Pulling hairpinned polynucleotide chains: Does base-pair stacking interaction matter?

Haijun Zhou[a)]
Max-Planck-Institute of Colloids and Interfaces, D-14424 Potsdam, Germany

Yang Zhang
Institute of Theoretical Physics, The Chinese Academy of Sciences, P.O. Box 2735, Beijing 100080, China

(Received 22 November 2000; accepted 7 March 2001)

Force-induced structural transitions both in relatively random and in designed single-stranded DNA (ssDNA) chains are studied theoretically. At high salt conditions, ssDNA forms compacted hairpin patterns stabilized by base pairing and base-pair stacking interactions, and a threshold external force is needed to pull the hairpinned structure into a random coiled one. The base-pair stacking interaction in the ssDNA chain makes this hairpin-coil conversion a discontinuous (first-order) phase transition process characterized by a force plateau in the force-extension curve, while lowering this potential below some critical level turns this transition into continuous (second-order) type, no matter how strong the base-pairing interaction is. The phase diagram (including hairpin-I, -II, and random coil) is discussed as a function of stacking potential and external force. These results are in quantitative agreement with recent experimental observations of different ssDNA sequences, and they reveal the necessity to consider the base-pair stacking interactions in order to understand the structural formation of RNA, a polymer designed by nature itself. The theoretical method used may be extended to study the long-range interaction along double-stranded DNA caused by the topological constraint of fixed linking number.
[DOI: 10.1063/1.1368401]

I. INTRODUCTION

Single-macromolecular manipulation techniques, such as atomic force microscopy and optical tweezer methods, were recently used by many authors to investigate the mechanical properties of double-stranded DNA (dsDNA) and to understand how the tension in dsDNA influences the interactions between DNA and proteins such as RNA polymerases, type II topoisomerases, and RecA proteins (for a brief review see Ref. 1). Many useful insights have been obtained through a detailed analysis of the experimental data generated by these precise measurements (see, for example, Refs. 2–9). These new experimental techniques were recently used also on RNA and single-stranded DNA (ssDNA) to explore the structure formation of these polynucleotides.[1,10–12] A RNA or ssDNA is a linear chain of nucleotide bases. At physiological conditions (concentration ~ 0.1 M Na^+ and temperature ~ 300 K), RNA molecules in biological cells can fold into stable native configurations (as in the case of proteins) to fulfill their various biological functions.[13] The higher-order structures in RNA and ssDNA are caused by the possibilities to form base pairs between the complementary nucleotide bases, A and U (T, in the case of ssDNA), and G and C. These structures can be further stabilized by the vertical stacking interactions between nucleotide base pairs. In a long polynucleotide chain, because there are a great many different ways to form base pairs, the accessible configurational space is very large, and the equilibrium configurations are determined by the competition between entropy and the interaction energy. To investigate how base pairing and base-pair stacking influence structure formations of polynucleotide is therefore of both biological and theoretical interest.

Recent measurements demonstrated that the structures of polynucleotides are sensitive to the ionic strength of the aqueous solution.[1,10–12] In a high salt environment (for example, 150 mM Na^+ or 5 mM Mg^{2+}), electrostatic repulsive forces between the negatively charged nucleotide phosphate groups are largely shielded and base-pairing is favored; furthermore, experimental observations suggested the formation of hairpinned configurations in ssDNA.[1] To understand the hairpin formation in ssDNA and RNA, Montanari and Mézard suggested very recently a model for these polynucleotides.[14] Base-pairing interaction is considered in their model, and their theoretical calculation can reproduce the experimental force-extension curve measured on a ssDNA chain whose nucleotide bases are arranged in relatively random order. It can be shown (see below and also Ref. 14) that the force-induced structural transition in ssDNA is a second-order continuous phase transition process, characterized by a gradual decrease in the number of base pairs as the external force is increased beyond a certain threshold value.

On the other hand, Rief *et al.*[15] observed a quite different phenomenon: When they pulled a ssDNA chain made of poly(dA-dT) or poly(dG-dC) sequence apart with an atomic force microscope, they found that the distance between its

a)Author to whom correspondence should be addressed; electronic mail: zhou@mpikg-golm.mpg.de

two ends suddenly changes from nearly zero to a value comparable to its total contour length during a very narrow force range.

A question thus arises: Why is the force-induced structural transition of ssDNA gradual and continuous in one case and highly cooperative in the other case, even though the solution conditions in these two kinds of experiments are comparable? We noticed that there may be strong base-pair stacking interactions in the designed ssDNA chains, while stacking interactions are absent or negligible in the relatively random sequences. We have previously shown that the short-ranged base-pair stacking potential in dsDNA dominates the energy contribution[16] and it is the reason both for the stability of dsDNA and its high extensibility at large forces.[7,8] Is the dramatically different experimental observation of Rief *et al.*[15] (compared with those of Refs. 1,11 and 12) caused by base-pair stacking interaction? Given the fact that the stacking potential is a kind of short-ranged (nearest-neighbor) interaction, does the cooperative transition reported in Ref. 15 for such an one-dimensional (1D) system really differ qualitatively from that observed in Refs. 1 and 11?

