

Theoretical and Computational Treatments of DNA and RNA

Molecules

Haijun Zhou

*Max-Planck-Institute of Colloids and Interfaces,
D-14424 Potsdam, Germany*

Phone: +49 331 567 9617

Fax: +49 331 567 9612

email: haijun.zhou@mpikg-golm.mpg.de

Yang Zhang

*Center of Excellence in Bioinformatics,
University at Buffalo, Buffalo,
New York 14203, USA*

Phone: +1 716 849 6716

Fax: +1 716 849 6747

email: zhang6@buffalo.edu

Zhong-Can Ou-Yang

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

Phone +86 10 6255 4467

Fax: +86 10 6256 2587

email: oy@itp.ac.cn

Corresponding author: Zhong-Can Ou-Yang

Contents

I. Introduction	4
A. Basics of DNA and RNA	4
B. A brief survey of single-molecule experimental studies on DNA and RNA	7
C. Outline of this chapter	11
II. DNA denaturation and unzipping	13
A. DNA denaturation: de Gennes-Peyrard-Bishop model	14
B. DNA denaturation: Montanari-Mézard model	16
1. DNA thermal melting	17
2. Force-induced DNA melting	21
C. Poland-Scheraga model and excluded-volume effect	23
1. Excluded-volume effect	25
D. Force-induced unzipping of real DNA molecules	27
1. Scaling relationship in DNA unzipping	31
E. Unzipping kinetics of DNA/RNA helix-loops under constant force	33
III. Secondary structures of RNA polymer	38
A. The homogeneous-sequence approximation	39
1. Elasticity of real homogeneous RNA chains	43
B. structural transitions in a weakly designed RNA polymer	44
C. Glassy transitions in RNA sequences at low temperature	47
D. Single-stranded DNA/RNA at low salt solution	50
1. Debye-Huckel theory of biopolymers	51
2. Elastic freely-jointed chain model and interplay with electrostatic force	52
3. Base-pairing interactions of ssDNA/RNA in solution	53
4. Sequence sensitivity of ssDNA/RNA elasticity	54
E. RNA secondary structure prediction through mechanical stretching	55
IV. DNA entropic elasticity	58
A. The Gaussian model and the freely-jointed-chain model	58
B. The wormlike-chain model	59

V. DNA over-stretching transition	62
A. Co-operative over-stretching transition viewed from a discrete model	62
B. Double-stranded polymer model of DNA and the over-stretching transition	64
VI. Elasticity of Supercoiled DNA	69
A. Analytical approximation	70
B. DNA supercoiling studied by Monte Carlo simulation	73
1. Discrete DNA model	74
2. Monte Carlo simulation procedure	75
3. Elasticity of supercoiled dsDNA	76
VII. Conclusion	79
Acknowledgments	81
Appendix: Path Integral method in Polymer statistical physics	82
References	85

I. INTRODUCTION

Since the first direct measurement of the elasticity of single DNA at 1992 [1], mechanical properties of individual DNA and RNA macromolecules and interactions between DNA and various proteins have been extensively investigated by single-molecule methods (see, for example, review articles [2–8]). These single-molecule manipulation techniques include atomic force microscopy, magnetic beads, micropipettes, optical tweezers, and many others. Many new mechanical characteristics of DNA and RNA molecules were revealed by applying these unprecedented instruments. Inspired by this huge experimental revolution, there were a lot of theoretical modeling efforts in recent years. These theoretical and computational studies, combined with single-molecule observations, have resulted in greatly improved understanding on the elastic and statistical mechanical properties of DNA and RNA polymers.

This chapter tries to give a comprehensive review of the methods used in these theoretical researches and the insights gained from them. In this introductory section, we first summarize some basic knowledge of DNA and RNA for those readers who are not yet familiar with these molecules. It is followed by a brief survey of recent single-molecule experiments. In § IC we outline the main topics covered in this chapter.

A. Basics of DNA and RNA

Molecular biology becomes the main focus of modern biological studies ever since the discovery of the double-helix structure of DNA in 1953 by Watson, Crick, Wilkins, Franklin and co-workers [9–11]. Deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein are the main themes of molecular biology; they constitute the components of the central dogma (see, for example, books [12, 13]). .

DNA is a polymer made of deoxyribonucleotides and RNA is a polymer of ribonucleotides. In biological cells, DNA is usually very long. For example, the DNA sequence of *Escherichia coli* bacterium contains more than nine million nucleotides, and the total number of nucleotides in one cell of our body is of the order of ten billion (10^{10}). Nucleotides in DNA polymer are organized into two polynucleotide linear chains of equal length (Fig. 1). A nucleotide monomer is an asymmetric molecule, with a 3'-end and a 5'-end. In each of the two strands of DNA, the 3'-end of one nucleotide is connected to the 5'-end of another nucleotide

by a phosphate-diester covalent bond [14]. A polynucleotide chain therefore has a direction defined from its 5'-end to its 3'-end. There are four different types of nucleotide residues (bases), namely Adenine (A), Guanine (G), Cytosine (C), and Thymine (T). For most organisms, DNA is the molecule where the hereditary information is stored and maintained. The genetic information of an organism is recorded by the particular order of appearances of the four types of bases A, C, G, and T along each strand of DNA.

In DNA, the two poly(deoxyribon)nucleotide strands are anti-parallel to each other and are twisted around a common central axis to form a right-handed double-helix structure, with the genetic information (the nucleotide residues) buried in the interior of the cylindrical molecule and the highly negatively charged sugar-phosphate back-bones exposed outside. Such a structure has the benefit of prohibiting genetic information from being lost or destroyed, but it causes severe constraints on the access of genetic information and the transfer of it from one generation to another. Over billions of years of evolution and selection, biological organisms have developed complicated mechanisms to facilitate gene duplication and transcription [12].

In physiological conditions, the double-helix structure of DNA is very stable. This stability is the result of the following three contributions.

Firstly, the two strands of DNA are bound together by many hydrogen-bonded base-pairs. The two polynucleotide strands of DNA are complementary to each other. Suppose each strand has n nucleotides, complementarity means that, if at position i of one strand (counting from its 5'-end) the nucleotide is A, then at position $n - i$ of the other strand the nucleotide is T (the same rule holds for nucleotides G and C). Adenine and Thymine can form a planar A-T base-pair through the formation of two hydrogen bonds between them, Cytosine and Guanine could form a planar G-C base-pair through the formation of three hydrogen bonds between them [14]. The sugar-phosphate backbone length of DNA between two consecutive base-pairs is about 0.68 nm.

Secondly, the folded double-helix is stabilized by base-pair stacking interactions between adjacent nucleotide base-pairs [14]. The base-pair stacking potential is a short-ranged attractive interaction, with a maximum strength when the two DNA base-pair planes are separated by about 0.34 nm. Therefore, in its natural conditions, DNA is a right-handed double-helix, each peach of which contains about 10.5 base-pairs, and two neighboring base-pairs are separated vertically about 0.34 nm, the pitch length of the double-helix along the

central axis is about 3.6 nm [15].

Thirdly, the existence of high concentration of counter-ions in biological cells screens out the negative charges of DNA. Therefore, electrostatic repulsive interactions among negatively charged phosphate groups are weak and can not destroy the compact double-helix structure.

