THEORY OF STRETCHING INDIVIDUAL POLYNUCLEOTIDE MOLECULE

YANG ZHANG\textsuperscript{1,2*}, HAIJUN ZHOU\textsuperscript{1,3}, AND ZHONG-CAN OU-YANG\textsuperscript{1}

\textsuperscript{1}Institute of Theoretical Physics, The Chinese Academy of Sciences, P. O. Box 2735, Beijing 100080, China
\textsuperscript{2}Laboratory of Computational Genomics Donald Danforth Plant Science Center, 893 North Warson Rd. St Louis, MO 63131, USA
\textsuperscript{3}MPI fuer Kolloid- und Grenzflaechenforschung, Am Muehlenberg, 14476 Golm, Potsdam, Germany
\textsuperscript{*}E-mail: yzhang@danforthcenter.org

Abstract

We review the recent results of experimental and theoretical investigations on the elasticity of both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) molecules. The focus is put on the applications of the path integral method and Monte Carlo technique to the explanation of the experimental measurements of stretching the torsionally relaxed and supercoiled dsDNA, and of pulling ssDNA with random and designed poly(dA-dT)/(dG-dC) nucleotide sequences.

1 Introduction

The elastic properties of polynucleotide molecules of double-stranded DNA and single-stranded RNA/DNA are of vital importance in many basic life processes. They affect, e.g., how DNA packs into chromosomes or serves as a template during the processes of transcription and replication, and how RNA folds into stable native patterns [1]. These mechanical properties have been longly studied experimentally through bulk methods, such as light scattering, sedimentation velocity, and ligase-catalyzed cyclization [2]. These macroscopic measurements represent ensemble averages over all accessible molecular configurations, thereby providing little information on the intermolecular and intramolecular forces that develop in the biopolymers during the course of their biological reactions.

In the past decade, the revolutionary progresses have been brought forth with the development of new micromanipulation techniques combining high force sensitivity [piconewtons (pN)] with accurate positioning (angstroms). These techniques allow researchers to stretch single biological macromolecule from both ends, and directly monitor its time-dependent elastic response under the external force. Since the first single macromolecule stretching experiment performed at 1992, both double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) molecules in different circumstances have been pulled in various laboratories (see, e.g. Refs. [3-9]). Depending on the concerned range of force and elasticity of the molecules, the applied force fields have been imposed through different mediums, for example, magnetic beads (0.01\textasciitilde10 pN), optical tweezers (0.1\textasciitilde100 pN), and atomic force microscopy (10\textasciitilde10,000 pN). Complex behavior has been revealed by these mechanical studies of the polynucleotide molecules. An excellent review of the experimental measurements can be found in Ref. [10].

In this talk, we will focus on the theoretical understanding of the above elasticity measurements of both dsDNA and ssDNA molecules. In our theoretical studies, the dsDNA molecule is modeled as two interwound worm-like chains bound tightly by permanent hydrogen bonds [11]. The structural deformation is characterized by the folding angle between side chains and the central axis. The stacking interaction of base pairs is taken into account by using Lennard-Jones potential with the result of quantum chemical calculations. For ssDNA, we model the molecule as a freely jointed chain (FJC) of elastic bonds with electrostatic, secondary structure (i.e. hairpins formed by the pairing of complementary base of A-T and G-C), and base-stacking interactions all taken into account. The stretching thermodynamics of our DNA models under applied force are calculated through path integral method of partition function and Metropolis Monte Carlo techniques. The computed force-elongation characteristics are in comprehensive agreement with above experimental measurements by different laboratory on both dsDNA and ssDNA in different circumstances.
Elastic properties of double-stranded DNA

2.1 Model of dsDNA

2.1.1 Bending and folding deformations

Natural DNA molecule in living cell is a double-stranded biopolymer, in which two complementary sugar-phosphate chains twist around each other to form a right-handed double helix. Each chain is a linear polynucleotide consisting of the following four bases: two purines (A, G) and two pyrimidines (C, T) [1,12]. The two chains are joined together by hydrogen bonds between pairs of nucleotides A-T and G-C.

As a mathematical model [11,13], the embeddings of two backbones can be defined by \( r_1(s) \) and \( r_2(s') \). The ribbon structure of DNA is enforced by having \( r_2(s') = r_1(s) + 2Rb(s) \) where the hydrogen-bond-director unit vector \( b(s) \) points from \( r_1(s) \) to \( r_2(s') \). As the result of the wormlike backbones, the bending energy of two backbones can be written as \( E_B = \frac{\kappa}{2} \int_0^L \left[ \left( \frac{d^2r}{ds^2} \right)^2 + \left( \frac{d^2r}{ds'^2} \right)^2 \right] ds \) (1)

The formation of basepairs leads to rigid constraints between the two backbones and at the same time they hinder considerably the bending freedom of DNA central axis because of the strong steric effect. In the assumption of permanent hydrogen bonds [14,15], \( |s'-s|=0 \). The relative sliding of backbones is prohibited and the basepair orientation lies perpendicular to the tangent vectors \( t_1 = \frac{dr_1}{ds} \) and \( t_2 = \frac{dr_2}{ds} \) of the two backbones and that of the central axis, \( t \): \( b \cdot t_1 = b \cdot t_2 = b \cdot t = 0 \). By defining the folding angle \( \theta \) as half of the rotation angle from \( t_2(s) \) to \( t_1(s) \), i.e., the intersection angle between tangent vector of backbones \( t_1(2) \) and DNA central axis \( t \), we have

\[
\begin{align*}
\mathbf{t}_1 &= \cos \theta \mathbf{t} + \sin \theta \mathbf{b} \times \mathbf{t} \\
\mathbf{t}_2 &= \cos \theta \mathbf{t} + \sin \theta \mathbf{b} \times \mathbf{t}.
\end{align*}
\]

Therefore, the bending energy of the two backbones can be rewritten as [11]

\( E_B = \int_0^L \left[ \kappa \left( \frac{dt}{ds} \right)^2 + \kappa \left( \frac{d\theta}{ds} \right)^2 + \kappa \frac{\sin^4 \theta}{R^2} \right] d\theta \) (3)

where \( ds \) denotes arc-length element of the backbones, \( L \) the total contour length of each backbone, and \( \kappa \) is the persistence length of one DNA backbone.

Bearing in mind that the pairing and stacking enthalpy of the bases significantly increase bending stiffness of polymer axis, the experimental value of persistent length of dsDNA polymer is considerably larger than that of a DNA single strand [5]. To incorporate the steric effect and also considering the typical experiment value of persistent length of dsDNA \( \rho=53 \) nm [16], we phenomenologically replace the bending rigidity \( \kappa \) in the first term of Eq. (3) with a new parameter \( \kappa^* \). It is required that \( \kappa^*>\kappa \), and the precise value of \( \kappa^* \) will then be determined self-consistently by the best fitting with experimental data (see below).

2.1.2 Base-stacking interactions

The vertical stacking interactions between basepairs originate from the weak van der Waals attraction between the polar groups in adjacent nucleotide basepairs. Such interactions are short-ranged and their total effect is usually described by a potential energy of the Lennard-Jones form (6-12 potential) [12]). In a continuum theory of elasticity, the summed total base-stacking potential energy is converted into the form of the following integration:

\( E_{ij} = \sum_{i=1}^{N_{bp}-1} U_{ij+1} = \int_0^L \rho(\theta) ds \) (4)

where \( U_{ij+1} \) is the base-stacking potential between the \( i \)-th and the \( (i+1) \)-th basepair, \( N_{bp} \) is the total number of basepairs, and the base-stacking energy density \( \rho \) is expressed as
In Eq. (5), the parameter $r_0$ is the backbone arclength between adjacent bases ($r_0=L/N$); $\theta_0$ is a parameter related to the equilibrium distance between a DNA dimer ($r_0\cos \theta_0 \sim 0.34$ nm); and $\epsilon$ is the base-stacking intensity which is generally base-sequence specific. Here we focus on macroscopic properties of long DNA chains composed of relatively random sequences, therefore we just consider $\epsilon$ in the average sense and take it as a constant, with $\epsilon \sim 14.0 \ k_B T$ as averaged over quantum-mechanically calculated results on all the different DNA dimers \cite{12}.