To address these questions, we extend the work of Montanari and Mézard[14] to include the effect of stacking interactions, and thus obtain a more realistic model for polynucleotides. We find that when the base-pair stacking potential is small, the ssDNA chain is in a hairpinned macro-state with the base pairs only weakly stacked (hairpin-I), while when the stacking potential is large, the ssDNA is in a hairpinned macro-state with almost complete pairing and stacking (hairpin-II). The force-induced transition between hairpin-II and random coil is first-order, while the transition between hairpin-I and random coil is continuous. There is also a continuous phase transition between hairpin-I and hairpin-II, which may be induced by changing the nucleotide arrangement or by environmental variations in temperature or ionic strength. A simplified treatment of the same question was reported earlier.[17]

The paper is organized as following: Section II introduces the theoretical model of ssDNA; and Sec. III presents some qualitative arguments concerned with the structural transitions in the model of ssDNA; in Sec. IV we perform a direct comparison with experimental observations. The main points of the paper are summarized in Sec. V, accompanied by some qualitative discussions on RNA secondary structure formation. We suggest the possibility to extend the theoretical method in this paper to address the long-range interactions along topologically constrained double-stranded DNA molecules, i.e., dsDNA with fixed linking number.

II. MODEL OF SSDNA WITH BASE-PAIR STACKING INTERACTION

The model used in the present study is similar with that developed earlier by Montanari and Mézard,[14] albeit with base-pair stacking potential included. We discuss a fictitious polymer chain made of N tiny beads (bases) with index i from 1 to N (each base in the model may correspond to several bases in a realistic polynucleotide). Between any two adjacent bases (i and $i+1$) there is an elastic bond of equilibrium length b and length variance l ($l \ll b$, i.e., the bond is considerably rigid). For any two bases i and j, if their mutual distance $|\mathbf{r}_{ij}|$ is less than an interaction radius a there could be a pairing potential $V^{i,j}_{\text{pair}}(\mathbf{r}_{ij})$ (and we say that an i-j base pair is formed); and if any two base pairs are nearest neighbors to each other [that is, i-j and $(i+1)$-$(j-1)$], there is an additional stacking interaction $J^{j,j-1}_{i,i+1}(\mathbf{r}_{ij},\mathbf{r}_{i+1,j-1})$. In this work, as what is assumed in other previous theoretical treatments,[18–20] only the secondary pairing patterns are considered, which leads to the following constraints on the pairing between bases: (1) Each base can be unpaired or be paired to at most one another base; (2) if base i is paired to base j (suppose $i<j$) and k to l ($k<l$), then either $i<j<k<l$ (''independent'') or $i<k<l<j$ (''nested'').

Because of these restrictions, the partition function can be calculated iteratively.[14,18,19] In the case when the pairing interaction radius a is far less than bond equilibrium length b ($a \ll b$), the total statistical weight $Z_{j,i}(\mathbf{r})$ for a polynucleotide segment (from base i to j) whose ends being separated by a distance $\mathbf{r}$ is expressed as

$$
\begin{aligned}
Z_{j,i}(\mathbf{r}) = {} & \int d\mathbf{u}\mu(\mathbf{u}) Z_{j-1,i}(\mathbf{r}-\mathbf{u}) + f_{ji}(\mathbf{r}) \int d\mathbf{u}_1 d\mathbf{u}_2 \mu(\mathbf{u}_1)\mu(\mathbf{u}_2) Z_{j-1,i+1}(\mathbf{r}-\mathbf{u}_1-\mathbf{u}_2) + \sum_{k=i+1}^{j-2} \int d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 d\mathbf{v} \mu(\mathbf{u}_1) \\
& \times \mu(\mathbf{u}_2)\mu(\mathbf{u}_3) f_{jk}(\mathbf{v}) Z_{k-1,i}(\mathbf{r}-\mathbf{u}_1-\mathbf{v}) Z_{j-1,k+1}(\mathbf{v}-\mathbf{u}_2-\mathbf{u}_3) + s(a-|\mathbf{r}|)\exp(-\beta V^{j,i}_{\text{pair}}(\mathbf{r})) \int d\mathrm{u}_1 d\mathrm{u}_2 \mu(\mathbf{u}_1) \\
& \times \mu(\mathbf{u}_2) g^{j,j-1}_{i,i+1}(\mathbf{r},\mathbf{r}-\mathbf{u}_1-\mathbf{u}_2) Z^{(p)}_{j-1,i+1}(\mathbf{r}-\mathbf{u}_1-\mathbf{u}_2) + \sum_{k=i+1}^{j-2} \int d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 d\mathbf{v} s(a-|\mathbf{v}|) \\
& \times \exp(-\beta V^{j,k}_{\text{pair}}(\mathbf{v})) g^{j,j-1}_{k,k+1}(\mathbf{v},\mathbf{v}-\mathbf{u}_2-\mathbf{u}_3) Z_{k-1,i}(\mathbf{r}-\mathbf{u}_1-\mathbf{v}) Z^{(p)}_{j-1,k+1}(\mathbf{v}-\mathbf{u}_2-\mathbf{u}_3). \qquad (1)
\end{aligned}
$$