DNA is transferred from a parent cell to its daughter cells through a **replication** process [12]. The parent DNA is duplicated into two identical copies. When the parent cell divides into two daughter cells, each of them contains a copy of the duplicated DNA molecule. During each cell cycle, the DNA sequence is read by RNA polymerases and genetic information is transcribed into a message RNA (mRNA) linear chain. This process is called **transcription**. The mRNA chain is then transmitted to another part of the cell and read by a protein synthesis machine called the ribosome. This last step is a **translation** process, where each word (of three nucleotides) in the input mRNA corresponds to an amino acid in the product protein chain.

During the replication and transcription processes, the DNA molecule is severely deformed: the double-helix is locally bended, untwisted, stretched, compressed, and the base-pair patterns are locally destroyed. The ability of DNA to perform reversible structural transitions in response to various external conditions is very important for the biological function of DNA as a hereditary material. DNA has a bending persistence length of about 50 nm [16], which means that if the bending deformation occurs over a length scale less than 150 base-pairs, a considerable bending energy must be overcome. DNA could also be twisted by applying a torsional stress on the molecule. The twist modulus of DNA is recently measured to be 440 ± 40 pN nm² [17] (pN means piconewton or 10^{-12} N). The stretching modulus of DNA is about 1000 pN [18]. However, when DNA is stretched with a force larger than 70 pN, its end-to-end extension suddenly increases to 1.7 times its relaxed contour length [18, 19]. The double-helix structure of DNA can also be severely changed by external torque [20–23].

The biological role of RNA is more versatile compared to that of DNA. Besides mRNA, there are tRNA (transfer-RNA), which brings amino acids to the protein-assembly machine, and rRNA (ribosomal RNA), which resides in the ribosome and is involved in protein synthesis. Some RNA chains can also have catalytic functions. Furthermore, in some kinds of viruses the hereditary information is stored in RNA rather than in DNA.

RNA exists as a single-stranded polymer. The nucleotide residues of RNA are also com-

posed of four types as in the case of DNA, but with Uracil (U) replacing Thymine. Although RNA is a single-stranded polymer, it can contain local right-handed double-helix structures. To form a local double-helix segment, one part of the RNA chain must bend at certain position to form a hairpin loop, and the hairpin structure is then stabilized by base-pairing and base-pair stacking interactions (for an example, see Fig. 2). Besides G-C base-pairs, an Adenine could form a A-U base-pair with a Uracil, and G-U base-pairs can also be formed.

The ability for some RNA chains (such as tRNA) to fold into stable three-dimensional structures are very important for their biological functions. Given a RNA sequence, a computational challenge is to predict its stable structures. On the other hand, for mRNA, it might be unfavorable to have stable three-dimensional configurations. If a mRNA forms a hairpin pattern with two distinct RNA segments bind to each other, this hairpin configuration may restrict the passage of mRNA through the ribosome molecule, leading to a blockage of protein synthesis. For a mRNA sequence, it is possible to have many transient hairpin patterns. These hairpin patterns may not be stable, therefore they can be destroyed by the force generated by the ribosome on mRNA during translation. In eukaryotes, it is also conceivable that these hairpin patterns might play some role in RNA intron splicing and alternative gene expression. The links between RNA sequence and biology is still not quite clear.

B. A brief survey of single-molecule experimental studies on DNA and RNA

Prior to the advent of single-molecule methods, properties of polynucleotides were inferred from bulk thermodynamic measurements. An example along this line is measuring DNA denaturation curves based on light absorption or calorimetry (see, for example, review article [24] and book [25]). Many thermodynamic parameters of DNA double-helix were estimated by these ways. However, bulk experiments with a certain concentration of macromolecules can only measure the polymer's average behavior. Fluctuations due to molecular individuality can not be traced. There is another drawback: it is not always easy to separate the effects due to intramolecular interactions from those resulting from intermolecular interactions.

Single-molecule methods, which can work in the broad force range of 10^{-15} - 10^{-9} N, directly record the response of a macromolecule under the action of environmental stimuli.

It enables researchers to gain detailed view of a macromolecule's mechanical properties. It also stimulates the need to perform theoretical modeling efforts on single-molecule systems with the aim of quantitatively reproducing experimental observations.

In 1992, the three authors S. B. Smith, L. Finzi and C. Bustamante [1] fixed one end of a lambda-phage DNA to a glass slide and attached a magnetic bead to the other end. They then recorded the force-extension of DNA by applying a magnetic force on the attached bead. Very novel elastic response of DNA was observed [1]. This research has stimulated a lot of efforts in single-molecule investigation on polynucleotides. Here we briefly mention some of these studies to give the reader an overview of the present experimental situations. For other recent experimental reviews, see [2–8].

Elasticity of double-stranded DNA Smith and co-workers [1] found that the force-extension curves of lambda-phage DNA do not follow predictions based on standard Gaussian polymer model or freely-jointed-chain model. Later it was realized that semiflexible worm-like-chain polymer could reproduce the observation of Smith et al. to a large extent [16].

Bensimon and co-workers [26] used a receding meniscus to measure the elasticity of DNA. They found that DNA could be stretched up to 2.14 times its relaxed contour length before breaking. Bustamante and co-workers also announced briefly in a conference abstract [27] that DNA could stretched to 1.85 times its B-form length before breaking and that the over-stretching was initiated suddenly at about 50 pN. In the experimental setup of Smith et al. [27] the ends of a lambda-DNA were attached to two micro-bead and moved by micropipettes. Later, Bustamante and co-workers repeated this experiment by using optical tweezers [18]. The unzipping force was more precisely measured to be about 65 pN and the over-stretched DNA was found to be 1.7 times its B-form length. Caron and co-workers [19] performed a similar experiment but with an optical fiber as the force level, their result was consistent with that of Smith et al. [18].

Croquette and co-workers [20] fixed the two strands of an lambda-DNA to a glass slide at one end and to a magnetic bead at the other end. They then applied a magnetic field to rotate and stretch the bead, thereby adding torsional as well as elongational stress in DNA double-helix [20]. Very interesting behaviors of DNA were observed in this and subsequent experiments [20–23]. The elastic response of an under-twisted DNA is quite different from that of an over-twisted DNA. Over-twisting might cause B-form DNA to transit into a new P-form DNA with exposed nucleotide bases [23], while under-twisted DNA might be in a

strand-separated form [21] or a left-handed Z-form [28–30].

The dynamics of single DNA macromolecules in elongational or shear flows were studied by Chu and co-workers [31–33].

Elasticity of single-stranded DNA Smith and co-workers [18] studied the elastic response of single-stranded (ss) lambda-phage DNA at the action of an external force. They found that the elasticity of ssDNA could be well modeled by a freely-jointed-chain model with Kuhn length 1.5 nm and stretching modulus 800 pN, or by a worm-like-chain model. Croquette and co-workers [34] also studied the elastic response of ssDNA using magnetic bead and magnetic force. At high salt conditions, both the modified freely-jointed-chain model and the worm-like-chain model could not reproduce the force-extension curve of ssDNA when the external force is less than 5 pN. Bustamante and co-authors [18, 35, 36] also observed similar behavior. This discrepancy between theory and experiment was attributed to the formation of hairpins in ssDNA chain.

At low salt conditions and low external forces, Bustamante et al. [2] also observed that the force-extension curve of ssDNA is much deviated from that of a freely-jointed-chain. This was caused by strong electrostatic repulsive interaction between negatively charged nucleotide monomers.

