, ground state dominance is valid, yielding

$$\langle r\rangle=\frac{\int_0^\infty dr\, ru_0^2(r)}{\int dr\, u_0^2(r)}. \tag{2.17}$$

In polymer adsorption, $\langle r\rangle$ corresponds to the thickness of the adsorbed layer, which is inversely proportional to the concentration of adsorbed monomers. Therefore the inverse of $\langle r\rangle$ is proportional to the order parameter, the fraction of bound base pairs, which is shown to vanish like $|T-T_c|$ on approaching the denaturation transition.

In dealing with other variables of interest, i.e., $\langle r(n)\rangle_{r_0}$, the average of the bp distance at the sequence n with the initial bp distance given by $r(0)=r_0$, and the related correlation function $\langle r(n)r(0)\rangle$, the approach using the Green's function is either conceptually unclear or practically cumbersome. For example, one might be tempted to generalize Eq. (2.16) and obtain

$$\langle r(n)\rangle_{r(0)}=\frac{\int\int dr_N dr\, rG(r_N,r;N-n)G(r,r_0;n)}{\int dr_N G(r_N,r_0;N)}. \tag{2.18}$$

As we shall see later, the above is not correct. We note that G, although it can be normalized, is not the probability that is conserved as n goes on. An alternative language we prefer is the transition probability $P(r,r_0;n)$, which is conserved and can remain finite, i.e., be stationary, as n approaches infinity. In terms of this transition probability the average $\langle r(n)\rangle_{r_0}$ and the correlation function are defined appropriately as is well known:

$$\langle r(n)\rangle_{r_0}=\int dr\, rP(r,r_0;n), \tag{2.19}$$

$$\langle r(n)r(0)\rangle=\int dr\, rr_0P(r,r_0;n)P_s(r_0). \tag{2.20}$$

Here, $P_s(r)$ is the stationary distribution. In the forthcoming section, we transform the Edwards equation for $G(r,r_0;n)$ into the Fokker-Planck equation for $P(r,r_0;n)$ with n now regarded as time. With the Brownian (Fokker-Planck) language, we can not only correctly evaluate the above quantities but also clearly provide the underlying physical picture.

III. THE BROWNIAN LANGUAGE DESCRIPTION

The Fokker-Planck equation

$$\frac{\partial}{\partial n}P(r,r_0;n)=\mathcal{D}\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}+\beta\frac{\partial V(r)}{\partial r}\right)P(r,r_0;n) \quad (3.1)$$

describes the stochastic motion of a Brownian particle of diffusivity $\mathcal{D}$ subject to an external potential $V(r)$, at a given time n. As is known, the equation with the substitution

$$P(r,r_0;n)=e^{-\beta[V(r)-V(r_0)]/2}G(r,r_0;n) \quad (3.2)$$

can be transformed to a Schrödinger-like equation for $G(r,r_0;n)$ [18]:

$$\frac{\partial}{\partial n}G(r,r_0;n)=\left[-\mathcal{D}\frac{\partial^2}{\partial r^2}+V_s(r)\right]G, \quad (3.3)$$

where $V_s(r)$ is obtained by

$$V_s(r)=\mathcal{D}\left[\frac{\beta^2}{4}\left(\frac{\partial V}{\partial r}\right)^2-\frac{\beta}{2}\frac{\partial^2 V}{\partial r^2}\right],$$

$$=\mathcal{D}e^{\beta V(r)/2}\frac{\partial^2}{\partial r^2}e^{-\beta V(r)/2}. \quad (3.4)$$

With the relations $\mathcal{D}=l^2/6$ and $V_s=\beta U$, we find that Eq. (3.3) is the Edwards equation. The inverse problem of finding $V(r)$ from $U(r)$ is given by rewriting Eq. (3.4) as

$$\left[-\frac{l^2}{6}\frac{\partial^2}{\partial r^2}+\beta U(r)\right]e^{-\beta V(r)/2}=0. \quad (3.5)$$