Here, $\mu(\mathbf{r}) \propto \exp(-(|\mathbf{r}|-b)^2/2l^2)$ is the probability density for the bond vector $\mathbf{r}$; $f_{ji}(\mathbf{r}) = \exp[-\beta V^{j,i}_{\text{pair}}(\mathbf{r}_{ij})] - 1$ characterizes the pairing interaction between bases i and j (this viral expansion form f is introduced because that, in the first integral on the right-hand side of the equality in Eq. (1), an vanishing pairing potential between the end base j and any another base is assumed and the contributions of all these possible base-paired configurations are included. Such

spurious contributions should be removed since the pairing potential is actually nonzero); $s(x)$ is a signal function, $s(x)=1$ if $x\geqslant 0$ and 0 otherwise; $g_{i,i+1}^{j,j-1}(\mathbf{r}_{i,j},\mathbf{r}_{i+1,j-1})=\exp[-\beta J_{i,i+1}^{j,j-1}(\mathbf{r}_{ij},\mathbf{r}_{i+1,j-1})]-1$ characterizes the base-pair stacking interaction (the reason for the introduction of the viral coefficient g can be similarly understood as that for the introduction of the viral coefficient f). $Z_{j,i}^{(p)}$ is the statistical weight for a ssDNA segment whose two end bases (i and j) forming a base pair

$$Z_{j,i}^{(p)}(\mathbf{r})=s(a-|\mathbf{r}|)\exp(-\beta V_{\text{pair}}^{j,i}(\mathbf{r}))\int d\mathbf{u}_1 d\mathbf{u}_2 \mu(\mathbf{u}_1)\mu(\mathbf{u}_2) \times Z_{j-1,i+1}(\mathbf{r}-\mathbf{u}_1-\mathbf{u}_2)+s(a-|\mathbf{r}|) \times\exp(-\beta V_{\text{pair}}^{j,i}(\mathbf{r}))\int d\mathbf{u}_1 d\mathbf{u}_2 \mu(\mathbf{u}_1)\mu(\mathbf{u}_2) \times g_{i,i+1}^{j,j-1}(\mathbf{r},\mathbf{r}-\mathbf{u}_1-\mathbf{u}_2)Z_{j-1,i+1}^{(p)}(\mathbf{r}-\mathbf{u}_1-\mathbf{u}_2). \quad (2)$$

The statistical property of this model system is determined by the coupled equations (1) and (2). In the general case where the pairing and stacking potentials are dependent on base index, it is certainly of no hope to investigate the property of the system analytically. Here we simplify our task by assuming that the pairing and stacking potentials between any bases have the same form, i.e., assuming the polymer chain to be homogeneous.[21] Then in Eqs. (1) and (2) all the subscripts specifying specific bases can be dropped, especially we can write $f_{ji}(\mathbf{r})$ as $f(\mathbf{r})$, $Z_{j,i}(\mathbf{r})$ as $Z_{j-i}(\mathbf{r})$ and $Z_{j,i}^{(p)}(\mathbf{r})$ as $Z_{j-i}^{(p)}(\mathbf{r})$. In the model, the stacking potential is a function of the interbase distances for the two base pairs concerned. Since the base pair's interbase distance is usually very small (less than a), we can approximate the stacking potential to be constant in this range, and hence we just denote the stacking potential to be a constant J and denote $g=g_{i,i+1}^{j,j-1}(\mathbf{r}_{i,j},\mathbf{r}_{i+1,j-1})=\text{const}$.

The Fourier transforms of the generating functions (Laplace transforms) of the statistical weights are defined as

$$\Xi(\zeta,\mathbf{p})=\int d\mathbf{r}\left(\sum_{n=0}^{\infty} Z_n(\mathbf{r})\zeta^n\right)\exp(i\mathbf{p}\cdot\mathbf{r}) \quad (3)$$

and

$$\Xi^{(p)}(\zeta,\mathbf{p})=\int d\mathbf{r}\left(\sum_{n=0}^{\infty} Z_n^{(p)}(\mathbf{r})\zeta^n\right)\exp(i\mathbf{p}\cdot\mathbf{r}). \quad (4)$$

Considering the iterative expressions for Z_n and $Z_n^{(p)}$ in Eqs. (1) and (2), we can show that

$$\Xi(\zeta,\mathbf{p})=[1+\zeta\sigma(\mathbf{p})\Xi(\zeta,\mathbf{p})] \times\left(1+\frac{\zeta^2 b^3}{(2\pi)^3}\left[\int d\mathbf{q}\sigma^2(\mathbf{q})\gamma(\mathbf{p}-\mathbf{q})\Xi(\zeta,\mathbf{q}) +g\int d\mathbf{q}\sigma^2(\mathbf{q})\gamma'(\mathbf{p}-\mathbf{q})\Xi^{(p)}(\zeta,\mathbf{q})\right]\right) \quad (5)$$

and

$$\Xi^{(p)}(\zeta,\mathbf{p})=\frac{\zeta^2 b^3}{(2\pi)^3}\int d\mathbf{q}\sigma^2(\mathbf{q})\gamma'(\mathbf{p}-\mathbf{q}) \times[\Xi(\zeta,\mathbf{p})+g\Xi^{(p)}(\zeta,\mathbf{q})], \quad (6)$$