The asymmetric base-stacking potential Eq. (5) ensures a relaxed DNA to take on a right-handed double-helix configuration (i.e., the B-form) with its folding angle $\theta \sim \theta_0$. To deviate the local configuration of DNA considerably from its B-form generally requires a free energy of the order of $\epsilon$ per basepair. Thus, DNA molecule will be very stable under normal physiological conditions and thermal energy can only make it fluctuate very slightly around its equilibrium configuration, since $\epsilon \gg k_B T$. Nevertheless, although the stacking intensity $\epsilon$ in dsDNA is very strong compared with thermal energy, the base-stacking interaction by its nature is short-ranged and hence sensitive to the distance between the adjacent basepairs. If dsDNA chain is stretched by large external forces, which cause the average inter-basepair distance to exceed some threshold value determined intrinsically by the molecule, the restoring force provided by the base-stacking interactions will no longer be able to offset the external forces. Consequently, it will be possible that the B-form configuration of dsDNA will collapse and the chain will turn to be highly extensible. Thus, on one hand, the strong base-stacking interaction ensures the standard B-form configuration to be very stable upon thermal fluctuations and small external forces (this is required for the biological functions of DNA molecule to be properly fulfilled \cite{1}); but on the other hand, its short-rangedness gives it considerable latitude to change its configuration to adapt to possible severe environments (otherwise, the chain may be pulled break by external forces, for example, during DNA segregation \cite{1}). This property of base-stacking interactions is very important in the determination of elasticity of dsDNA and the conformation of secondary structure in single-stranded DNA/RNA molecules, as will shown in following sections.

2.1.3 External forces field

In the previous two subsections, we have described the intrinsic energy of DNA double-helix. Experimentally, to probe the elastic response of linear DNA molecule, the polymer chain is often pulled by external force fields. Here we constrained ourselves to the simplest situation where one terminal of DNA molecule is fixed and the other terminal is pulled with a force $F=fx_0$ along direction of unit vector $z_0$ \cite{3-9}). (In fact, hydrodynamic fields or electric fields are also frequently used to stretch semiflexible polymers \cite{17}, but we will not discuss such cases here.) The end-to-end vector of a DNA chain is expressed as $\int_0^L \hat{t}(s) \cos \theta(s) ds$. Then the total ``potential'' energy of the chain in the external force field is

$$E_f = -\int_0^L \hat{t} \cos \theta ds \cdot F = -\int_0^L \hat{t} \cdot z_0 \cos \theta ds. \quad (6)$$

To conclude this section, the total energy of a dsDNA molecule under the action of an external force is expressed in our model as

$$E = E_b + E_{13} + E_f$$

$$= \int_0^L \left[ \kappa \left( \frac{dt}{ds} \right)^2 + \kappa \left( \frac{d\theta}{ds} \right)^2 + \frac{\kappa}{R^2} \sin \theta \right] ds + \rho(\theta) - \hat{t} \cdot z_0 \cos \theta ds. \quad (7)$$

2.2 Extensibility and entropic elasticity

According to the path integral method of polymer chain (see Appendix A of Ref. \cite{13}), the Green equation of Eq. (7) is
\[
\frac{\partial \Psi(t,\theta; s)}{\partial s} = \frac{\partial^2}{4l_p^2 \partial t^2} + \frac{\partial^2}{4l_p^2 \partial \theta^2} + \frac{f \cos \theta}{k_B T} \cdot \mathbf{t} \cdot \mathbf{z}_0 - \frac{\rho(\theta)}{k_B T} \cdot \frac{l_p}{R^2} \sin^4 \theta |\Psi(t,\theta; s)|, \tag{8}
\]

where \( l_p = \kappa / k_B T \), \( l = \kappa / k_B T \), and \( \Psi(t,\theta; s) \) is an auxiliary function for the configuration of dsDNA system. The spectrum of the above Green equation is discrete and for a long dsDNA molecule its average extension can be obtained either by differentiation of the ground-state eigenvalue, \( g_0 \), of Eq. (7) with respect to \( f \):

\[
\langle Z \rangle = \int_0^L \langle \mathbf{t} \cdot \mathbf{z}_0 \cos \theta \rangle ds = L k_B T \frac{\partial g_0}{\partial f}, \tag{9}
\]

or by a direct integration with the normalized ground-state eigenfunction, \( \Phi_0(t,\theta) \), of Eq. (8):

\[
\langle Z \rangle = L \int \hat{\Phi}_0 \mathbf{t} \cdot \mathbf{z}_0 |\cos \theta| dt d\theta. \tag{10}
\]

Both \( g_0 \) and \( \Phi_0(t,\theta) \) can be obtained numerically through standard diagonalization methods and identical results are obtained by Eqs. (9) and (10).

**Figure 1.** Force-extension relation of torsionally relaxed DNA molecule. Experimental data taken from Ref. [4] (symbols). Theoretical curve obtained by the following considerations: (a) \( l_p = 1.5 \text{ nm} \) and \( \epsilon = 14.0 \ k_B T \); (b) \( l_p = 0.34 / \langle \cos \theta \rangle f = 0 \text{ nm} \), \( \eta_0 = 0.34 / \langle \cos \theta \rangle f = 0 \text{ nm} \) and \( R = (0.34 \times 10.5 / \epsilon |\tan \theta| f = 0 \text{ nm} \); (c) adjust the value of \( \theta_0 \) to fit the data. For each \( \theta_0 \), the value of \( \langle \cos \theta \rangle f = 0 \) is obtained self-consistently. The present curve is drawn with \( \theta_0 = 62.0^\circ \) (in close consistence with the structural property of DNA), and \( \langle \cos \theta \rangle f = 0 \) is determined to be 0.573840. DNA extension is scaled with its B-form contour length \( L \langle \cos \theta \rangle f = 0 \).

In Fig. 1 is shown the calculated force vs extension relation in the whole relevant force range and comparison with the experimental observations [4,5]. The theoretical curve in this figure is obtained with just one adjustable parameter (see caption of Fig. 1); the agreement with experiment is excellent. Figure 1 demonstrates that the highly extensibility of DNA molecule under large external forces can be quantitatively explained by the present model.

To further understand the force-induced extensibility of DNA, in Fig. 2 we show the folding angle distribution of dsDNA molecule at different external force, i.e. [47]

\[
P(\theta) = \int |\Phi_0(t,\theta)|^2 dt. \tag{11}
\]
Figures 1 and 2, taking together, demonstrate that the elastic behaviors of dsDNA molecule are radically different under the condition of low and large applied forces:

The low-force region When external force is low (≤ 10 pN), the folding angle is distributed narrowly around the angle of $\theta \sim 57^\circ$, and there is no probability for the folding angle to take on values less than $0^\circ$ (Fig. 3), indicating that DNA chain is completely in the right-handed B-form configuration with small axial fluctuations. This should be attributed to the strong base-stacking intensity, as pointed out above. Consequently, the elasticity of DNA is solely caused by thermal fluctuations in the axial tangent $t$ (Fig. 1), and DNA molecule can be regarded as an inextensible chain. This is the physical reason why, in this force region, the elastic behavior of DNA can be well described by the wormlike chain model [16,18]. Indeed, as shown in Fig. 3, at forces ≤ 10pN, the wormlike chain model and the present model give identical results. Thus, we can conclude with confidence that, when external fields are not strong, the wormlike chain model is a good approximation of the present model to describe the elastic property of dsDNA molecules.