The equation shows that $e^{-\beta V(r)/2}$ is the eigenfunction corresponding to the zero eigenvalue in the stationary Edwards equation, Eq. (2.8), $u_0(r)=c^{1/2}e^{-\beta V(r)/2}$, where c is a normalization constant. Then Eq. (3.2) can be rewritten as

$$P(r,r_0;n)=\frac{u_0(r)}{u_0(r_0)}\sum_k e^{-n\epsilon_k}u_k(r)u_k(r_0). \quad (3.6)$$

In order to find $V(r)$, we note that our information concerning the bp conformation remains invariant with respect to a shift of the potential in the original Edwards problem. We shift the potential $U(r)$ to $U(r)+\Delta$, where the shifted lowest eigenvalue is zero, $\overline{\epsilon_0}=0$ [Fig. 2(b)], so that the transition probability, which can be rewritten as $P(r,r_0;n)=[u_0(r)/u_0(r_0)][u_0(r)u_0(r_0)+\Sigma_c e^{-n\overline{\epsilon_c}}u_c(r)u_c(r_0)]$, retains the stationarity:

$$P(r,r_0;n\to\infty)=P_s(r)=u_0^2(r)=ce^{-\beta V(r)}. \quad (3.7)$$

The eigenfunction $u_0(r)$, independent of the shift, readily yields

$$\beta V(r)=\left(\frac{T_c}{T}\right)^{1/2}e^{-a(r-r_m)}+a\left[\left(\frac{T_c}{T}\right)^{1/2}-1\right](r-r_m), \quad (3.8)$$

apart from a constant term. As we see above, $V(r)$, the potential under which the Brownian particle moves, has a suggestive temperature dependence. Figure 3 shows that, for temperatures far below T_c, the Brownian motion is manifestly confined in a narrow well around a small distance $\sim r_m$. As the temperature approaches T_c, however, the potential tends to be flat for large r, so that the Brownian motion becomes unbounded, with the average base pair distance diverging. In this language, the central ansatz Eq. (2.15) is rewritten as

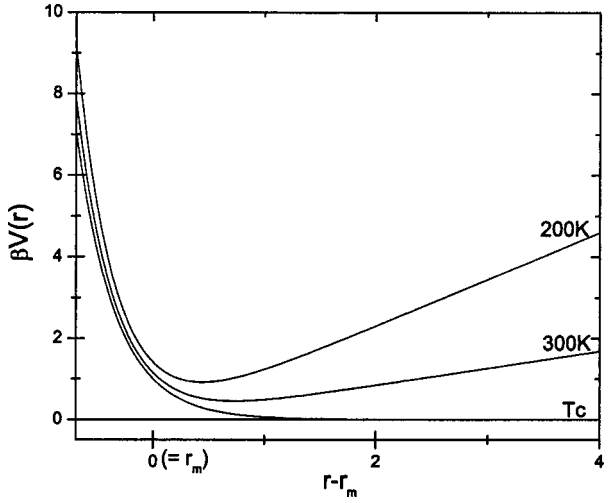

FIG. 3. Plots of temperature-dependent potential $V(r)$. The x axis (Å) indicates the relative distance from r_m. The confinement effect by the potential completely disappears at T_c. The parameter values are the same as in the previous section.

$$P(r,r_0;n)=\frac{u_0(r)}{u_0(r_0)}[u_0(r)u_0(r_0)+e^{-n\Delta}G_0(r,r_0;n)]. \quad (3.9)$$