where $\sigma(\mathbf{p})=\int d\mathbf{r}\mu(\mathbf{r})\exp(i\mathbf{p}\cdot\mathbf{r})=(\sin pb/pb)\exp(-p^2l^2/2)$, with $\sigma(\mathbf{0})=1$; $\gamma(\mathbf{p})=b^{-3}\int_{|\mathbf{r}|\leqslant a}d\mathbf{r}f(\mathbf{r})\exp(i\mathbf{p}\cdot\mathbf{r})$ and $\gamma'(\mathbf{p})=\gamma(\mathbf{p})+(a/b)^3$. Since $a\ll b$, we have $\gamma'=\gamma$ and both γ and γ' can be regarded as independent of momentum $\mathbf{p}$. Then Eqs. (5) and (6) lead to

$$\Xi(\zeta,\mathbf{p})=\frac{D(\zeta)}{1-\sigma(\mathbf{p})D(\zeta)}, \quad (7)$$

where

$$D(\zeta)=\zeta+\zeta\Xi^{(p)}(\zeta)=-\eta_1\zeta^3+\eta_2\zeta^2+\zeta, \quad (8)$$

with coefficient η_1 a constant and η_2 related to D:

$$\eta_1=g\gamma b^3(2\pi)^{-3}\int d\mathbf{q}\sigma^2(\mathbf{q})=g\gamma(4\pi)^{-3/2}(b/l), \quad (9)$$

$$\eta_2(D)=D\eta_1\left[1+(b^2l/g\pi^{3/2})\int d\mathbf{q}\frac{\sigma^2(\mathbf{q})}{1-\sigma(\mathbf{q})D}\right]. \quad (10)$$

Equations (7) and (8) are the central equations of this article.

When a external force $\mathbf{F}$ is applied at the end of the ssDNA chain, the total partition function is $Z_N^F=\int d\mathbf{r}Z_N(\mathbf{r})\exp(-\beta\mathbf{F}\cdot\mathbf{r})$. The Laplace transform of this partition function is calculated to be

$$\sum_{N=0}^{\infty} Z_N^F\zeta^N=\Xi(\zeta,-i\beta\mathbf{F}), \quad (11)$$

where $\Xi(\zeta,-i\beta\mathbf{F})$ is determined by Eq. (7).

For a linear polymer system, the free energy is an extensive quantity proportional to the number of monomers N in the thermodynamic limit. This indicates that the smallest positive singularity point of the function $\Xi(\zeta,-i\beta\mathbf{F})$ in the variable ζ corresponds to the free energy density of the ssDNA chain.[22,23] This point is used in the next section.

III. QUALITATIVE ANALYSIS OF THE HAIRPIN-COIL TRANSITION

It is evident that $\Xi(\zeta,-i\beta\mathbf{F})$ has a pole ζ_{pole} determined by

$$D(\zeta_{\text{pole}})=1/\sigma(-i\beta\mathbf{F})=\frac{\beta Fb}{\sinh(\beta Fb)}\exp(-\beta^2F^2l^2/2), \quad (12)$$

where $F=|\mathbf{F}|$. The function at the right side of the equality monotonously decreases with F from 1 at $F=0$ to 0 as $F\to\infty$. On the other hand, the function $D(\zeta)$ is related to the Laplace transform of the hairpinned configurations as indicated by Eq. (8), therefore, $D(\zeta)$ has a finite convergence radius, and it is an increasing function of ζ before this radius is reached. The singularity of $D(\zeta)$ is related to the roots of Eq. (8). Although analytical expressions for the roots of this third-order equation are available, they are lengthy and here we discuss the statistical property of the system through an alternative routine. First we consider two extreme cases:

Case A: the base-pair stacking potential $J=0$. In this case $g=0$ and $\eta_1=0$ and Eq. (8) reduces to second-order. This situation has been studied by Montanari and Mézard[14] and they found that $D(\zeta)$ has a second-order branching point at $\zeta_{\rm bp}$ equaling to the maximum of the expression $(-1+\sqrt{1+4D\eta_2})/2\eta_2$, which is reached at $D=D_{\rm bp}<1$. When the external force is less than the threshold value $F_{\rm cr}$ determined by Eq. (12) at $D=D_{\rm bp}$, the polymer resides in the hairpinned phase with zero extension, and the free energy density equals to $\phi(F)=(1/\beta)\ln\zeta_{\rm bp}$ and is force-independent. At $F=F_{\rm cr}$ there is a second-order continuous hairpin-coil phase-transition (because $d\zeta/dD=0$ at $D_{\rm bp}$), and the free energy density is changed to $\phi(F)=(1/\beta)\ln(\zeta_{\rm pole})$.

Case B: the stacking potential J so large that $g\gg 1$. In this case, Eq. (10) indicates that $\eta_2=D\eta_1$ and Eq. (8) is equivalent to

$$(\zeta+1/\sqrt{\eta_1})(\zeta-D)(\zeta-1/\sqrt{\eta_1})=0. \tag{13}$$

We readily see that when $D\leqslant 1/\sqrt{\eta_1}$ the smallest positive root of this equation is $\zeta=D$; and when $D>1/\sqrt{\eta_1}$, the smallest positive solution is a constant $\zeta=1/\sqrt{\eta_1}\propto(g\gamma)^{-1/2}$. Most importantly, we have $d\zeta/dD=1$ as D approaches $1/\sqrt{\eta_1}$ from below. As a consequence, for $F<F_{\rm cr}$ which is determined by Eq. (12) with $D=1/\sqrt{\eta_1}$, the polymer is in the hairpinned state and the free energy density is $\phi(F)=-(\frac{1}{2}\beta)\ln\eta_1\propto-(\frac{1}{2}\beta)\ln(g\gamma)$ (again independent of F). For $F>F_{\rm cr}$ the system is in the random coil state, and $\phi(F)=(1/\beta)\ln(\zeta_{\rm pole})$. Here $\zeta_{\rm pole}$ is determined by Eq. (12) (with $D=\zeta_{\rm pole}$). At $F=F_{\rm cr}$ there is a *first-order* hairpin-coil phase-transition, resulted from the fact that $d\zeta/dD|_{D=1/\sqrt{\eta_1}}=1$.