DNA unzipping and rupture Heslot and co-workers [37] separated the two strands of a lambda-phage DNA by fixing the 5'-end of one strand to a glass slide and the 3'-end of the other strand to a soft force level and moving the force level at constant velocity. They found that the resulting force-extension is directly related to DNA sequence [37–40]. DNA unzipping experiment were also performed by Danilowicz et al. [41] under the condition of fixed external force.

Lee and co-authors [42] fixed the 5'-end of one strand of a DNA oligomer of 12 base-pairs to a surface and the 5'-end of the other strand to the tip of an atomic force microscopy and then stretched the molecule. The rupture force between the two strands was measured to be 410 pN in 0.1 M NaCl [42, 43]. Similar experiments on DNA oligomers of various lengths and sequences were undertaken by Noy and co-workers [44], Strunz and co-workers [45], and Pope and co-workers [46] under different ionic conditions. Very surprisingly, the rupture forces for DNA oligomers are typically much lower than the melting forces of long DNA polymers. Rief and co-workers [47, 48] measured the melting force for lambda-phage DNA at room temperature to be about 150 pN. This dramatic difference may be due to the

following fact: In lambda-phage DNA there are many single-stranded breaks, these nicks enable DNA's torsional strain induced by stretching to be released. However, single-stranded breaks are absent in the DNA oligomers studied in Refs. [42–46].

The unzipping kinetics of short DNA chains was studied by Sauer-Budge et al. [49] by letting single-stranded but not double-stranded DNA segment to pass through a nano-pore. They found that the strand separation for perfectly matched DNA double-helices of tens of base-pairs an all-or-nothing process. However, there are unzipping intermediate states when a DNA oligomer with a single mismatch was forced to traverse the nano-pore. The experimental setup of Ref. [49] is therefore capable of discriminate between minute differences in DNA sequences.

Even the unzipping forces between single nucleotide base-pairs were measured by atomic force microscopy. Boland and Ratner [50] found that the force to separate a single A-T base-pair is 54 pN. This value, however, is much higher than the unzipping force of 9-12 pN measured in Refs. [37, 47].

RNA unfolding The unfolding processes of several native RNA molecules were studied by many authors [51–54]. These non-equilibrium force-extension measurements were used to construct the energy landscape of RNA unfolding [52, 55, 56]

DNA-protein interactions The interaction between DNA and RNA polymerases were studied by Block and co-workers [57–60] and also by Bustamante and co-authors [61]. These authors found that, at low external forces, RNA polymerases move along DNA template with an average velocity of 16 nucleotides per second. The stalling force for RNA polymerases is about 25 pN.

DNA replication dynamics was studied by Maier, Bensimon, and Croquette [34] and by Bustamante and co-workers [36] using single-molecule methods. It was found that the rate of DNA replication depends strongly on the applied force. When the external force exceeds 40 pN, DNA polymerase began to move back along the DNA template [36].

The working mechanisms of DNA topoisomerases were probed in Refs. [62].

The mechanical property of single chromatin fibers were studied by Cui and Bustamante [63]. One of the findings is that, at physiological ionic conditions, condensation-decondensation transition could be induced by an external force of 5 pN, suggesting the chromatin fiber to be in a dynamic structure because of thermal fluctuations. The dynamic structure of individual nucleosomes was also explored by Wang and co-authors [64]

by stretching an tandem array of nucleosomes. The typical force needed to destroy the nucleosome structure was measured to be about 25 pN. The kinetics of chromatin assembly was recorded by Viovy and co-authors [65].

Poirier and Marko [66] demonstrated convincingly that, in a mitotic chromosome there is no contiguous protein scaffold. Covalent connection in a chromosome is therefore due to the DNA chain.

The interaction of RecA protein with double-stranded and single-stranded DNA were studied experimentally by Chatenay and co-authors [67]. These researchers found that structural fluctuations in double-stranded DNA are coupled to the RecA-DNA binding reaction. Related experiments were also undertaken by Libchaber and co-workers [68, 69] and by Bustamante and co-authors [35].

DNA sequencing Quake and co-workers [70] reported that sequence information of DNA could be obtained by performing experiments on just one single DNA molecule. The four types of nucleotides are labeled with four kinds of fluorescent agents. The incorporation of these nucleotides by a DNA polymerase which moves along the DNA chain is then recorded in real-time. The authors [70] found that single-nucleotide resolution could be reached by this method.

It is anticipated that more and more single-molecule experimental observations on the interaction between protein and DNA will be reported in the near future.

C. Outline of this chapter

On the theoretical side, analytical as well as computational methods have been developed in understanding single-molecule experimental observations. In §II we outline some models on the melting phenomenon of double-stranded DNA. In § III theoretical work on the structural transitions in RNA secondary structures will be mentioned in detail. In § IV we review the entropic elasticity theory of DNA. In § V we discuss DNA over-stretching transition. In section § VI we briefly review works done on the supercoiling property of DNA. Finally, § VII is reserved for conclusion.

There is a huge theoretical literature on the study of DNA and its interaction with proteins. The materials presented in this chapter by no means are complete. They are inevitably biased by our own expertise and taste. For example, theoretical investigations on

the interaction between DNA and protein are completely neglected in this chapter. Recent theoretical reviews include Refs. [71, 72].

II. DNA DENATURATION AND UNZIPPING

As mentioned in § I A, DNA is a double-stranded polymer, with the two strands bound together by hydrogen bonds. Compared to a covalent bond, a hydrogen bond is very weak. The energy needed to break a hydrogen bond is comparable to thermal energy $k_B T$ at room temperature of $T \simeq 300\text{ K}$, where k_B is Boltzmann's constant. An A-T base-pair contains two hydrogen bonds and a G-C pair contains three. Because of thermal fluctuations, at physiological temperatures a base-pair may transit frequently between its paired state and the unpaired open state. Researchers are now quite interested in the biological significance of such local dynamical behavior of DNA double-helix (see, for example, [73–75]). Here we focus on the global equilibrium properties of DNA double-helix and review theoretical works on the denaturation transition in long DNA chains.

Experimentally, it has been known for many years that when a solution of DNA macromolecules is heated to about $80\text{ }^\circ\text{C}$, the base-pairs in the DNA double-helix break up cooperatively and the two DNA strands dissociate from each other to form two separated random coils. This phenomenon is referred to as DNA *denaturation* or *thermal DNA melting* [14]. Even at relatively low temperature where thermal DNA melting does not occur, the two strands of DNA could be separated by applying an oppositely directed force on the two strands at the DNA terminal point [37, 41]. This phenomenon is called *force-induced DNA melting* or *DNA unzipping*. DNA unzipping is also referred to as *directional DNA melting* [37], because the strand separation propagates from one end of the polymer chain to the other end.

The two strands of a DNA can also be separated by pulling its two opposite 5' ends with a force (see, for example, Ref. [42–46]). This process is called *DNA rupture*. A theoretical investigation of DNA rupture could be found in Ref. [76]. In this section, we discuss only DNA unzipping and will not treat DNA rupture.

We investigate the statistical mechanics of thermal and directional DNA melting. In § II A we discuss the de Gennes-Peyrard-Bishop model. In § II B we introduce the Montanari-Mézard model and outline its main results. In § II C the renowned Poland-Scheraga model is recalled, followed by reviewing the calculations of Kafri and co-workers on the order of DNA denaturation transition. In § II D we discuss how the sequence information in DNA will be manifested in the force-extension curves in an equilibrium DNA unzipping experiment. In

§ II E, we discuss kinetic issues related to DNA unzipping.