The large-force region With the continuous increase of external pulling forces, the axial fluctuations becomes more and more significant. For example, at forces ~ 50pN, although the folding angle distribution is still peaked at $\theta \sim 57^\circ$, there is also considerable probability for the folding angle to be distributed in the region $\theta \sim 0^\circ$ (Fig. 2). Therefore, at this force region, DNA polymer can no longer be regarded as inextensible. At $f \sim 65pN$, another peak in the folding angle distribution begins to emerge at $\theta \sim 0^\circ$, marking the onset of cooperative transition from B-form DNA to overstretched S-form DNA [4,5]. This is closely related to the short-ranged nature of the base-stacking interactions [12]. At even higher forces ($f \geq 80pN$), the DNA molecule becomes completely into the overstretched form with its folding angle peaked at $\theta \sim 0^\circ$. This threshold $f_i$ of over-stretch force is also consistent with a plain evaluation from base-stacking potential of $\varepsilon \sim f_t r_0$, i.e. $f_i \sim 90pN$. 

![Figure 2. Folding angle distributions for negatively supercoiled DNA molecule pulled with a force of 1.3 pN.](image)
Figure 3. Low-force elastic behavior of DNA. Here experimental data is from Fig. 5B of [3], the dotted curve is obtained for a wormlike chain with bending persistence length 53.0 nm and the parameters for the solid curve are the same as those in Fig. 1.

It should be mentioned that in the experiments [4,5] the transition to S-DNA occurs even more cooperatively and abruptly than predicted by the present theory (see Fig. 1). This may be related to the existence of single-stranded breaks (nicks) in the dsDNA molecules used in the experiments. Nicks in DNA backbones can lead to strand-separation or relative sliding of backbones [4,5], and they can make the transition process more cooperative. However, the comprehensive agreement achieved in Figs. 1 and 3 indicates that such effects are only of limited significance. The elasticity of DNA is mainly determined by the competition between folding angle fluctuation and tangential fluctuation, which are governed, respectively, by the base-stacking interactions ($\epsilon$) and the axial bending rigidity ($\kappa'$) in Eq. (7).

## 2.3 Elongation of supercoiled DNA

The number of times the two strands of DNA double helix are interwound, i.e., the link number $L_k$, is a topologic invariant quantity for closed DNA molecule. It is also topologically invariant for linear DNA polymer in case that the orientations of two extremities of the linear polymer are fixed and any part of polymer is forbidden to go round the extremities of the polymer (as performed in Strick et al's supercoiling experiments [6]). An unstressed B-DNA molecule has one right-handed twist per 3.4nm along its length, i.e., $L_k=3L_B/3.4$. Under some twist stress, the link number of DNA polymer may deviate from its torsionally relaxed value [20]. In all case when $\Delta L_k=|L_k-L_k^0|\neq 0$, the DNA polymer is called “supercoiled”. The relative difference in link number, $\sigma=(L_k-L_k^0)/L_k^0$, signifies the degree of supercoiling which is independent upon the length of DNA polymer. In this section we investigate the elasticity of supercoiled DNA double-helix through Monte Carlo simulation, based on the same model but in a discrete form [21].

### 2.3.1 Discrete dsDNA model

In the simulation, the double-stranded DNA molecule is modeled as a chain of discrete cylinders, or two discrete wormlike chains constrained by basepairs of fixed length $2R$. The conformation of the chain is specified by the space positions of vertices of its central axis, $\mathbf{r}_i=(x_i,y_i,z_i)$ in 3-D Cartesian coordinate system, and the folding angle of the sugar-phosphate backbones around the central axis, $\theta_i$, $i=1,2,\ldots,N$. Each segment is assigned the same amount of basepairs, $n_{bp}$, so that the length of the $i$th segment satisfies $\Delta s_i = |\mathbf{r}_i-\mathbf{r}_{i-1}| = 0.34n_{bp}\langle \cos \theta_i \rangle_{\tau_0}$, where $\langle \cdot \rangle_{\tau_0}$ means the thermal average for a relaxed DNA molecule.

According to Eq. (7), the total energy of dsDNA molecule with $N$ segments in our discrete computery model is expressed as
\[
E = \alpha \sum_{i=1}^{N-1} \gamma_i^2 + \alpha' \sum_{i=1}^{N-1} (\theta_{i+1} - \theta_i)^2 + \frac{k}{R^2} \sum_{i=1}^{N} \Delta \theta_i \sin^3 \theta_i \tan \theta_i + \sum_{j=1}^{N_{ex}} U(\theta_j) - f z_N, \tag{12}
\]

where \( \gamma_i \) is the bending angle between the \((i-1)\)th and the \(i\)th segments, and \( z_N \) is the total extension of the DNA central axis along the direction of the external force \( f \) (assumed in the \( z \)-direction). Here, the bending rigidity constant \( \alpha \) corresponds to the persistent length \( p=53 \) nm of dsDNA according to the direct discretization of Eq. (1) or (3), i.e.

\[
\alpha = \frac{p}{2b} k_B T, \tag{13}
\]

where \( b \) is the average length of segments.

In earlier approaches [21, 22], the bending rigidity constant \( \alpha \) of discrete chain was determined according to the Kuhn statistic length of wormlike chain, which is twice of the persistent length \( p \). In fact, the Kuhn length \( l_{\text{kuhn}} \) of discrete wormlike chain with rigidity \( \alpha \) is written as (see, e.g. [21])

\[
l_{\text{kuhn}} = b \frac{1 + \langle \cos \gamma \rangle}{1 - \langle \cos \gamma \rangle}, \tag{14}
\]

where

\[
\langle \cos \gamma \rangle = \frac{\int_0^\pi \cos \gamma \exp(-\alpha \gamma^2 / k_B T) \sin \gamma d\gamma}{\int_0^\pi \exp(-\alpha \gamma^2 / k_B T) \sin \gamma d\gamma}. \tag{15}
\]

The rigidity constant \( \alpha \) defined in this way is only the function of \( m=l_{\text{kuhn}}/b \), the number of links within one Kuhn length. The dependence of \( \alpha \) versus \( m \) obtained from Eqs. (14) and (15) is shown numerically in Fig. (4). The rigidity constant \( \alpha \) follows very well the linear dependence upon \( m \), especially in the reasonable region of \( m>5 \). As a comparison, we also show the line of Eq. (13), i.e. \( \alpha=(m/4) k_B T \). It is obvious that the rigidity constants of discrete chain are numerically equivalent in two algorithms. However the processes of discretization of Eq. (3) is more convenient to be used, especially to the models with complicated elasticity [46].

**Figure 4.** The bending rigidity constant \( \alpha \) of discrete wormlike chain as function of number of segments within one Kuhn length \( m=l_{\text{kuhn}}/b \). The circle points denote the result calculated from Eqs. (14) and (15); the solid line from direct discretization of Eq. (3).