Comparing case A and case B, we have the impression that the inclusion of base-pair stacking interaction may dramatically change the statistical property of the ssDNA system, even the order of the hairpin-coil phase transition. This is understandable qualitatively. The stacking interaction has two effects: (1) It makes base-pairing even more favorable; and (2) it causes the formed base pairs to aggregate into large stacked blocks (see Ref. 24). Since the order of the hairpin-coil transition is related to the intensity of the stacking potential, the hairpin macro-state at low stacking intensity is anticipated to be different to the hairpin macro-state at high stacking intensity, and a continuous phase transition between these two hairpin macro-states can be predicted as the average base-pair stacking intensity changes. This insight is confirmed by observing how the order parameter of this system changes with stacking potential, as shown in Fig. 1. Later we will refer to the hairpin macro-state at low stacking intensity as hairpin-I and that at high stacking intensity as hairpin-II.

In the general case, the average number of base-pairs (in units of $N/2$) is calculated according to $N_{\rm bp}=-2\partial\phi/\partial\ln\gamma$, and the average number of stacked base pairs (also in units of $N/2$) is $N_{\rm sbp}=-2\partial\phi/\partial\ln g$, and the relative extension of each ssDNA bond along the direction of the external force is obtained by $Ex=-\partial\phi/\partial F$.

We demonstrate in Fig. 1 how the values of $N_{\rm bp}$, $N_{\rm sbp}$, and the ratio $N_{\rm sbp}/N_{\rm bp}$ change with strength of base-pair stacking potential. In this figure γ is fixed to 1.9 (just serves as an example) and the external force is set to zero. We see that as the stacking potential J increases, all these three quantities increase and they approach 1 as J reaches about $J_{\rm cr}\simeq 6k_BT$. The ratio $N_{\rm sbp}/N_{\rm bp}$ could serve as an order parameter. Figure 1 shows that when stacking potential is high, almost all the nucleotide bases are paired and stacked; lowering stacking potential from $J>J_{\rm cr}$ to $J<J_{\rm cr}$ there is a continuous phase transition where such a highly stacked configuration is gradually melted out and a growing fraction of bases becomes unpaired and unstacked. This is a structural transition between a highly stacked hairpin macro-state (hairpin-II) and a loosely stacked (or irregular) hairpin macro-state (hairpin-I). This hairpin I–II transition is induced by changing the average effective base-pair stacking interaction in the polymer. Giving a polynucleotide chain, macroscopically the effective base-pairing interaction is determined mainly by its base composition, while the effective stacking potential is determined mainly by the particular arrangement of the bases along the chain. Hence, for different polynucleotides with the same base composition, the folded hairpin configurations may be quite different and fall into two gross catalogue, dependent on their particular base sequence arrangement. This is consistent with the observation that certain RNA chains have unique stable native configurations. The transition between hairpins-I and -II can also be induced by temperature changes or variations in solution ionic concentrations.

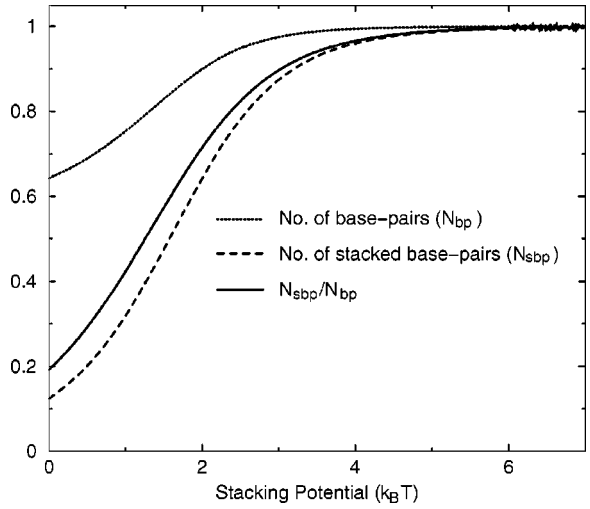

FIG. 1. The relationship between the order parameter $N_{\rm sbp}/N_{\rm bp}$ and the effective stacking potential. Here $N_{\rm bp}$ is the total number of paired bases (in units of $N/2$), and $N_{\rm sbp}$ is the total number of stacked base pairs (also in units of $N/2$). In the numerical calculation of this figure, γ (it accounts for the pairing intensity) is set to 1.9 and external force $F=0$.

It should be mentioned that the critical stacking potential $J_{\rm cr}$ is not sensitive to the pairing potential. This can be seen from the fact that the hairpin I–II transition occurs when the second term in the square brackets of Eq. (10) approaches zero, a condition which is solely satisfied when the stacking potential (and hence g) becomes large.