Earlier reviews on DNA thermal melting could be found in Refs. [24, 25, 77]

A. DNA denaturation: de Gennes-Peyrard-Bishop model

A simplified continuous model of DNA denaturation was investigated first by de Gennes [78] and later by Peyrard and Bishop [79]. This model regards that base-pairing and base-pair stacking interaction both to be important to understand DNA denaturation. These two types of interactions are considered by the following Hamiltonian:

$$\mathcal{H} = \sum_{n=1}^N V(x_n) + \sum_{n=1}^{N-1} \frac{k}{2} (x_{n+1} - x_n)^2. \quad (1)$$

In Eq. (1), x_n denotes the bond length of the n -th base-pair, and $V(x)$ is the base-pairing potential. $V(x)$ is a short-ranged potential, and Peyrard and Bishop modeled it by a Morse potential for analytical convenience:

$$V(x) = D [e^{-ax} - 1]^2. \quad (2)$$

D is set to $5.29 \times 10^{-22} J$ and $a^{-1} = 1.8 \text{ \AA}$, which corresponds to mean value for the hydrogen bonds in A-T and G-C base-pairs.

The second term of Eq. (1) models couplings between two adjacent base-pairs caused by base-pair stacking. The coupling strength k is of the order of $4.5 \times 10^{-24} J/\text{\AA}$ [79]. In general, the values of the parameters D , a , and k should all be sequence-dependent.

The partition function is

$$\mathcal{Z}(N) = \int_{-\infty}^{\infty} \prod_{n=1}^N dx_n \exp[-\beta \mathcal{H}(N)], \quad (3)$$

where $\beta = 1/k_B T$. Because base-pair stacking introduces only nearest-neighbor coupling in Eq. (1), the partition function Eq. (3) could be evaluated by transfer-matrix method [79]. Denote ϵ_i and $\phi_i(x)$ as, respectively, the i -th eigenvalue and eigenfunction of the integral equation

$$\int_{-\infty}^{+\infty} dx' \exp \left[-\frac{k}{2} (x - x')^2 + V(x) \right] \phi_i(x') = e^{-\beta \epsilon_i} \phi_i(x), \quad (4)$$

then Eq. (3) is re-written as

$$\mathcal{Z}(N) = \sum_{i=0}^{\infty} e^{-(N-1)\beta \epsilon_i} \int_{-\infty}^{+\infty} dx_N \phi_i(x_N) \int_{-\infty}^{+\infty} dx_1 e^{-\beta V(x_1)} \phi_i(x_1). \quad (5)$$

As long as the temperature T is low enough such that $d = (\beta/a)(2kD)^{1/2} > \frac{1}{2}$, Eq. (4) has a confined ground-state eigenfunction of

$$\phi_0(x) = a^{1/2} \frac{(2d)^{d-1/2}}{[\Gamma(2d-1)]^{1/2}} \exp(-de^{-ax}) \exp\left(-\left(d - \frac{1}{2}\right)ax\right). \quad (6)$$

In the thermodynamic limit of $N \rightarrow \infty$, Eq. (4) reduces to $\mathcal{Z}(N) \propto \exp[-\beta N \epsilon_0]$. The average separation of a base-pair is equation to

$$\langle x \rangle = \int_{-\infty}^{+\infty} x' \phi_0^2(x') dx' \quad (7)$$

Peyrard and Bishop [79] calculated the average separation between a DNA base-pair based on Eq. (7). They found that when the temperature is increased, the hydrogen-bond stretching has a sigmoidal form and changes gradually from about 2 Å to about 10 Å over a temperature range of about 50 K. They also found that the transition temperature is greatly influenced by the coupling constant k in Eq. (1). The work of Peyrard and Bishop [79] suggested that base-pair stacking interactions as well as the short-ranged nature of base-pairing interactions are quite important to understand DNA denaturation. More refined models following Ref. [79] were also constructed and investigated detailedly [80–87].

The experimentally observed DNA denaturation is much more cooperative than that predicted in Ref. [79]. For this purpose, Dauxois and co-authors [81, 86] introduced non-harmonic nearest-neighbor stacking interactions. This was achieved by assuming the coupling k in Eq. (1) to be a separation-dependent quantity. Their calculations leads to a much sharper behavior than the original model [81, 86]. In Ref. [85], twisting degrees of freedom are also incorporated into the model to improve the agreement between theory and experiment.

The original calculations of Ref. [79] considered only the (localized) ground-state eigenvector of Eq. (4). For short DNA molecules where the thermodynamic limit does not hold, it is necessary to consider both the localized and extended eigenfunctions of Eq. (4). Extensive numerical work of Zhang and co-authors [84] demonstrated that, after this improvement, actually the Peyrard-Bishop model could reproduce the melting curves of some naturally occurring short DNA chains with high precision.

Zhou discussed the phase-transition issue of DNA denaturation based on the de Gennes-Peyrard-Bishop model. In his calculation, the Morse potential Eq. (2) was replaced by an

asymmetric δ potential for analytical convenience

$$V(x) = \begin{cases} +\infty, & x < 0; \\ -\gamma\delta(x - a), & x \geq 0. \end{cases} \quad (8)$$

In addition, an energy term $-fx_N$ is added into Eq. (1) to mimic the effect of external unzipping. By taking into all the eigenstates of this model system, the Laplace transform of the partition function is exactly worked out [87]

$$G(z) = \sum_{N=0}^{\infty} \exp(-zN)\mathcal{Z}(N) = \frac{8\sqrt{z(2k\beta)^3}a[\exp(\beta fa/2) - \exp(-\sqrt{2zk\beta}a)]}{[8zk\beta - \beta^2 f^2][2\sqrt{2zk\beta}a - \tau(1 - \exp(-2\sqrt{2zk\beta}a))]} \quad (9)$$

where $\tau = 2ak\gamma\beta^2$. The largest solution of the equation $1/G(z) = 0$ corresponds to the linear free energy density of the polymer system.

Reference [87] showed that denaturation is a localization-delocalization transition process. When $T < T_m = \sqrt{2k\gamma a}/k_B$, there is one localized eigenstate, corresponding to base-paired DNA. The phase-transition occurs at T_m when the localized ground-state disappears. When there is no external force, the denaturation transition is second-order, provided that the hydrogen-bonding interaction $V(x)$ is asymmetric (for symmetric potential $V(x)$ there is no true phase-transition). When the external force is non-zero, the denaturation is first-order. The phase-diagram based on this simplified model is shown in Fig. 3.

The de-Gennes-Peyrard-Bishop model is essentially one-dimensional. The configurational fluctuations in DNA strands are all neglected. In the next section, we discuss a more realistic model of DNA denaturation which considers DNA as a three-dimensional object.

B. DNA denaturation: Montanari-Mézard model

Montanari and Mézard [88] constructed a three-dimensional polymer model to study the secondary structures in RNA polymers (see § III). However, as we show here, this model is also suitable to study DNA denaturation and unzipping.

In the Montanari-Mézard model, each DNA strand is regarded as a deformable freely-jointed chain of N bonds and $N + 1$ beads (see Fig. 4). The $2(N + 1)$ beads along the two strands are located at spatial positions $\{\mathbf{r}_0^\alpha, \mathbf{r}_1^\alpha, \dots, \mathbf{r}_N^\alpha\}$, with $\alpha = 1$ for the first strand and $\alpha = 2$ for the second strand. In each strand, the bond linking two consecutive beads i and

$i + 1$ has the following bond length distribution [88]:

$$\mu(\mathbf{u}_i) \propto \exp\left(-\frac{(|\mathbf{u}_i| - b)^2}{2l_0^2}\right) \quad (10)$$

where $\mathbf{u}_i = \mathbf{r}_{i+1}^\alpha - \mathbf{r}_i^\alpha$ is the bond vector, b is the intrinsic length of the bond, and l_0 is the typical magnitude of bond length fluctuations. For single-stranded DNA, $b \simeq 1.7$ nm [2, 18] and $l_0 \sim 0.1$ nm.