The constant \( \alpha' \) in the second term of Eq. (12) should be associated with stiffness of the DNA single strand. As an crude approximation, we have taken here \( \alpha' = \alpha \). Our unpublished data show that, the amount of second term of Eq. (12) is quite small compared with other four terms. And the result of simulation is not sensitive to \( \alpha' \).
2.3.2 Simulation procedure

The equilibrium sets of conformations of dsDNA chain are constructed using the Metropolis MC procedure [23]. The conformational space is sampled through a Markov chain process. Three kinds of movements are considered, i.e. (1) the length of a randomly chosen segment is modified; (2) a portion of the chain is rotated around the axis connecting the two ends of rotated chain; (3) The segments from a randomly chosen vertex to the free end are rotated around an arbitrary orientation axis that passes the chosen vertex [21]. A trial move from a conformation (or state) $i$ to a conformation (or state) $j$ is accepted on the basis of the probability $\rho_{i \rightarrow j} = \min(1, \rho_j / \rho_i)$, where $\rho_i$ is the probability density of conformation $i$. Energetic importance sampling is realized in the Metropolis MC method by choosing the probability density $\rho_i$ as the Boltzmann probability: $\rho_i = \exp(-E_i/k_B T)$, where $E_i$ is the energy of conformation $i$ calculated according to Eq. (12).

The starting conformation of chain is unknotted. To avoid knotted configuration in the Markov process, we calculate the Alexander polynomial of each trial conformation [24,25]. In a case where the trial movement knots the chain, the trial movement will be rejected. To incorporate the exclude-volume effect, for each trial conformation, we calculate the distance between any point on the axis of a segment and any point on the axis of another nonadjacent segment. If the minimum distance for any to chosen segment is less than the DNA diameter $2R$, the energy of trial conformation is set infinite and the movement is then rejected.

During the evolution of DNA chain, the supercoiling degree $\sigma$ may distribute around all the possible values. In order to avoid the waste of computation events and also for the comparison of Strick et al's supercoiling experiments [6,19], we bound the supercoiling $\sigma$ of DNA chain inside the region of $-0.12 \leq \sigma \leq 0.12$. When the torsion degree of trial conformation is beyond the chosen range, we simply neglect the movement and reproduce a new trial movement again.

The linking number $L_k$ of each conformation is calculated according to White's formula [26],

$$L_k = T_w + W_r,$$

where

$$T_w = \frac{1}{2\pi R} \sum_{i=1}^{N} \Delta \theta_i \tan \theta_i.$$  \hspace{1cm} (16)

To enclose the linear DNA molecule without changing its linking number, we add three long flat ribbons to the two ends for each conformation and keep the ribbons in the same planar. The writhing number $W_r$ can therefore be calculated through Gaussian integral

$$W_r = \frac{1}{4\pi} \iint dsds' \left| \mathbf{\hat{r}}(s) \times \mathbf{\hat{r}}(s') - [\mathbf{r}(s) - \mathbf{r}(s')] \right| \left| \mathbf{r}(s) - \mathbf{r}(s') \right|^3.$$ \hspace{1cm} (17)

2.3.3 Elasticity of supercoiled dsDNA

To obtain equilibrium ensemble of DNA evolution, $10^7$ elementary displacements are produced for each chosen applied force $f$. The relative extension $x$ and supercoiling degree $\sigma$ of each accepted conformation of DNA chain are calculated. When the trial movement is rejected, the current conformation is count up twice (see Ref. [23]).
Figure 5. Force versus relative extension curves for negatively (a,b) and positively (c,d) supercoiling DNA molecule. Left two plots (a) and (c) are the results of our Monte Carlo simulation, and the horizontal bars of points denote the statistic error of relative extension in our simulations. Right two plots (b) and (d) are the experimental data from Ref. [19]. The solid curves serve as guides for the eye.

In order to see the dependence of mechanics property of dsDNA upon supercoiling degree, the whole sample is partitioned into 15 subsamples according to the value of the supercoiling degree $\sigma$. For each subsample, we calculate the averaged extension

$$\langle z_N \rangle_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{z_N(i)}{L_B}, \quad j = 1, \cdots, 15$$

and the averaged torsion

$$\langle \sigma \rangle_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \sigma_i, \quad j = 1, \cdots, 15,$$

where $N_j$ is the number of movements supercoiling of which belong to $j$th subsample.

We display the force versus relation extension for all positive and negative supercoiling in Fig. 5a and c respectively. As a comparison, the experimental data [6,19] are shown in Fig. 5b and d. In Fig. 6 is shown the averaged extension as a function of supercoiling degree for 3 typical applied forces. In spite of quantitative difference between Monte Carlo results and experimental data, the qualitative coincidence is striking. Especially, three evident regimes exist in both experimental data and our Monte Carlo simulations:

1) At a low force, the elastic behavior of DNA is symmetrical under positive or negative supercoiling. This is understandable, since the DNA torsion is the cooperative result of hydrogen-bond constrained bending of DNA backbones and the base-stacking interaction in our model. At very low force, the contribution from applied force and the thermodynamic fluctuation perturbs the folding angle $\theta$ of basepair to derive very little from the equilibrium position $\theta_0$. Therefore, the DNA elasticity is achiral at this region. For a fixed applied force, the
increasing torsion stress tends to produce plectonemic state which shortens the distance of two ends, therefore, the relative extension of linear DNA polymer. These features can be also understood by the traditional approaches with harmonic twist and bending elasticity [22,27].

2) At intermediate force, the folding angle of basepairs are pulled slightly further away from equilibrium value $\theta_0$ where van der Waals potential is not symmetric around $\theta_0$. So the chiral nature of elasticity of the DNA molecule appears. In negative supercoiling region, i.e. $\theta<\theta_0$, the contribution of applied force dominates that of potential because of the low plateaus of $U(\theta)$ (see Eq. (5)). So the extension is insensitive to negative supercoiling degree. On the other hand, the positive supercoiling still tends to contract the molecule.

3) At higher force, the contribution of the applied force to the energy dominates that of van der Waals potential in both over- and underwound DNA. The extension of DNA accesses to its B-form length. Therefore, the plectonemic DNA is fully converted to extended DNA, the writhe is essentially entirely converted to twist and the force-extension behavior reverts to that of untwisted $\sigma=0$ DNA as expected from a torsionless worm-like chain model [3,11,18].

![Figure 6](image)

**Figure 6.** Relative extension versus supercoiling degree of DNA polymer for three typical stretch forces. Open points denote the experimental data[19], and solid points the results of our Monte Carlo simulation. The vertical bars of the solid points signify the statistic error of the simulations, and the horizontal ones denote the bin-width that we partition the phase space of supercoiling degree. The solid lines connect the solid points to guide the eye.

It should be mentioned that there is an up-limit of supercoiling degree for extended DNA in current approach, i.e. $\sigma_{\text{max}} \sim 0.14$, which corresponds to $\theta=90^\circ$ of folding angle. In recent experiments, Allemand et al. [28] twisted the plasmid up to the range of $-5<\sigma<3$. They found that at this `unrealistically high' supercoiling, the curves of force versus extension for different $\sigma$ split again at higher stretch force (>3 pN). As argued by Allemand et al., in the extremely under- and overwound torsion stress, two new DNA forms, denatured-DNA and P-DNA with exposed bases, will appear. In fact, if the deviation of the angle which specifies DNA twist from its equilibrium value exceeds some threshold, the corresponding torsional stress causes local distraction of the regular double helix structure. So the emergence of these two striking forms is essentially associated with the broken processes of some basepairs under super-highly torsional stress. In this case, the permanent hydrogen constrain will be violated and the configuration of base stacking interactions be varied considerably.