The force-induced hairpin-coil transition is then second-order or first-order depending on whether the hairpinned configuration is type I or type II. In Fig. 2 the force vs ex-

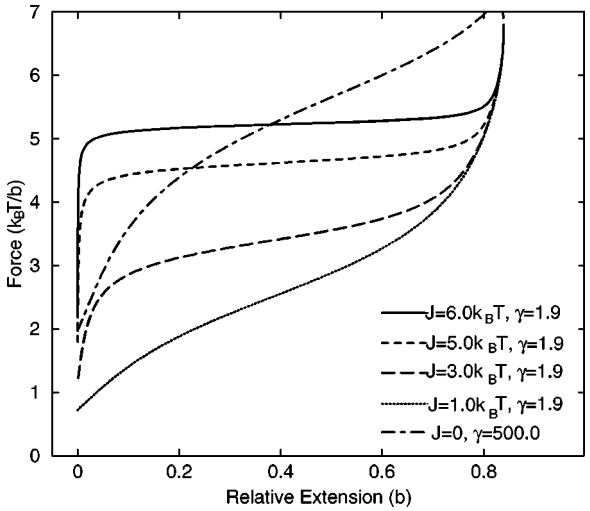

FIG. 2. The relationship between the average end-to-end extension of a polynucleotide and the applied force at different effective base-pairing and base-pair stacking intensities. Here b denotes the Kuhn length of a ssDNA segment, γ accounts for the intensity of the pairing potential between nucleotide bases, and J is the effective stacking potential between nucleotide base pairs.

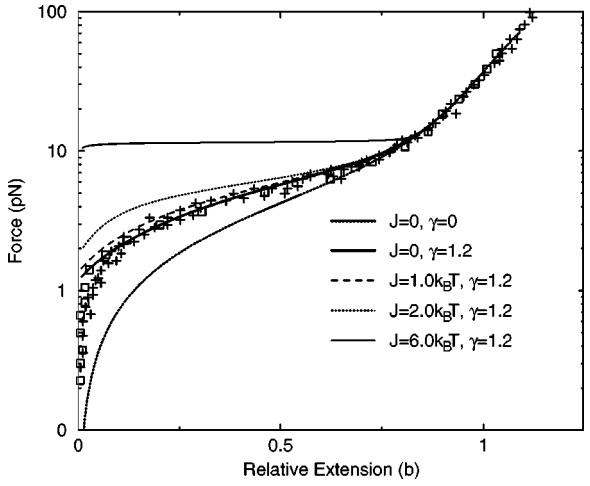

FIG. 4. Force vs extension curves for a ssDNA polymer with its bases randomly ordered. The experimental data come from Refs. 1 (pluses) and 12 (squares), and the ionic concentration is 5 mM Mg^{2+}. The theoretical calculations are performed with $b=1.7$ nm and $l=1.105$ Å, and the other two fitting parameters (γ and J) are listed in the figure.

tension relationship for ssDNA at different pairing and stacking interaction intensities are shown. As is expected, at large stacking intensity, there is a broad force-plateau and this force-plateau disappears as stacking potential is lowered. It is striking to notice that only the inclusion of base pair stacking interaction can lead to the appearance of a highly cooperative hairpin-coil transition. As shown also in Fig. 2 (the dot–dashed curve), if $J=0$ the force-extension curve is always continuous and gradual no matter how strong the base-pairing interaction is.

These theoretical predictions are summarized in the qualitative phase-diagram depicted in Fig. 3.

IV. QUANTITATIVE COMPARISON WITH EXPERIMENTAL OBSERVATIONS

In the last section we have analyzed the main qualitative predictions of the present polynucleotide elastic model. In this section we apply these results to attain a quantitative understanding of some reported experimental findings.

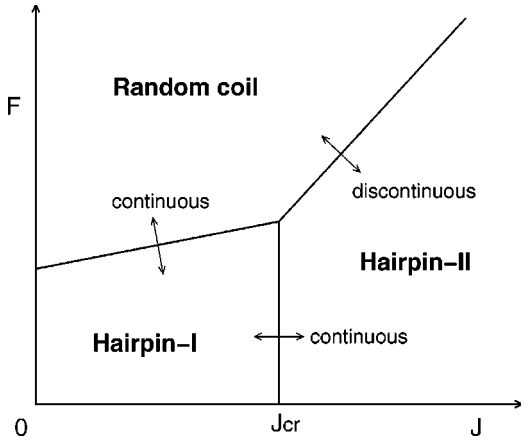

FIG. 3. The qualitative phase-diagram of a polynucleotide at high-salt conditions. The phases hairpin-I and hairpin-II are defined in the main text. F denotes the intensity of the external force.