A bead of the model is an effective interaction unit and corresponds to about three nucleotide bases of a real DNA strand. If the distance $\mathbf{r}_i = \mathbf{r}_i^1 - \mathbf{r}_i^2$ between the i -th pair of beads is within some range a , there is an attractive interaction ϵ_i . The attractive interaction is base-pair dependent, a G-C pair is stronger than an A-T pair. We are interested here in the general physics, so we assume all the pairing interactions to be sequence-independent: $\epsilon_i \equiv -\epsilon_0$, with ϵ_0 being the average pairing energy between two complementary DNA segments of length corresponding to one bead of the model. The effect of sequence heterogeneity to DNA unzipping will be discussed later. Cule and Hwa have also performed Monte-Carlo simulation to study the changing of denaturation behavior due to sequence heterogeneity, see Ref. [83].

Apart from this kind of “on-site” base-pairing interactions, the stacking interaction between two consecutive base-pairs is also considered. If both $|\mathbf{r}_i| \leq a$ and $|\mathbf{r}_{i+1}| \leq a$, then there is an additional stacking energy gain of $-\epsilon_s$. In principal, ϵ_s should also depend on sequence, but this complication is dropped here.

In the following theoretical treatment, excluded volume effect is neglected, and the local helical structure of DNA is not considered. The model could be investigated by many means. Here we follow the graph-counting and generating-function method used in Refs. [88–93]. § IIB1 investigates DNA denaturation and § IIB2 treats unzipping.

1. DNA thermal melting

The partition function of a double-stranded polymer of $2n$ bonds is denoted by $Z_{\text{ds}}(n, \mathbf{r})$, where \mathbf{r} is the separation between the last $[(n + 1)$ -th] pair of beads (see Fig. 4 (left)). We need to distinguish between two situations: (i) the last pair of beads are not base-paired, i.e., $|\mathbf{r}| > a$; (ii) the last pair of beads are base-paired, i.e., $|\mathbf{r}| \leq a$. Denote the partition function for the first case as $Z_{\text{ds}}^{(\text{u})}(n, \mathbf{r})$ and that of the second case as $Z_{\text{ds}}^{(\text{p})}(n, \mathbf{r})$. The total

partition function is

$$Z_{\text{ds}}(n, \mathbf{r}) = \Theta(a - |\mathbf{r}|)Z_{\text{ds}}^{(\text{p})}(n, \mathbf{r}) + [1 - \Theta(a - |\mathbf{r}|)]Z_{\text{ds}}^{(\text{u})}(n, \mathbf{r}). \quad (11)$$

In Eq. (11), $\Theta(x)$ is the Heaviside step function: $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ if $x < 0$.

Based on the graph representation of Fig. 4 (left), the following recursive equations are written down for $n \geq 1$:

$$Z_{\text{ds}}^{(\text{p})}(n, \mathbf{r}) = \Theta(a - |\mathbf{r}|)e^{\beta\epsilon_0} \int \left[\prod_{i=1}^3 d\mathbf{u}_i \right] \mu(\mathbf{u}_1)\mu(\mathbf{u}_2)\delta\left(\sum_{i=1}^3 \mathbf{u}_i - \mathbf{r}\right) \times \\ \times \left[Z_{\text{ds}}^{(\text{u})}(n-1, \mathbf{u}_3) + e^{\beta\epsilon_s} Z_{\text{ds}}^{(\text{p})}(n-1, \mathbf{u}_3) \right], \quad (12)$$

$$Z_{\text{ds}}^{(\text{u})}(n, \mathbf{r}) = [1 - \Theta(a - |\mathbf{r}|)] \int \left[\prod_{i=1}^3 d\mathbf{u}_i \right] \mu(\mathbf{u}_1)\mu(\mathbf{u}_2)Z_{\text{ds}}(n-1, \mathbf{u}_3)\delta\left(\sum_{i=1}^3 \mathbf{u}_i - \mathbf{r}\right). \quad (13)$$

Equations (12) and (13) are supplemented with the initial condition $Z_{\text{ds}}^{(\text{p})}(0, \mathbf{r}) = \delta(\mathbf{r})$ and $Z_{\text{ds}}^{(\text{u})}(0, \mathbf{r}) = 0$.

Hydrogen-bonding is a short-ranged interaction, with its range of interaction a much shorter than the Kuhn length b of a single-stranded DNA chain: $a \ll b$. Take into account this fact, a good approximation is to write $\Theta(a - |\mathbf{r}|) \simeq \frac{4\pi}{3}a^3\delta(\mathbf{r})$. This approximation makes our following analytical calculations much easier to perform.

The Fourier transform of a function $f(\mathbf{r})$ is indicated by an over-tilde and is defined as

$$\tilde{f}(\mathbf{p}) = \int d\mathbf{r} \exp(i\mathbf{p} \cdot \mathbf{r})f(\mathbf{r}). \quad (14)$$

Equations (11), (12) and (13) are Fourier transformed to

$$\tilde{Z}_{\text{ds}}^{(\text{p})}(n, \mathbf{p}) = \frac{a^3}{6\pi^2}e^{\beta\epsilon_0} \int d\mathbf{q}\tilde{\mu}^2(\mathbf{q})\tilde{Z}_{\text{ds}}(n-1, \mathbf{q}) + \\ + \frac{a^3}{6\pi^2}e^{\beta\epsilon_0}(e^{\beta\epsilon_s} - 1) \int d\mathbf{q}\tilde{\mu}^2(\mathbf{q})\tilde{Z}_{\text{ds}}^{(\text{p})}(n-1, \mathbf{q}), \quad (15)$$

$$\tilde{Z}_{\text{ds}}^{(\text{u})}(n, \mathbf{p}) = \tilde{\mu}^2(\mathbf{p})\tilde{Z}_{\text{ds}}(n-1, \mathbf{p}) - \frac{a^3}{6\pi^2} \int d\mathbf{q}\tilde{\mu}^2(\mathbf{q})\tilde{Z}_{\text{ds}}(n-1, \mathbf{q}), \quad (16)$$

$$\tilde{Z}_{\text{ds}}(n, \mathbf{p}) = \tilde{Z}_{\text{ds}}^{(\text{p})}(n, \mathbf{p}) + \tilde{Z}_{\text{ds}}^{(\text{u})}(n, \mathbf{p}). \quad (17)$$

In the above equations, $\tilde{\mu}$ is the Fourier transform of the bond vector distribution Eq. (10),

$$\tilde{\mu}(\mathbf{p}) = \int d\mathbf{r} \exp(i\mathbf{p} \cdot \mathbf{r})\mu(\mathbf{r}) = \frac{\sin b|\mathbf{p}|}{b|\mathbf{p}|} \exp\left(-\frac{l_0^2|\mathbf{p}|^2}{2}\right). \quad (18)$$