3 **Mechanics of pulling single-stranded DNA**

By attaching dsDNA between beads and melting off the unlabeled strand with distilled water or formaldehyde, a single stranded can be obtained [10]. Because of its thin diameter and high flexibility, ssDNA is more contractile than dsDNA in low force. However, it can be stretched to a greater length at
high force since it no longer forms a helix. In 150 mM NaCl solution, the force/extension curve of ssDNA, melted from a λ phage DNA molecule, was found to be able to be fit with a simple freely jointed chain (FJC) of Kuhn length of 1.5 nm with including a stretch modulus [5]. However, more detailed measurements showed that the elongation characteristics of ssDNA is very sensitive to the ionic concentration of solution, and the FJC is not valid in both high ionic (e.g. 5 mM MgCl₂) and low ionic (e.g. 2 mM NaCl) solutions [10,9]. On the other hand, the measurements by Rief et al. [7] shows that the force/extension characteristics of ssDNA is strongly sequence-dependent. When a single designed poly(dA-dT) or poly(dG-dC) strand is pulled with an atomic force microscope, they found that, at some stretched force [9 pN for poly(dA-dT) and 20 pN for poly(dG-dC)], the distance of two ends of the designed molecules suddenly elongates from nearly zero to a value comparable to its total contour length in a very cooperative manner, which is drastically different from the gradual elongation of the nature ssDNA (within a relative random sequence).

Here, we present our recent Monte Carlo calculations of a modified freely jointed chain with elastic bonds. In order to attain an unified understanding of reported force/extension data of ssDNA molecule in different ionic atmospheres and for different nucleotide sequences, we have incorporated three possible interactions of base-pairing [29], stacking [30] and electrostatic interactions in our calculations. In the next section, we at first determine the electrostatic potential between DNA strands through numerically solving the nonlinear Poisson-Boltzmann equation.

3.1 Electrostatic interaction between ssDNA

Under the assumptions of (1) the solute in a solution of strong electrolyte is completely dissociated into ions; (2) all deviations from the properties of an ideal solution (ions are uniformly distributed) are due to the electrostatic forces which exist between the ions, the electrostatic potential \( \psi(\mathbf{r}) \) at a space point \( \mathbf{r} \) can be submitted to the Poisson-Boltzmann equation [32]:

\[
\nabla^2 \psi(\mathbf{r}) = -\frac{4\pi}{D} \sum_{i=1}^{n} v_i c_i \exp\left(-v_i e \psi(\mathbf{r}) / k_B T \right).
\]

Here the solution is assumed to contain \( N_1, \ldots, N_n \) different ions with valences \( v_1, \ldots, v_n \), and \( c_i (=N_i/V) \) is the bulk concentration of the ionic species \( i \), where \( V \) and \( D \) are the volume and dielectric constant of the solution, and \( e \) is the protonic charge.

Equation (20) cannot be solved in closed form. Here, we calculate the electrostatic potential of ssDNA cylinder immersed in the solutions of NaCl and MgCl₂, through numerically solving Eq. (20) according to the series expansion method used earliest by Pierce [31,32]. As illustrative examples, we show in Fig. 7 the electrostatic potential of ssDNA cylinder in 2 mM NaCl and 5 mM MgCl₂ solutions, where the potential function is expanded up to 17th order for symmetrical electrolyte (NaCl) and 14th order for unsymmetrical electrolyte (MgCl₂) in our calculations.

**Figure 7.** Electrostatic potential of ssDNA cylinder versus the radial distance from the cylinder axis in the solutions of 2 mM NaCl (black) and 5 mM MgCl₂ (grey). The solid curves are the numerical solutions of Poisson-Boltzmann
(P-B) up to the expansion $17^{th}$ order for NaCl and $14^{th}$ order for MgCl$_2$; the dashed curves denote the Debye-Hückel approximation (D-H) with effective linear charge density $\nu$ along the axis listed in Table I. The dotted-line corresponds to the surface of ssDNA cylinder of $r_0=0.5$ nm.

However, the numerical solution of straight charged cylinder of Poisson-Boltzmann equation can not be directly used in the calculations of ssDNA molecule, since the real molecule actually takes a variety of irregular configurations. To approach the problem, we consider the first-order approximation of Equation (20), i.e. linear Poisson-Boltzmann equation, the solution of which can be implicitly expressed. Around a point charge $q$, the electrostatic potential in the linear equation can be written in Debye-Hückel form as

$$\psi_{D-H}(r) = \frac{q}{D|r|} \exp(-\kappa \cdot |r|),$$

where $r$ is the position vector from $q$, and the inverse Debye length $\kappa = \left(\frac{8\pi \epsilon_0 e^2}{Dk_B T}\right)^{1/2}$ for NaCl solution, and $\kappa = \left(\frac{24\pi \epsilon_0 e^2}{Dk_B T}\right)^{1/2}$ for MgCl$_2$.

In order to count the influence of higher expansion terms of Poisson-Boltzmann equation, one can phenomenologically change the amplitude of the Debye-Hückel potential of Eq. (21) to match the numerical solution of Poisson-Boltzmann equation according to Brenner and Parsegian [33] and Stigter [34]. According to Eq. (21), the electrostatic potential of a straight charged cylinder of infinite length can be written as

Figure 8. The effective linear charge density $\nu$ of both ssDNA and dsDNA as function of ionic concentrations of NaCl and MgCl$_2$ solutions. The solid circles are the results in present calculations; The opened circles are Stigter's results [34], where electrophoretic charge of $-0.73e$ were used, which is required to fit Stigter's electrophoresis theory to experimental data. In present calculations, the full charge per phosphate group is assumed. The curve is a fit of equation (23) with fit parameters listed in Table I.
\[
\psi_{\text{DH}}(r) = \int_{-\infty}^{\infty} \frac{\nu d\lambda}{D} \exp(-\kappa \sqrt{\lambda^2 + r^2}) = \frac{2\nu}{D} K_0(\kappa r),
\]
where the integral of \(\lambda\) is along the cylinder axis, \(r\) is the radial distance from cylinder axis, \(\nu\) is the linear charged density, and \(K_0\) is the first-order modified Bessel function. By comparing the Eq. (22) with Poisson-Boltzmann solution in the overlap region far from the cylinder surface, we can determine the effective linear charge density \(\nu\) in different bulk ionic concentrations \(c\) of both NaCl and MgCl\(_2\) (see Table I). In Table I we also show the effective charge density of dsDNA.

As shown in Fig. 8, all the data of \(\nu\) can be very well fitted by the formula of

\[
\nu = \exp(\alpha + \beta \lambda^{2/3}),
\]
with the fit parameters \(\alpha\) and \(\beta\) listed in Table I. As a comparison, Stigter's calculation for dsDNA in NaCl solution, where 73% of electrophoretic charge was assumed [34], is also shown in Fig. 8.

### Table I

<table>
<thead>
<tr>
<th>Ionic Concentration (c_0) (M)</th>
<th>SsDNA NaCl</th>
<th>SsDNA MgCl(_2)</th>
<th>DsDNA NaCl</th>
<th>DsDNA MgCl(_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>4.18</td>
<td>9.50</td>
<td>91.85</td>
<td>993.16</td>
</tr>
<tr>
<td>0.75</td>
<td>3.50</td>
<td>6.74</td>
<td>56.15</td>
<td>410.67</td>
</tr>
<tr>
<td>0.5</td>
<td>2.84</td>
<td>4.51</td>
<td>31.22</td>
<td>144.10</td>
</tr>
<tr>
<td>0.2</td>
<td>2.04</td>
<td>2.31</td>
<td>11.73</td>
<td>24.52</td>
</tr>
<tr>
<td>0.15</td>
<td>1.89</td>
<td>1.97</td>
<td>9.29</td>
<td>16.22</td>
</tr>
<tr>
<td>0.1</td>
<td>1.73</td>
<td>1.64</td>
<td>7.02</td>
<td>9.82</td>
</tr>
<tr>
<td>0.05</td>
<td>1.53</td>
<td>1.27</td>
<td>4.78</td>
<td>4.98</td>
</tr>
<tr>
<td>0.02</td>
<td>1.37</td>
<td>0.99</td>
<td>3.29</td>
<td>2.66</td>
</tr>
<tr>
<td>0.01</td>
<td>1.29</td>
<td>0.86</td>
<td>2.66</td>
<td>1.91</td>
</tr>
<tr>
<td>0.005</td>
<td>1.23</td>
<td>0.78</td>
<td>2.26</td>
<td>1.45</td>
</tr>
<tr>
<td>0.002</td>
<td>1.17</td>
<td>0.71</td>
<td>1.93</td>
<td>1.3</td>
</tr>
<tr>
<td>0.001</td>
<td>1.14</td>
<td>0.67</td>
<td>1.76</td>
<td>1.00</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.0338</td>
<td>-0.577</td>
<td>0.300</td>
<td>-0.505</td>
</tr>
<tr>
<td>(\beta)</td>
<td>1.36</td>
<td>2.80</td>
<td>4.18</td>
<td>7.33</td>
</tr>
</tbody>
</table>