Bustamante *et al.* and others[1,11,12] have pulled a plasmid ssDNA fragment of 10.4 kilo-bases under different ionic concentrations. They found that the elastic response of ssDNA deviates from that of a random coil both at high and at low ionic conditions. It was suspected that in high salt environments there is partial hairpin formation in the ssDNA chain.[1,11] As is consistent with this insight, Fig. 4 demonstrates that at high salt conditions, the ssDNA made of such a relatively random sequence can be well modeled by the present model with an effective base-pairing interaction characterized by the dimensionless quantity $\gamma=1.2$ and Kuhn length $b=1.7$ nm and bond length variation $l=0.065b=1.105$ Å. The effective base pair stacking interaction for random sequences seems to be quite small and is set to $J=0$ in the fitting (the thick solid line in Fig. 4). Such a comparison has already been performed by Montanari and Mézard earlier.[14] For random sequences the reason that the effective stacking interaction is negligible may be explained as follows: In the polynucleotide chain whose bases arrange randomly, the formed base pairs are usually separated from each other and the possibility for two base pairs to be adjacent to each other is quite small.

Even when the effective stacking potential J between two consecutive node pairs becomes comparable with thermal energy k_BT (the thin dashed line in Fig. 4), the gradual force-extension pattern is only slightly changed. It indicates that the stacking interaction still plays a less important role than the pairing interaction. But the fitting deteriorates much for $J\geqslant 2k_BT$ if the pairing potential is kept constant (thin dotted line): For example, at $J=6K_BT$ (thin solid line) there is a wide plateau in the theoretical force-extension curve.

In another experiment, Rief *et al.*[15] synthesized single-stranded poly(dG-dC) and poly(dA-dT) DNA chains and investigated their elastic responses under external force field. They observed a highly cooperative elongation process in

both these two kinds of sequences, with the transition force being 20 pN for poly(dG-dC) and 9 pN for poly(dA-dT). Figures 5(A) and Fig. 5(B) shows the experimental records as well as the theoretical fittings based on the present model.

In the case of poly(dG-dC), the theoretical curve is obtained by calculating the total extension of a polymer connected by a dsDNA segment of 290.0 nm (with persistence length 53.0 nm and stretch modulus $S=1000.0$ pN as determined by previous experiments[1]) and a ssDNA segment of 230.0 nm (with Kuhn length 1.7 nm and bond length variance 1.105 Å as determined by the data of Fig. 4). (We have included a segment of dsDNA simply because the experiment in Ref. 15 was performed by inserting a ssDNA segment between two dsDNA segments). In the hairpinned state at zero force, the average free energy per Kuhn length is thus determined by the transition force to be $\phi=-(\frac{1}{2}\beta)\ln(\eta_1)=-5.59k_BT$. Similarly, in the case of poly(dA-dT), the fitting is obtained with a dsDNA segment of 80.0 nm and a ssDNA segment of 445.0 nm ($b=1.7$ nm, $l=1.53$ Å) with the average free energy per Kuhn length being $\phi=-1.763k_BT$.

Assuming each Kuhn length contains three nucleotide bases (i.e., each base in the model corresponds to three realistic nucleotides), we can infer that the free energy per base pair in the poly(dG-dC) chain is $3.72k_BT$, while in the poly(dA-dT) chain is $1.18k_BT$. These values are close to what we estimated earlier[17] and are close to the phenomenological parameters $E_{\mathrm{G-C}}\sim 3.0k_BT$and $E_{\mathrm{A-T}}=1.3k_BT$chosen by Bochelmann *et al.*[25] to interpret their experimental results of separating the two strands of a dsDNA apart by mechanical force.

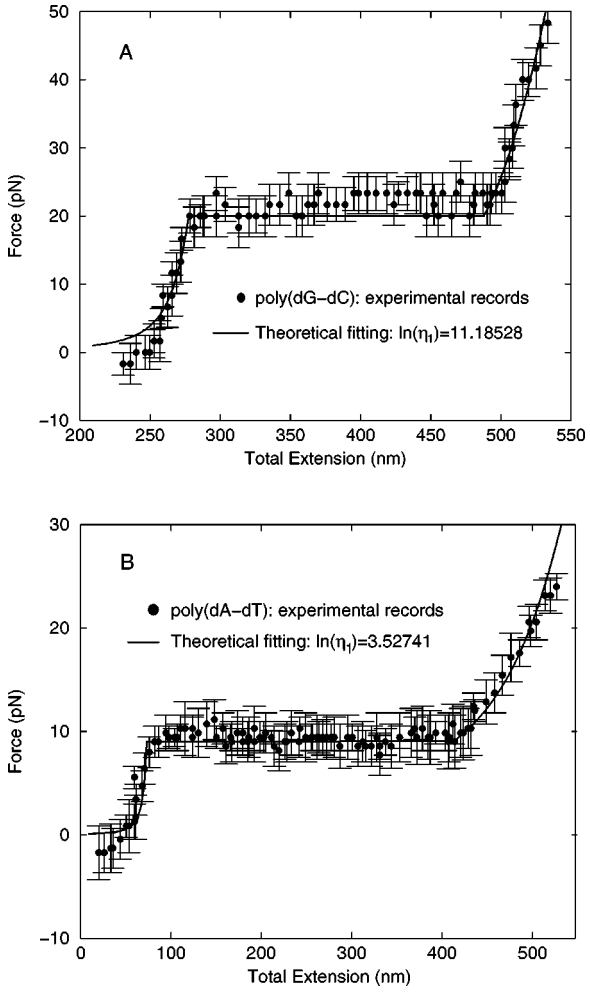

FIG. 5. Force vs extension curves for (A) poly(dG-dC) and (B) poly(dA-dT) ssDNA chains. The experimental data are from Ref. 15 and the theoretical fittings are done with $b=1.7$ nm, and $l=1.105$ Å in (A) and $l=1.530$ Å in (B). The parameter $\ln\eta_1$ [see Eq. (9)] is determined by the transition force, and is set to 11.185 28 and 3.527 41, respectively. To account for the low-extension data a segment of double-stranded DNA is included in the fitting (since the experiment was performed by inserting a ssDNA segment in between two dsDNA segments).