The generating function $G(z, \mathbf{p})$ of $Z_{\text{ds}}(n, \mathbf{p})$ is defined by

$$G(z, \mathbf{p}) = \sum_{n=0}^{+\infty} \tilde{Z}_{\text{ds}}(n, \mathbf{p})z^n. \quad (19)$$

Based on Eqs. (15), (16), and (17) we arrive at the following expression:

$$G(z, \mathbf{p}) = \frac{1}{1 - z\tilde{\mu}^2(\mathbf{p})} \times \frac{1}{1 - \frac{za^3}{6\pi^2} e^{\beta\epsilon_0} (e^{\beta\epsilon_s} - 1) \int d\mathbf{q} \tilde{\mu}^2(\mathbf{q}) \left[1 + \frac{za^3}{6\pi^2} \int d\mathbf{q} \frac{\tilde{\mu}^2(\mathbf{q})}{1 - z\tilde{\mu}^2(\mathbf{q})} \right] - \frac{za^3}{6\pi^2} (e^{\beta\epsilon_0} - 1) \int d\mathbf{q} \frac{\tilde{\mu}^2(\mathbf{q})}{1 - z\tilde{\mu}^2(\mathbf{q})}}. \quad (20)$$

The total partition of a double-stranded polymer of $2n$ bonds is

$$Z_{\text{ds}}(n) = \int d\mathbf{r} Z_{\text{ds}}(n, \mathbf{r}) = \tilde{Z}_{\text{ds}}(n, \mathbf{0}). \quad (21)$$

The generation function of $Z_{\text{ds}}(n)$ is nothing but $G(z, \mathbf{0})$. For a linear polymer system with only short-range interactions, the total free energy is a linear function of system size n , plus correlation terms which scale logarithmically with n in leading order [89]. We expect to find the following asymptotic expression for the partition function $Z_{\text{ds}}(n)$ when $n \gg 1$:

$$Z^{(n)} \sim n^{-\gamma} \exp\left(-\frac{ng}{k_B T}\right), \quad (22)$$

where g is the free energy linear density of the system and γ is a scaling constant. From Eq. (19) and Eq. (22) we know that $g = k_B T \ln z_c$, where z_c is the smallest positive singular point of the function $G(z, \mathbf{0})$.

Function $G(z, \mathbf{0})$ has a singular point at $z = z_{\text{coil}} = 1$ and another singular point at the point z_{native} determined by $y(z_{\text{native}}) = 0$, with

$$y(z) = 1 - \frac{za^3 e^{\beta\epsilon_0} (e^{\beta\epsilon_s} - 1)}{6\pi^2} \int d\mathbf{q} \tilde{\mu}^2(\mathbf{q}) \left[1 + \frac{za^3}{6\pi^2} \int d\mathbf{q} \frac{\tilde{\mu}^2(\mathbf{q})}{1 - z\tilde{\mu}^2(\mathbf{q})} \right] - \frac{za^3 (e^{\beta\epsilon_0} - 1)}{6\pi^2} \int d\mathbf{q} \frac{\tilde{\mu}^2(\mathbf{q})}{1 - z\tilde{\mu}^2(\mathbf{q})}. \quad (23)$$

Function $y(z)$ monotonously decreases with z in the range $0 \leq z \leq 1$, with $y(0) = 1$. Because $1 - \tilde{\mu}^2(\mathbf{q}) \propto \mathbf{q}^2$ as $|\mathbf{q}| \rightarrow 0$, we know from Eq. (23) that, for three-dimensional systems of our interest, $y(1)$ is finite but $\left. \frac{dy(z)}{dz} \right|_{z=1} = -\infty$.

When environment temperature T is low, $y(1) < 0$ and $y(0) > 0$. There must be a root of $y(z_{\text{native}}) = 0$ with $0 < z_{\text{native}} < 1$. The smallest positive singularity of $G(z, \mathbf{0})$ is at z_{native} . When T is high enough such that $T > T_m$ (the melting temperature), $y(1) > 0$ and $y(0) > 0$. There is no root of $y(z) = 0$ in the range $0 \leq z \leq 1$. The smallest singularity of $G(z, \mathbf{0})$ is at $z_{\text{coil}} = 1$. There is a phase transition at the melting temperature T_m when $y(1) = 0$.

At any temperature T , the fraction of hydrogen-bonded base-pairs $n_{\text{bp}}(T)$ is determined by

$$n_{\text{bp}}(T) = -k_B T \frac{d \ln z_c}{d \epsilon_0} = \begin{cases} -k_B T \frac{d \ln z_{\text{native}}}{d \epsilon_0}, & (T \leq T_m) \\ 0. & (T > T_m) \end{cases} \quad (24)$$

In Fig. 5 the number of intact base-pairs n_{bp} as a function of T is shown for a double-stranded DNA made of 50% G-C base-pairs and 50% A-T base-pairs (the sequence is assumed to be random). The melting temperature is set to $T_m = 363.82$ K. Different curves correspond to different (ϵ_0, ϵ_s) values, while the sum of $\epsilon_0 + \epsilon_s$ is fixed to about 11.0 kJ/mol.

When temperature $T < T_m$, DNA is in the native phase, the fraction of intact base-pairs is non-zero. As the temperature approaches T_m from below, the number of intact base-pairs vanishes to zero; and when $T > T_m$ DNA is in the denatured phase, with very few hydrogen-bonded base-pairs. Figure 5 shows that the cooperativity of this denaturation transition depends on the relative importance of base-pairing and base-pair stacking interactions. If the energy of double-stranded DNA is mainly contributed by base-pairing interactions, then the denature transition will be a gradual processes (the bottom curve of Fig. 5); while if base-pair stacking energies contribute the main part of the energy, then the transition will be very cooperative (the top curve of Fig. 5). Experimentally, it was observed that the DNA thermal melting is a highly cooperative transition process [24, 94, 95], with a melting curve resembling the upper curves Fig. 5. When the temperature is slightly below T_m , most of the base-pairs in DNA is intact, while when T is slightly above T_m , most of the base-pairs are denatured. The Montanari-Mézard model with base-pair and base-pair stacking interactions is able to explain this highly cooperative phenomenon, if we assume that in DNA double-helix, base-pair stacking interactions are more important than base-pairing interactions.

One way to separate the contributions of base-pairing and base-pair stacking to the stability of DNA double-helix is to observe the melting curves of an ensemble of DNA molecules of the same nucleotide composition but different sequences. The total base-pairing interaction in these molecules are the same. Therefore, different melting behaviors must be due to differences in the base-pair stacking interactions. Recently, Montrichok and co-workers [96] combined UV spectroscopy and rapid quenching to demonstrate that the cooperativity of DNA denaturation is strongly sequence-dependent. For two DNA oligomers with similar A, T, G, and C compositions but with different sequence orders, one oligomer melts in a

two-state way, while the other melted through many intermediates.

A question of academic interest is the order of the denature transition. Although the transition could be very cooperative as shown in Fig. 5, it is predicted to be *continuous* according to the model of this subsection. No latent heat is needed to complete the melting transition process. This is also consistent with the result of Ref. [87] based on the de Gennes-Peyrard-Bishop model of the preceding subsection. The transition is continuous because both the free energy and its first-order derivative with respect to temperature T are continuous at T_m . DNA thermal melting is driven by the competition between energy and entropy. A denaturation bubble (which is composed of two unpaired segments) could have many different configurations and hence are entropy-favored; while base-paired regions are energy-favored. When temperature is high enough, the gain in configurational entropy out-performs the loss in pairing energy and the two strands are therefore separated.