### 3.2 Model and method of calculations

#### 3.2.1 Model of single-stranded DNA

In the simulation, the single-stranded DNA molecule is modeled as a freely jointed chain with \(N\) elastic bonds. The conformation of the chain is specified by the space position of its vertices, \(\mathbf{r}_i=(x_i, y_i, z_i)\), \(i=0, 1, \ldots, N\), in three-dimensional Cartesian coordinate system with \(\mathbf{r}_0\) fixed at the original point. The equilibrium features of stretched ssDNA in salt solution are determined by the interplay of following five energies within the frame of canonical Boltzmann statistical mechanics.

The first energy, called \(E_{\text{ele}}\), is the electrostatic interaction energy between strands. As discussed in the last section, the electrostatic energy of ssDNA molecule can be calculated according to Debye-Hükel approximation

\[
\frac{E_{\text{ele}}}{k_B T} = \frac{\nu^2}{k_B T D} \int d\lambda_i \int d\lambda_j \exp(-\kappa |\mathbf{r}_i - \mathbf{r}_j|) \frac{\exp(-\kappa |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|}
\]
where the effective charged density \(\nu\) is taken from Table I, the integration done along the strand, and \(|\mathbf{r}_i - \mathbf{r}_j|\) is the distance between the current positions at the strand to which the integration parameters \(\lambda_i\) and \(\lambda_j\) correspond.

The second energy describe pairing interactions of complementarity bases in ssDNA. Two elementary pairings in ssDNA are the G-C and A-T base pairs of Watson and Crick. As a G-C pair is
formed through three hydrogen bonds while an A-T one involves only two hydrogen bonds, the basepairing potential should be sequence dependent. Since the Kuhn length of ssDNA is rather longer than its sugar-phosphate backbone between two bases, each node in our model actually includes several bases. We approximate the basepairing interaction by the node-pairing energy of $E_{\text{pair}} = \sum_{i=1}^{N_p} V_p$, where $N_p$ is the number of node-pairs and $V_p$ is a sequence-dependent parameter. The pairing rule in our simulation is similar with the standard one which keeps only nested structure [35,36,29,37,30]: (a) Two nodes $(i,j)$ can be paired only when their distance $|\mathbf{r}_i - \mathbf{r}_j|$ is less than 2 nm, which corresponds approximately to the interaction range of Watson-Crick hydrogen bond in dsDNA [12]; (b) each node can be paired to at most one other node; (c) two-node pairs $(i,j)$ and $(k,l)$ can coexist only when they are either nested (i.e. $i<k<l<j$) or independent (i.e. $i<j<k<l$); this condition neglect the formation of pseudoknots which are known to be very rare in real RNA structure. (d) $|j-i| \geq 4$; this restriction permits flexibility of the chain and it is also necessary to rule out entirely the influence of phase space on the number of pairings, as confirmed by our following MC simulations.

The third energy describes the vertical stacking interactions of neighboring base-pairs. We approximate the basepairing energy by $E_{\text{sta}} = \sum_{i=1}^{N_s} V_s$, where $N_s$ is the number of stacked node-pairs and $V_s$ is the stacking potential between two neighboring node-pairs. Two node-pairs are considered as 'stacked' only when they are nearest neighbors to each other, e.g. pairs of $(i,j)$ and $(i+1,j-1)$.

The fourth energy in our model is the deformation energy of the chain when the length of its individual linker deviates from its equilibrium length $b$. According to Hooke's law for small rod deformation, the elastic energy of the ssDNA can be written as

$$E_{\text{ela}} = \frac{1}{2} \sum_{i=1}^{N} Y(|\mathbf{r}_i - \mathbf{r}_{i-1}| - b_0)^2$$

where $b_0$ is the Kuhn statistic length of ssDNA and $Y$ characterizes the stretch stiffness of the ssDNA backbone.

The fifth energy, i.e. work done by the external force $F$, is written as $E_{\text{for}} = -F z_N$, where $z_N$ is the coordinate of the last vertex, i.e. the distance of two terminus of ssDNA molecule in the directory of external force $F$. Here we have chosen the orientation of $F$ along the $z$-axis of the Cartesian system.

3.2.2 MC procedure of ssDNA

The Monte Carlo procedure of ssDNA is similar with that of dsDNA described above. Beside those three updates used in dsDNA, however, an additional movement is here used, which involves a permutation between a subchain of 2 segments at position $i$ and another subchain of 3 segments at position $j$. This move was firstly adopted by Vologodskii et al. [38] in the MC calculation of supercoiled dsDNA. Before the permutation, the conformations of both subchains of doublet and triplet should be deformed so that the length of the subchains could be incorporated to their new positions. The net result of this permutation is a translation of random chosen subchain of $(i,j)$ by 1 segment along the chain axis. Even though the acceptance probability of this movement can be quite low, it can substantially increase the probability of extrusion and resorption of hairpinned structures and help the simulation to go out of some local traps.

As mentioned above, the trial movements in the canonical Metropolis algorithm is accepted or rejected according to the Boltzmann weight of $p_i = \exp(-E_i/k_B T)$, where the total energy $E_i = E_{\text{ele}} + E_{\text{pair}} + E_{\text{sta}} + E_{\text{ela}} + E_{\text{for}}$ in the case of ssDNA. However, the energy landscape of ssDNA with hairpin structures is characterized by numerous local minima separated by energy barriers, and the probability of canonical Metropolis procedure to cross the energy barrier of height $\Delta E$ is proportional to $\exp(\Delta E/k_B T)$. When the pairing and stacking energies are rather large, e.g. in the case of poly(dG-dC), the energy barriers around some special conformations can be very high so that the simulation tends to get trapped in these conformation although they are by no means of the lowest energy. During the finite CPU time, only small parts of the canonical ensemble of DNA conformations can therefore be explored, rendering the calculation of physical quantities unreliable.