To visualize and simulate the stochastic motion embodied in the above Fokker-Planck description, we consider the equivalent Langevin equation

$$\gamma\frac{dr(n)}{dn}=-\frac{\partial V(r)}{\partial r}+\xi(n), \quad (3.10)$$

where $\gamma=T/\mathcal{D}$ is the friction coefficient, and $\xi(n)$ is the Gaussian and white noise satisfying $\langle\xi(n)\rangle=0$, $\langle\xi(n)\xi(n')\rangle=2\gamma T\delta(n-n')$. By integrating the Langevin equation, one can make a move algorithm from which the trajectories of the Brownian particle as a function of n are simulated. In Fig. 4, we show ensembles of the base pair distance of 500-bp-long dsDNA via a Langevin simulation with initial distance put to $r_0=r_m$ and $\Delta n=0.0001$. We use the parameters introduced in Sec. II, with $T_c\approx 400$ K. Remarkably, the thermally broken pairs appear to be ubiquitous even far below T_c. As the temperature increases approaching T_c, the loops tend to be larger and eventually become comparable to DNA in size.

Using the definition Eq. (2.19), the average distance is given by

$$\langle r\rangle_{r_0}=\langle r\rangle_\infty+\langle\delta r\rangle_{r_0}. \quad (3.11)$$

Here $\delta r=r-\langle r\rangle_\infty$ and

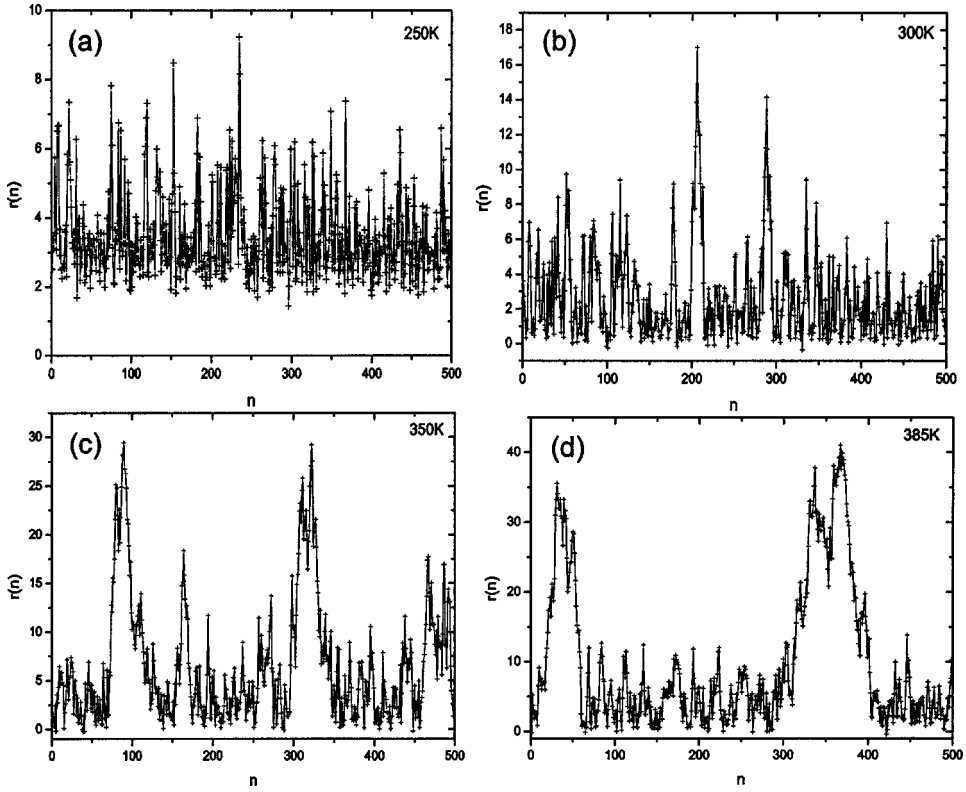

FIG. 4. Profiles of base pair distances of 500-bp-long dsDNA molecules obtained via Langevin simulation for varying temperatures: (a) 250, (b) 300, (c) 350, and (d) 385 K. All simulations are accomplished with $r(0)=r_m$ (Å) and $\Delta n=0.0001$. The parameters are the values presented in Sec. II, resulting in $T_c \approx 400$ K, where lots of thermally broken base pairs are ubiquitous even far below T_c, and a few large loops can be seen at high temperatures. Near T_c the size of the loops is the order of the length of the DNA itself.