In § IIC we will show that if self-avoidance effect is properly taken into account in the theoretical calculations, the melting transition could become first-order. Before doing this, however, let us first discuss force-induced DNA melting based on the present model.

2. Force-induced DNA melting

When there is an external force \mathbf{F} pulling apart the double-stranded polymer (see Fig. 4 (right)), the total partition function Eq. (11) should be changed to $\exp(\beta\mathbf{F} \cdot \mathbf{r})Z_{\text{ds}}(n, \mathbf{r})$. The generating function of the new system has the same form as the expression in Eq. (20). The only difference is that \mathbf{p} in Eq. (20) is replaced with $i\beta\mathbf{F}$.

For the new system, $z_{\text{coil}} = \tilde{\mu}^{-2}(i\beta\mathbf{F}) < 1$, and z_{native} is still given by the root of Eq. (23). When either the force or the temperature or both are large enough, the system is in the stretched coil state with two separated strands; when both the force and the temperature is small, the system is in the native paired state. The phase diagram in the force-temperature plan for a model DNA polymer made of 50% G-C pairs is shown in by the dashed line of Fig. 6. In this phase diagram, there is a critical force-temperature curve separating the native and the denatured phase. This phase diagram is determined by the equation

$$y(z_{\text{coil}}) = 1. \tag{25}$$

When the external force is nonzero, there exists a *discontinuous* first-order phase struc-

tural transition when the critical force-temperature curve is crossed. The phase transition is first-order because the free energy density is continuous, while its first-order derivative respect to temperature or force is discontinuous. The force-induced first-order phase transition is consistent with the de Gennes-Peyrard-Bishop model [87, 97]. Marenduzzo and co-authors [98] have studied an elegant three-dimensional directed lattice model of DNA force-induced unzipping. They also predicted a first-order phase transition.

According to the the phase-boundary (the dashed line) of Fig. 6, one would anticipate that that room temperature of $T = 300$ K, an stretching force of about 4 pN is enough to pull the two strands of DNA apart. This is not the case. According to the experimental work of Essevaz-Roulet et al. [37] and Rief et al. [47], at room temperature, an external force in the range 9 – 20 pN is needed to separated the DNA double-helix.

There are many reasons that may related to this quantitative disagreement between theory and experiment. For example, in the model we mentioned earlier, there is only an energetic contribution to base-pair stacking, this is denoted by the constant energy parameter ϵ_s . However, in the actual case, for the two consecutive base-pairs to stack onto each other, the DNA backbones must be twisted and folded. This confinement costs entropy, but this entropic effect is not considered in the model. Similarly, the free energy contribution to base-pairing is also assumed to be enthalpic in nature in our previous calculations. Unfortunately, to separate the enthalpic and entropic contributions to the free energy of pairing and stacking has not been an easy task and much work still needs to be done (see, for instance, Refs. [14, 25, 99, 100] and references cited therein). Here, to demonstrate the effect of taking into account of entropic contribution to base-pair stacking free energy, we separate the free energy of base-pair stacking into two parts, one part comes from the average stacking energy, and the other part comes from entropic penalties:

$$\epsilon_s = \delta H_s - T\delta S_s. \quad (26)$$

In the above equation, δH_s is the enthalpic difference between a stacked and an un-stacked pair of base-pairs, and δS_s is the entropic difference between a stacked and un-stacked pair of base-pairs [14, 25]. Under this modification, the new phase boundary is shown by the solid line of Fig. 6, which is in better agreement with experimental observations. In general, one should also consider the temperature dependence of the pairing free energy ϵ_0 .

A method of combining external force and temperature to measure various thermody-

dynamic parameters of DNA could be suggested. One can fix the environment at various temperatures and measure the unzipping critical force of DNA. The resulting phase-diagram can then be fitted with theoretical results based on the model of this section to determine a set of free energy parameters that leads to a good fit to the experimental phase-diagram. We anticipate to see such progress in the near future.

C. Poland-Scheraga model and excluded-volume effect

It is quite difficult to consider excluded-volume interactions in the de Gennes-Peyrard-Bishop model and the Montanari-Mézard model. To incorporate the effect of excluded volume on the melting transition of DNA, we first mention a simplified two-state model, which is due to Poland and Scheraga [77, 90]. In the the Poland-Scheraga model, each base-pair exists either in a base-paired state or in an un-paired open state. A segment of DNA that contains only paired base-pairs is called a double-helical segment, and a segment that is composed completely of opened base-pairs is referred to as a denatured segment. DNA could be regarded as a linear chain of double-helical segments and denatured segments, with these two types of segments occur in alternative order. When two DNA segments come close together each other, there are excluded volume interactions. Excluded volume interactions also exist between nucleotide bases in a single denatured DNA segment.

At first, let us discard any excluded volume effect and focus only on the base-pairing and base-pair stacking interactions. Without loss of generality, we assume the first and the last pair of bases are always in the paired state. Denote the partition function of a DNA double-helical segment of of i ($i \geq 1$) base-pairs by $Z_{\text{helix}}(i)$, and that of a denatured segment of j ($j \geq 1$) base-pairs as $Z_{\text{bubble}}(j)$. The whole DNA polymer may in the double-helix configuration, or there may be many double-helix segments and denaturation bubbles (see Fig. 7). The total partition function of a DNA of N base-pairs is obtained by summarizing the contributions from all possible configurations:

$$\begin{aligned}
 Z(N) = & Z_{\text{helix}}(N) + \sum_{s=1} \sum_{i_0=1} \sum_{j_1=1} \sum_{i_1=1} \cdots \sum_{j_s=1} \sum_{i_s=1} \delta_{\sum_k i_k + \sum_l j_l}^N \times \\
 & \times \prod_{m=1}^s [Z_{\text{helix}}(i_m) Z_{\text{bubble}}(j_m)] Z_{\text{helix}}(i_0). \tag{27}
 \end{aligned}$$

The generating function of Eq. (27) is

$$G(z) = \sum_{N=1} Z(N)z^N = \frac{G_{\text{helix}}(z)}{1 - G_{\text{helix}}(z)G_{\text{bubble}}(z)}, \quad (28)$$

where $G_{\text{helix}}(z) = \sum_{i=1} Z_{\text{helix}}(i)z^i$ and $G_{\text{bubble}}(z) = \sum_{j=1} Z_{\text{bubble}}(j)z^j$.

For a helical segment of i base-pairs, the free energy is $-i\epsilon_0 - (i-1)\epsilon_s$ due to base-pairing (ϵ_0) and base-pair stacking (ϵ_s). Therefore

$$G_{\text{helix}}(z) = \frac{ze^{\beta\epsilon_0}}{1 - ze^{\beta(\epsilon_0 + \epsilon_s)}}. \quad (29)$$

The partition function $Z_{\text{bubble}}(j)$ is estimated as follows. As mentioned in § II B, when neglecting all the possible interactions with a single strand, a single-stranded DNA could be modeled as a freely-jointed chain of bond length $b \simeq 1.7$ nm, with each bond corresponding to about three nucleotide bases. A DNA denaturation bubble of j base-pairs therefore corresponds to a freely-jointed chain of $2(j+1)/3$ bonds, with the additional constraint that the two ends of the chain should be separated no more than $2a$ apart, where a is the base-pairing interaction range (see § II B). Then

$$Z_{\text{bubble}}(j) = \Omega^{2(j+1)/3} \times \text{Probability}(|\mathbf{r}_1 + \dots + \mathbf{r}_{2(j+1)/3}| \leq 2a) \quad (30)$$

$$\simeq c_0 \frac{e^{js_{\text{coil}}}}{(j+1)^\gamma}, \quad \text{for } j \gg 1. \quad (31)$$

In Eq. (30), Ω denotes the total number of microscopic configurations of a bond; in Eq. (31), $\gamma = \frac{3}{2}$, $s_{\text{coil}}/2k_B$ is the configurational entropy of a single-stranded DNA base, and c_0 is a numerical constant. The scaling of the partition function of a denaturation bubble with bubble size, $j^{-3/2}$ is completely caused by entropic constraint of chain closure.