In order to overcome this problem of "ergodicity breaking" of poly(dG-dC) ssDNA, we produce an artificial ensemble according to a modified weight factor [39]
\[ \rho(E) = \exp\left[\left(\frac{E + \sqrt{2}}{\sigma} \mid E - \langle E \rangle \right) / k_B T\right], \tag{26} \]

where \(\sigma^2 = n_F/2\) is the mean squared derivation of energy of the canonical thermodynamic system, and \(n_F\) the number of degree of freedom of the chain, \(\langle E \rangle\) is the averaged energy of system which can be calculated in a simple iteration procedure \([39,40]\). In Eq. (26), the probability of both high- and low-energy are exponentially reinforced and sharp peak of the canonical ensemble around \(\langle E \rangle\) is damped, which can efficiently help the simulation to jump out from local energy basins. Since one configuration in the artificial ensemble of Eq. (26), in fact, represent \(n(E) = \exp[\sqrt{2} \mid E - \langle E \rangle \mid / (k_B T \sigma)]\) configurations in canonical system, we should reweight the artificial sample to obtain the expectation value of considered quantity. For example, the averaged extension \(z_N\) should be calculated by \([41]\)

\[ \langle z_N \rangle = \frac{\sum_{i=1}^{N_{MC}} n(E_i) z_N(E_i)}{\sum_{i=1}^{N_{MC}} n(E_i)} \tag{27} \]

where \(N_{MC}\) is the number of sweeps of the artificial sample.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>(V_s(k_B T))</th>
<th>(V_p(k_B T))</th>
<th>(b)</th>
<th>(Y(k_B T/nm^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>4.6</td>
<td>0</td>
<td>1.6</td>
<td>123.5</td>
</tr>
<tr>
<td>Poly(dA-dT)</td>
<td>4.1</td>
<td>4.0</td>
<td>1.6</td>
<td>123.5</td>
</tr>
<tr>
<td>Poly(dG-dC)</td>
<td>10.4</td>
<td>6.0</td>
<td>1.6</td>
<td>123.5</td>
</tr>
</tbody>
</table>

3.3 MC results of ssDNA

3.3.1 Force/extension of ssDNA

Until now, there are three groups who have pulled ssDNA of both random and designed sequence and presented their force/extension data in different salt environments \([5,7,9,10]\). These data offer us good opportunity to check the theoretical model and meanwhile determine the 4 main parameters in our model, i.e. the Kuhn length \(b\), stiffness of ssDNA backbone \(Y\), pairing potential \(V_p\), and stacking potential \(V_s\) (see Table II). In the following presented calculations, we make 20,000,000 Monte Carlo runs with \(N=100\) nodes at each given external force for each case, according to our CPU capacity. We have also confirmed that the larger value of \(N\), e.g. \(N=200\), with more MC runs will not lead to different results.

We at first calculate force/extension characteristics of FJC with elastic bonds in different solutions, but without including pairing and stacking potentials. As shown in Fig. 9a, the electrostatic interactions tend to swell the volume of chain and make the segments more easily aligned along the force direction. This is formally equivalent to enlarge the Kuhn length of ssDNA. The lower the ionic concentration becomes, the larger the effective Kuhn length is, and the more rigid the molecule looks like. Since ssDNA can be entirely pulled back in very high force and all data become the same in this force range, we can determine accordingly the stiffness of sugar-phosphate backbone of ssDNA as \(Y=123.5\ k_B T/nm^2\) in our calculations.

In Fig. 9b is the force/extension data of a plasmid ssDNA fragment of 10.4 kilobases under 5 mM MgCl\(_2\) and 2 mM NaCl solutions. Bearing in mind that the sequence is relatively random and the formed base-pairs in ssDNA are usually separated spatially from each other along the molecule, the stacking interaction is negligible in this case. We therefore take the stacking potential \(V_s=0\) in our calculations. The best fit to the data is \(V_p=4.6\ k_B T\) and \(b=1.6\ nm\). In lower salt case of 2 mM NaCl, the force-elongation curve is not influenced much by the pairing potential compared with Fig. 9a, since the dominant electrostatic repulsive potential exclude the bases from getting close to pairing interaction range and the node-pairing probability is very rare. However, in high salt solution of 5 mM MgCl\(_2\), the
force/extension property is the result of completion of two opposite interactions of pairing and electrostatic repulsive interactions. The experimental data suggest that the pairing effect is slight larger in this case.

Figure 9. External force as the function of distance of two ends of ssDNA, scaled by its B-form length \( L_B \). \( L_B = 3N_B/1.58769 \) in the modeling calculation, and other model parameters used in our Monte Carlo calculation are listed in Table II. (a), Monte Carlo results of pure freely jointed chain (FJC) without electrostatic, pairing and stacking potential (dash-doted line), and that with electrostatic interaction considered in 2 mM NaCl (dashed line) and 5 mM MgCl\(_2\) (solid line) solutions, but \( V_P = V_S = 0 \). (b), \( \lambda \)-phage DNA in 2 mM NaCl and 5 mM MgCl\(_2\) solutions. Data are taken from Ref. [10] (triangles and squares), and from Ref. [9] (circles). (c) and (d), for designed poly(dA-dT) and poly(dG-dC) sequences respectively in 150 mM NaCl solution. Data are from Ref. [7].

In the cases of designed poly(dA-dT)/poly(dG-dC) ssDNA, two types of unitary basepairs of A-T and G-C can be formed in respective sequences. Any two bases are possible to be paired and strong stacking interaction exists between two basepairs when they are nearest neighboring (i.e. at the van der Waals distance range [12]). The stacking potential can dramatically change the conformation of secondary structure of ssDNA in low force. When the stacking potential is absent, the base-pairs are formed quite randomly with formation of many interior-loops and branched structures. So the dispersed hairpin structure can be easily pulled open in a medium force [42]. When the stacking interaction exist, on the other hand, the base-pairs are encouraged energetically to be neighboring, and therefore lead to the formation of bulk hairpin structure. Both experimental data and modeling calculation suggest that it need a threshold force to pull back the bulk hairpin structure with existing of stacking interaction. The value of threshold force, i.e. the height of plateau in force/extension curve, is dependent on the pairing and stacking potentials of the nodes. The best fit to the data, as in Fig. 9c and d, are \( V_P = 4.1 k_B T \) and \( V_S = 4 k_B T \) in poly(dA-dT) sequence, \( V_P = 10.4 k_B T \) and \( V_S = 6 k_B T \) in poly(dA-dT) sequence.

3.3.2 Criticalness of force-induced phase transition

It has been shown from force/extension data that the elongation of nature ssDNA with random sequence under external force is gradual, but that of designed sequence is in a cooperative and discontinuous manner when stacking interaction is involved. This first-order like phase transition of the designed ssDNA system can also manifest itself in other aspects.

In Fig. 10a we present the average number of pairings scaled by the maximum pairing number of \((N-2)/2\) for poly(dA-dT) sequence in 150 mM NaCl. When the external force is smaller than the critical force \((F_c \sim 9.5 \text{ pN})\), the number of pairing \(2N_P/(N-2)\) is almost all the bases are paired. The number of stacked pairs \(N_S\) has its maximum value, signifying that all the pairs aggregate into a compact pattern. Around the external force of \(F_c\) \((-9.5 \text{ pN})\), the bulk hairpin is suddenly pulled back and
the numbers of pairings and stackings sharply decrease from their maximum to zero. This behavior is a reminiscence of temperature induced first-order transition in, e.g., a 2-dimensional spin system as described by Ising model. But there the order parameter is magnetization or number of spins with specified orientation. And here the order parameter is number of paired nodes, and the transition is force induced and takes place in the 1-dimensional system.

Figure 10. The order parameters and autocorrelation time as the function of external force for poly(dA-dT) sequence in 150 mM NaCl solution. (a), number of node-pairs ($N_P$), stackings ($N_S$), and the ratio of pairings and stackings, all of which are scaled be the maximum value of pairings of $(N-2)/2$. (b), electrostatic Debye-Hükel potential. (c), The autocorrelation time of extension of ssDNA polymer calculated by Eqs. (28) and (29). The MC time is scaled by the number of nodes $N$.

The electrostatic energy of ssDNA changes with external force also in a cooperative manner (Fig. 10b), because of discontinuous jump of the averaged distance between the ssDNA backbone at critical point. One can notice that there are irregular doglegs for the values of averaged distance of two ends, number of pairings and the electrostatic energy, when the external force approaches to the critical point. This is because of so-called critical fluctuation in our simulations of ssDNA chain at the critical point. In Fig. 11 is shown the time series of some order quantities such as the extension $z_N$ and number of pairings. The order quantities keep staying around their values of thermodynamic equilibrium when the external force is away from the critical point. When the external force approach $F_c$, however, the fluctuation of order quantities sharply enlarge, since the correlation length may diverge at this point.