$$\langle r\rangle_\infty = \int dr\, r P_s(r) = \frac{\int_0^\infty dr\, r e^{-\beta V(r)}}{\int_0^\infty dr\, e^{-\beta V(r)}} \tag{3.12}$$

is the average over the stationary state achieved over long time (large n). It is identical to Eq. (2.17). Figure 5 shows the analytically calculated value of $\langle r\rangle$ as a function of temperature. It is nearly identical to the result $\langle r\rangle_{r_0}$ that was obtained by simulating a long chain $n\sim 10^4-10^5$ bps. $\langle r\rangle^{-1}$ is found to be close to the experimental plot of the order parameter (fraction of bounded complementary base pairs) [3]. The correction due to finite n can be obtained by

$$\langle \delta r\rangle_{r_0} = e^{-n\Delta} \int_0^\infty dr\, r \frac{u_0(r)}{u_0(r_0)} G_0(r, r_0; n). \tag{3.13}$$

Figure 6 shows $\langle r\rangle_{r_0}$ at a physiological temperature with three different values of r_0. The simulation data using Langevin dynamics are shown to be in close agreement with the results obtained by Eq. (3.13), which demonstrates that our approximation (3.9) or (2.15) is excellent. To evaluate the integral we note that $u_0(r)$ is nonvanishing only over the molecular distance, independent of n, over which G_0 does not vary appreciably. Considering n large enough to assume $|r \pm r_0| \ll (nl^2)^{1/2}$ and temperature much lower than T_c where the integral below is bounded, Eq. (3.13) can be rewritten as

$$\langle \delta r\rangle_{r_0} \approx e^{-n\Delta} \left(\frac{3}{2\pi n l^2}\right)^{1/2} \int_0^\infty dr\, r \frac{u_0(r)}{u_0(r_0)} \frac{6 r r_0}{n l^2},$$
$$\sim \frac{e^{-n\Delta}}{n^{3/2}}. \tag{3.14}$$

This function, also the correlation function shown next, decays exponentially as well as in a power-law-like fashion. The presence of a large gap Δ shows the stability of the DNA duplex against spontaneous loop formation. When some sequence is unzipped by a certain microscopic manipulation, the disturbance along the neighboring sequences propagates over the correlation length Δ^{-1}, which is about 5 bps distance at a physiological temperature. On the other hand, the power-law decay is characteristic of the chain cooperativity (connectivity). The power-law and exponential decays in n figure most naturally in the related correlation function $\langle \delta r(n)\, \delta r(0)\rangle = \langle r(n) r(0)\rangle - \langle r\rangle_\infty^2$. From Eq. (2.20),

$$\langle \delta r(n)\, \delta r(0)\rangle = \int dr_0 P_s(r_0) r_0 \langle \delta r\rangle_{r_0},$$
$$\approx e^{-n\Delta} \left(\frac{3}{2\pi n l^2}\right)^{1/2} \frac{6}{n l^2} \left[\int dr\, u_0(r) r^2\right]^2, \tag{3.15}$$