V. CONCLUSIVE REMARKS

In this paper, we have presented an elastic model for single-stranded DNA and RNA polymers, where both the base-pairing and base-pair stacking interactions were considered. The theoretical calculations demonstrated that, depending on the intensity of base-pair stacking potential, ssDNA or RNA can form two kinds of hairpinned structures, hairpin-I and hairpin-II, with the base-pairs in hairpin-I only slightly stacked and those in hairpin-II almost completely stacked. The force-induced hairpin-coil transition is a second-order process if the polymer was originally in hairpin-I macro-state and a first-order process if it was in hairpin-II macro-state. The phase diagram and the force-extension relationship for this polymer system were obtained and the theoretical results achieved good agreement with experimental observations. This work indicates the significance of base-pair stacking interaction to the structural property of polynucleotide chains. The existence of such a hairpin-II macro-state in a designed polynucleotide also justified our previous phenomenological work on the same system by considering only single-looped hairpin structures.[17]

The present work has the following implications. Given a polynucleotide chain, if the nucleotide bases are randomly arranged, then the possibility to forming stacked patterns of base-pairs is very small, since even forming the smallest stack of two base-pairs requires the correlations among four nucleotide bases. Therefore, in random polynucleotide chains, stacking potential is not important. The chains can only form loose hairpins (hairpin-I), and structural fluctuations between these hairpins may be large and quick, making it difficult for the existence of any well-defined stable native configurations. However, things may be dramatically changed if the polynucleotide chain composed of the same bases as the random chain rearranges these nucleotide bases carefully. If the bases in the chain is arranged in such a way that forming of base pairs results in stacking of base pairs, then it is very likely that the polymer will fold into certain stable native structures which are much lower in structural energy than other configurations. Is this part of the reason why some polynucleotide sequences are folded in a particular manner (such as transfer RNA or t-RNA[13])? The nucleotide sequences in RNA molecules of biological cells are the results of millions of years of natural evolution and selection, and therefore in some sense, they should all be well designed.

Finally, we just mention that the theoretical method used

in this work to count for the long-range interactions in ssDNA may be applied to investigate the tertiary structures (supercoils) in double-stranded DNA caused by the topological constraint of fixed linking numbers. The topological constraint leads to a long-range interaction along the dsDNA chain and make the polymer to fold into compacted plectonemic structures.[26] Similar force-extension profiles as the theoretical curves in Fig. 4 have been observed.[27]

ACKNOWLEDGMENTS

We are grateful to Professor Z.-C. Ou-Yang. One of the authors (H.Z.) appreciates a helpful correspondence with A. Montanari, the helps of U. Bastolla and Jian-Jun Zhou, as well as the financial support of the Alexander von Humboldt Foundation. H.Z. is also grateful to Professor R. Lipowsky for a critical reading of the manuscript and comments.

[1] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, Curr. Opin. Struct. Biol. **10**, 279 (2000).
[2] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, Science **265**, 1599 (1994).
[3] J. Marko and E. Siggia, Macromolecules **28**, 8759 (1995).
[4] J. D. Moroz and P. Nelson, Proc. R. Soc. London, Ser. A **94**, 14 418 (1997).
[5] T. R. Strick, J. F. Allemand, D. Bensimon, and V. Croquette, Biophys. J. **74**, 2016 (1998).
[6] C. Bouchiat and M. Mézard, Phys. Rev. Lett. **80**, 1556 (1998).
[7] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **82**, 4560 (1999).
[8] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. E **62**, 1045 (2000).
[9] Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **78**, 1979 (2000).
[10] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[11] G. L. Wuite, S. B. Smith, M. Young, D. Keller, and C. Bustamante, Nature (London) **404**, 103 (2000).
[12] B. Maier, D. Bensimon, and V. Croquette, Proc. Natl. Acad. Sci. USA **97**, 12 002 (2000).
[13] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner, *Molecular Biology of the Gene*, 4th ed. (Benjamin/Cummings, Menlo Park, CA, 1987).
[14] A. Montanari and M. Mézard, e-print cond-mat/0006166; Phys. Rev. Lett. **86**, 2178 (2001).
[15] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Nat. Struct. Biol. **6**, 346 (1999).
[16] P. J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[17] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **86**, 356 (2001).
[18] P. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[19] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[20] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).
[21] Since this work is concerned with the macroscopical elastic property of polynucleotides and each fictitious base here represents an average of several real bases of the polynucleotide, this can be a well justified treatment.
[22] S. Lifson, J. Chem. Phys. **40**, 3705 (1964).
[23] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[24] Y. Zhang, H. Zhou, and Z.-C. Ou-Yang (in preparation).
[25] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997); Phys. Rev. E **58**, 2386 (1998).
[26] J. F. Marko and E. D. Siggia, Phys. Rev. E **52**, 2912 (1995).
[27] T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, Science **271**, 1835 (1996).