As a confirmation, we calculate integrated autocorrelation time $\tau$ of the distance $z_N$ of two ends of the ssDNA, i.e. [50]

$$\tau = \int_0^\infty \frac{\chi(t)}{\chi(0)} dt,$$

(28)

where time $t$ is the Monte Carlo steps, which has been scaled by the magnitude $N$ of the chain, and the time-displaced autocorrelation function $\chi(t)$ is calculated as

$$\chi(t) = \int_0^{\infty} dt' [z_N(t') - \langle z_N \rangle][z_N(t'+t) - \langle z_N \rangle].$$

(29)
Here $t_{\text{max}}$ is the total time sweeps of Monte Carlo simulation, and $\langle \cdots \rangle$ denotes the time average along the Monte Carlo series.

**Figure 11.** Monte Carlo time series of scaled extension $z_N/L_p$, scaled pairing numbers $2N_p/(N-2)$, scaled stacking pairs $2N_s/(N-2)$, and the ratio of pairs and stacking numbers $N_s/N_p$ for poly(dA-dT) sequence in 150 mM NaCl solution. All the order parameters fluctuate around their thermal equilibrium values when force is far away from critical force $F_c=9.42$ pN, however the fluctuations diverge around the critical point.

In Fig. 10c is the time correlation length as the function of external force. At critical point, the correlation time indeed diverges. So the number of independent measurements, $n=t_{\text{max}}/2\tau$, is very small in the simulation, which renders the MC calculation at critical point unreliable. This effect, known as critical slowing down, is an inherent property of Monte Carlo algorithm used to perform the simulation for phase transition system. Some techniques, such as cluster-flipping algorithm [43,44], have been proposed to alleviated the problem of critical slowing down for a number of spin systems. However, an efficient algorithm for that of biopolymer system still lack. Having in mind the pronounced double-peak structure of the sample action density near critical point, which is the main reason of critical slowing down in our simulation, it is possible to construct a new weight factor and enhance the tunneling between these two metastable states at the critical point. The detail into this problem is being discussed somewhere else [45].

### 4 Summary and Conclusions

We summarize detailed theoretical studies of the elasticities of DNA molecules, and comparisons of modeling calculations and experimental measurements of stretching single- and double-stranded DNA molecules. In various laboratories, the dsDNA molecule has been stretched through different mediums in high- and low-force ranges with torsionally relaxed and supercoiled statuses [3-6,10,48], while the
ssDNA was pulled in different ionic atmosphere with relatively random and designed poly(dA-dT)/(dG-dC) sequences [7,8,9,10,48]).

In order to attain a comprehensive understanding of the experimental data, we model the dsDNA molecule as two interwound wormlike chain, bound by dominant hydrogen bonds. The structural deformation of dsDNA is characterized by the folding angle of sugar-phosphate strand and the molecule central axis. The stacking interaction is calculated through Lennard-Jones potential, determined by the distance of neighboring basepairs or the folding angle of backbones. The ssDNA molecule in our calculation is modeled as a freely jointed chain with elastic bonds. Base-pairing (i.e. hairpins formed by the pairing of complimentary bases of ssDNA) and vertical base-stacking potentials are incorporated in the FJC. In order to calculate the electrostatic interaction, Poisson-Boltzmann equation is numerically solved in both mono- and bi-valent ionic atmosphere. The amplitude, or effective charge density, of Debye-Hückel electrostatic potential is determined through matching the first-order modified Bessel function with the Poisson-Boltzmann potential in the overlap region.

The thermodynamics of DNA molecules in external fields of force and torque is calculated by both path integral method and Monte Carlo simulations. We conclude by summarizing the main results of the calculations.

**Double-stranded DNA:**

1. In low force range (0 ~ 10 pN), the elasticity of the molecule is entropy-dominated. A simple inextensible wormlike chain model with persistent length of 53 nm can give excellent description of the experimental results in this range [16,18,11,13].
2. In high force region (starting as several tens of pN), the back-stacking potential can be overcome and the helically stacked base-pairs are pulled apart at this stretched force, and therefore a structural transition from canonical B-form to a new overstretched conformation called S-DNA can be triggered [11,13]. This structural transition has been manifested by the cooperative elongation of dsDNA at about 70 pN, as observed by Cluzel et al. [4] and Smith et al. [5].
3. The elasticity of supercoiled DNA is decided by the interplay of external force/torque field and the stacking potential of basepairs. In low force and torque, the inherent helix structure is not perturbed, the elastic behavior of DNA is symmetrical under positive or negative supercoilings [21,22,27]; in intermediate force, the chirality of the elasticity appear since the basepairs are pulled back a little and the van der Waals stacking potential is asymmetric in this region [21]; in high force field, the contribution of the external field dominate that of van der Waals potential in both over- and underwound DNA, and the plectonemic DNA is fully converted to extended DNA. All these results are in qualitative agreements with experimental measurements of Strick et al. [6,19].

**Single-stranded DNA:**

1. In high salt solution, abounding secondary structure exists and hairpins can be formed, when the ssDNA bends back onto itself and complementary bases A-T and G-C are paired, gaining an energy of several $k_B T$ per pair. This interaction makes it need a slightly larger external force to pull back a $\lambda$-ssDNA than that expected in a pure FJC [9,10,29,30,51].
2. In low salt solution, the electrostatic repulsive interaction dominates, there are little formed basepairs since the strong electrostatic repulsive potential exclude the bases from getting close into the Watson-Crick basepairing range. This interaction make the chain more easily to be aligned and more subject to be stretched, which is formally equivalent to enlarge the stiffness of the chain.
3. For designed poly(dA-dT)/poly(dG-dC) chain, the stacking interaction between basepairs encourage base pairs to aggregate into a compact pattern, and it need a threshold force to pull back the bulk hairpin structure, characterized by a plateau in the force/extension curve. The height of plateau is determined by the pairing potential $V_P$ and stacking potential $V_S$ in our model. The best fit to the experimental data shows that $V_P=4.1 k_B T$ and $V_S=4 k_B T$ for poly(dA-dT) sequence, $V_P=10.4 k_B T$ and $V_S=6 k_B T$ for poly(dG-dC) sequence. Bearing in mind that each Kuhn length (~1.6 nm) contains about three nucleotide bases, we can infer that the pairing energy of each A-T base-pair is about 1.37 $k_B T$ and that for each G-C base-pair is about 3.47
These values are comparable with the measurements of Bockelmann et al. [49] when they pulled apart the two strands of a double-stranded DNA helix.

4. On the opposite to the gradual elongating of ssDNA of random sequence, the hairpin-coil transition of designed ssDNA is discontinuous. The calculated thermodynamics of stretching designed ssDNA sequence shows typical critical characteristics, as happens in well-known 2-dimensional Ising spin model. However, the transition in our model is force-induced and takes place in one-dimensional system. All the order parameters, such as distance of two ends of the ssDNA chain, number of pairings, and electrostatic potential, discontinuously jump when the external force pass through the critical force. At the critical point, the fluctuation of the order parameters and the autocorrelation time diverge. This effect, known as critical slowing down, renders the canonical Metropolis Monte Carlo calculations unreliable. All these features make the designed ssDNA an excellent laboratory for the study of first-order phase transition in one-dimensional system.

5 Acknowledgement

Y. Zhang is grateful to V. Sa-yakanit for his invitation and arrangement of the talk, and for his organizing of such an enjoyable workshop. He also thanks Drs. H. Frauenfelder, N. Go, L. Matsson, and W. Wiegel for helpful discussions. H. J. Zhou likes to thank the Alexander von Humboldt Foundation for financial support.

References