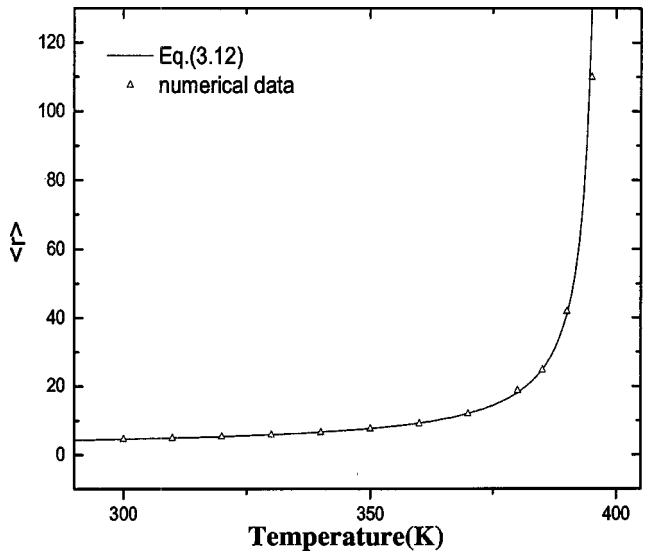

FIG. 5. $\langle r\rangle$ vs temperature. The numerical simulation values (triangles) agree remarkably well with the analytical result (line) $\langle r\rangle_\infty$ [Eq. (3.12)]. To assure that $\langle r(n)\rangle$ was independent of $r(0)$, we considered very large values of n in a long chain $N=10^4-10^5$.

which shows that under the identical conditions it thus shares the same dependence on n as $\langle \delta r\rangle_{r_0}$ studied above.

Finally, let us focus upon the loop size distribution. The loop is defined by a random walk of base pair distance that returns, for the first time, to its initial distance r_0, which is set infinitesimally small. To calculate the distribution of this first passage time (i.e., the loop size), we apply the absorbing boundary condition $P(r_0,n|r_0)=0$ for $n>0$, i.e., wherever a bp returns to its initial distance r_0, the excursion is finished. The probability that the bp is separated by a distance larger than r_0 is given by

$$P(n,r_0)=\int_{r_0}^{\infty} dr\, P(r,n|r_0), \tag{3.16}$$

in terms of which the probability that the random walker arrives at r_0 in dn, i.e., the loop distribution, is given by [19]

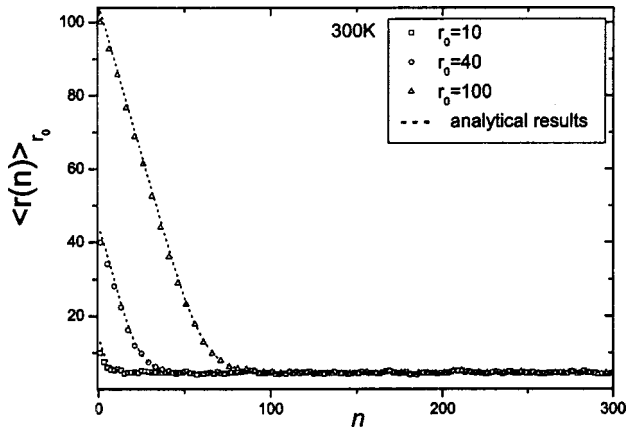

FIG. 6. $\langle r(n)\rangle_{r_0}$ with several r_0 values (10, 40, 100 Å). The curves obtained from Eqs. (3.11) and (3.13) agree closely with the simulation data.

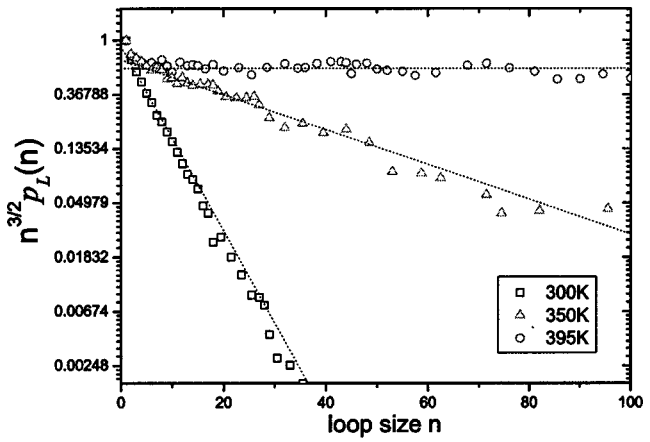

FIG. 7. Loop size distributions for several temperatures. The straight lines are the outcome of Eq. (3.18), with the slopes given by Δ (note the vertical logarithmic scale). Near T_c ($T=395$ K), $\Delta\approx 0$, where the distribution follows a power law.

$$p_L(n)=-\frac{\partial}{\partial n}P(n,r_0). \tag{3.17}$$

$p_L(n)$ is solely given by the unbound (second) contribution in Eq. (3.9), which naturally reflects the process of excitation (crossing) over the gap caused by base pair bonding. Using the same approximation used before for finding $\langle \delta r\rangle_{r_0}$, Eq. (3.14), one again obtains $P(n,r_0)\sim e^{-n\Delta}/n^{3/2}$ for n larger than the order of unity far below T_c. Near T_c, however, the approximation is not allowed, but instead Eq. (3.9) yields $P(n,r_0)$, which is proportional to $1/n^{1/2}$, not to $n^{3/2}$. Therefore we find that the loop distribution function for large n is given by

$$p_L(n)\sim\begin{cases}\dfrac{e^{-n\Delta}}{n^{3/2}} & \text{far below } T_c,\\[2ex] \dfrac{1}{n^{3/2}} & \text{near } T_c.\end{cases} \tag{3.18}$$

We confirm the validity of the above results against our numerical simulation for a number of temperatures. The loop distribution at each temperature in Fig. 7 is given for 500 500-bp-long DNAs along with the theoretical expectation represented by the dotted lines. As shown in the figures, $p_L(n)$ given by Eq. (3.18) is in good agreement with the data even for relatively small n values and for a broad range of temperatures. The combined exponential and power-law decay of $p_L(n)$ closely resembles the recent result of the Poland-Scheraga- (PS-) type model [9,20], where the power is larger than 3/2 due to the excluded volume effect and chain stiffness. We find the average size of the loop as

$$\langle n\rangle=\frac{\Sigma_n n p_L(n)}{\Sigma_n p_L(n)}=\frac{\int_{n_c}^{\infty} e^{-n\Delta}n^{-1/2}dn}{\int_{n_c}^{\infty} e^{-n\Delta}n^{-3/2}dn}. \tag{3.19}$$

For T near T_c, it is found that $\langle n\rangle\sim\Delta^{-1/2}\sim|T-T_c|^{-1}$, irrespective of the cutoff n_c, the scaling behavior obeyed by $\langle r\rangle$. This shows indeed that, as the denaturation transition is ap-

proached, the loops tend to give way to fewer but larger loops until the size grows to the size of the DNA molecule itself, which was also demonstrated in the simulation result (Fig. 4).

IV. SUMMARY

In this paper, we introduced a Brownian (stochastic) description version of flexible polymer Green's function theory to study the interstrand (base pair) distance of dsDNA, and several associated problems. Within our approach the base pair distance $r(n)$ is regarded as the coordinate of a Brownian particle moving under a temperature-dependent potential, with n indicating the segment number. The Fokker-Planck and the equivalent Langevin equations are formulated from the Edwards (imaginary time Schrödinger) equation of base pair distance distribution via the well-known transformation rule. The interstrand distribution function is a sum of the bound and unbound base pair contributions; the latter is approximated by the free chain distribution in half space.

Using the Langevin equation, we performed several simulations of the DNA base pair distances with initial bp distance fixed as $r(0)=r_0$ in a broad range of temperature. We observed that loop formation can be excited at temperatures even far below the denaturation temperature (T_c), with the average size of the loops growing as the temperature increases toward T_c. We also calculated the average and the correlation function of the base pair distance $[\langle r(n)\rangle_{r_0}$ and $\langle r(n)r(0)\rangle]$ by analytically solving the Fokker-Planck equation. According to the results, the correlations $[\langle \delta r(n)\rangle_{r_0}, \langle \delta r(0)\delta r(n)\rangle]$ decay as the product of exponential ($e^{-n\Delta}$) and power-law ($n^{-3/2}$) behaviors provided that n is large enough and the temperature is far below T_c. This result is in close agreement with our simulation data. Interestingly, the loop size distribution is found to have the same n dependency, which, unlike the correlation function, is appliable to broad ranges of n and temperature. Approaching T_c, the loop size diverges like $|T-T_c|^{-1}$, and its distribution eventually follows the power law $p_L(n)\sim n^{-3/2}$. These behaviors, also corroborated by our Langevin simulation, are consistent with the simulation result [20] recently obtained by using the PS-type model that includes the chain semiflexibility and the excluded volume effect.

Our methodology of treating the problem in the language of stochastics allows us to treat naturally not only quantities like the correlation function and average loop size, but also sequence heterogeneity and randomness, which can be treated as, in general, a colored noise changing in time n. The stochastic dynamics paradigm that has recently been obtained for a variety of time-dependent fluctuations can be a guide to study DNA conformation. However, this methodology is limited to the flexible-chain model; adaptation to a more realistic model inclusive of chain semiflexibility and helical structure remains as a future challenge.

ACKNOWLEDGMENTS

The authors thank D. S. Koh for helpful discussions. This work was supported by KRF (Grant No. ds0008) and the National RND program under MOST (Grant No. M10214000230).

[1] B. Alberts *et al.*, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York, 2000).
[2] C. Bustamante *et al.*, Curr. Opin. Struct. Biol. **10**, 279 (2000); H. Clausen-Schaumann *et al.*, Biophys. J. **78**, 1997 (2000); U. Bokelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. E **58**, 2386 (1998); I. Rousina and V. A. Bloomfield, Biophys. J. **80**, 882 (2001); S. Cocco, R. Monasson, and J. F. Marko, Phys. Rev. E **65**, 041907 (2002).
[3] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[4] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1464 (1966); Michael E. Fisher, *ibid.* **45**, 1469 (1966); M. Ya. Azbel, Phys. Rev. A **20**, 1671 (1979).
[5] D. K. Lubensky and D. R. Nelson, Phys. Rev. E **65**, 031917 (2002).
[6] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997); L. Tang and H. Chate, *ibid.* **86**, 830 (2001); T. Hwa, E. Marinari, K. Sneppen, and L. Tang, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003).
[7] M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000); Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000); T. Garel, C. Monthus, and H. Orland, Europhys. Lett. **55**, 132 (2001).
[8] N. Theodarakopoulos, T. Dauxois, and M. Peyrard, Phys. Rev. Lett. **85**, 6 (2000); T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[9] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).
[10] S. Cocco and R. Monasson, Phys. Rev. Lett. **83**, 5178 (1999); J. Rudnick and R. Bruinsma, Phys. Rev. E **65**, 030902(R) (2002).
[11] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
[12] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989); T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993); M. Peyrard and J. Farago, Physica A **288**, 199 (2000).
[13] T. Lipniacki, Phys. Rev. E **64**, 051919 (2001).
[14] The usual kinetic term $\int dn\, m(\dot{\mathbf{r}}_1^2+\dot{\mathbf{r}}_2^2)/2$ is neglected here since it is irrelevant to the equilibrium state of DNA we study.
[15] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).
[16] P. M. Morse, Phys. Rev. **34**, 57 (1929); C. L. Pekeris, *ibid.* **45**, 98 (1934).
[17] P. J. Park and W. Sung, Phys. Rev. Lett. **77**, 783 (1996).
[18] H. Risken, *The Fokker-Planck Equation*, 2nd ed. (Springer-Verlag, Berlin, 1989).
[19] L. E. Reichl, *A Modern Course in Statistical Physics* (University of Texas Press, Austin, 1980).
[20] M. Baiesi *et al.*, Phys. Rev. E **67**, 021911 (2003).