Based on Eq. (29) and Eq. (31), Eq. (28) is re-written as

$$G(z) = \frac{ze^{\beta\epsilon_0}}{1 - ze^{\beta(\epsilon_0 + \epsilon_s)} - zc_0e^{\beta\epsilon_0} \sum_{j=1}^{\infty} \frac{(ze^{s_{\text{coil}}})^j}{j^{3/2}}}. \quad (32)$$

According to Eq. (28), the free energy density of the system is

$$g_{\text{total}} = k_B T \ln z_s, \quad (33)$$

where z_s is the smallest positive singular point of function $G(z)$.

$G(z)$ has two singular points: one is at $z = \exp(-s_{\text{coil}})$ and the other is at the point where $1/G(z) = 0$. At low temperatures, the free energy of the whole system is determined by the

second singular point. As temperature is increased, the second singular point approaches the first one, marking the transition from a native DNA into a denatured DNA. The transition is second-order, because at the transition point g_{total} is continuous and its first-order derivative with respect to temperature is also continuous [90].

1. Excluded-volume effect

How will the phase-diagram of DNA denaturation be changed if excluded-volume effect is taken into account? From our preceding discussion, it is obvious that the value of γ in Eq. (31) could influence the order of the denaturation transition. If $\gamma > 2$ the transition is *first-order*, as Eq. (32) predicts.

Many analytical and computational efforts have been made to include excluded-volume effect in the Poland-Scheraga model. In 1966, Fisher [101] suggested a modified Poland-Scheraga model to include the excluded-volume effect between the two strands of each denaturation bubble. He found that $\gamma \simeq 1.75$ by this modification, so the transition is still second-order. Grassberger and co-workers [102] and Carlon and co-workers [103] performed numerical simulations on a lattice double-stranded polymer model. They have considered the self-avoidance effect in the simulation. The result of Refs. [102, 103] shows that, when excluded volume effect is fully considered, the DNA denaturation transition is *first-order*. The scaling exponent $\gamma \simeq 2.10$ as estimated by Carlon et al. [103]. Zhou [87] mapped the DNA melting problem to a quantum-mechanical problem and related the melting point with the vanishing of a confined state in the ground-state eigen vectors. However, because self-avoidance effect is only partially considered in the work of [87], the denaturation transition is still predicted to be continuous. In the model studied by Garel and co-authors [104], excluded volume interaction between the two chains of DNA was considered, and intro-chain repulsions were neglected. This partially simplified model was found to predict a first-order denaturation transition *in all dimensions* [104] (but see comment by Bhattacharjee [105]). To further understand the results of Ref. [104], Baiesi and co-authors [106] considered the statistics of two mutually avoiding random walks and found that the reunion scaling exponent of such a pair of chains is $\gamma > 2$ both in two and three dimensions.

Here, we mainly review the work of Kafri, Mukamel and Peliti [107, 108]. In the original Poland-Scheraga model, no excluded volume effect is considered. When excluded volume

effect is taken into account, the available configuration of a DNA loop will be decreased. This decrease in the available configurations comes from two sources: first, the excluded volume effect between the two strands of the loop [101]; second, the excluded volume effect between the loop and the rest of the DNA chain. Fisher has considered the excluded volume interaction between the two strands of the same loop, but it is not enough to account for the first-order phase transition. Kafri, Mukamel, and Peliti extended the work of Fisher to consider the interaction between one denaturation bubble and the rest of the DNA chain. They found that if this kind of interaction is considered, the scaling exponent $\gamma \geq 2.115$ [107], leading to a first order phase transition. They have arrived at this conclusion taken using of the scaling properties of polymer networks developed by Duplantier and co-workers [109–111].

The arguments of Kafri and co-authors are as follows [108]. Consider the configuration of Fig. 8. When the polymer is considerably long, all the segments in Fig. 8 could be regarded as flexible. These segments form a polymer network of four chains and one loop. There are also two vertices (A and D) of degree unit and two vertices (B and C) of degree three. Generally, Duplantier and co-workers [109–111] showed that, for a polymer network made of M segments of length l_1, l_2, \dots, l_M (the total length is $L = \sum_i l_i$), the total partition function has the following scaling form:

$$Z_g = L^{\gamma_g - 1} z\left(\frac{l_1}{L}, \frac{l_2}{L}, \dots, \frac{l_M}{L}\right), \quad (34)$$

where function $z(\dots)$ is a scaling function and γ_g is obtained by

$$\gamma_g = 1 - 3\nu N_{\text{loop}} + \sum_{k=1} n_k \sigma_k. \quad (35)$$

In Eq. (35), ν is the scaling exponent related with the radius of gyration of a self-avoiding chain in three dimensions, $\nu \simeq 0.588$ [111, 112]; N_{loop} is the total number of independent loops in the polymer network; n_k is the total number of vertices of degree k ; and σ_k is a scaling exponent associated with a vertex of degree k .

For the polymer network of Fig. 8, when $n \ll N$, Eq. (34) leads to

$$Z_g = (N + n)^{2\sigma_1 + 2\sigma_3 - 3\nu} z(n/N). \quad (36)$$

When $n/N \rightarrow 0$, Z_g should approach the partition function of a single polymer. Therefore, $g(n/N) \sim (n/N)^{2\sigma_3 - 3\nu}$. The total partition function then has the following scaling form:

$$Z_g = n^{2\sigma_3 - 3\nu} N^{2\sigma_1}, \quad (37)$$

indicating that the total partition function is the partition function for a single loop, n^γ (with $\gamma = 3\nu - 2\sigma_3$), and that of the rest of the polymer, $N^{2\sigma_1}$.

It was known that $\sigma_3 = -0.175$ [111], so $\gamma \simeq 2.115$ in Eq. (31). Since it is larger than 2, according to the Poland-Scheraga model we mentioned earlier, the denaturation phase transition will be first-order.

Because the loop size distribution Eq. (31) is geometric with $2 < \gamma < 3$, the average loop size will be finite when the temperature approaches the melting temperature from below. However, the variance of the loop size becomes infinite as the melting temperature is approached.

It should be remarked that, in the literature there is still debate concerned with the excluded volume effect between a denaturation loop and the remaining part of the DNA molecule (see, for example, [113, 114]). The theoretical reasoning of Refs. [107, 108] assume the polymer under study is extremely long, but cooperative denaturation behaviors were usually observed with DNA chains with about 10^2 - 10^3 base-pairs. As our analysis in § II B (and also the discussion in the book of Poland and Scheraga [77]) demonstrated, DNA denaturation could be very cooperative when base-pair stacking interaction is considered. It might be suggested that for DNA chains with length of the order of 10^3 base-pairs, base-pair stacking instead of excluded volume interaction is the main contribution to the cooperativity of denaturation. More experimental investigations are certainly needed to resolve this issue.

D. Force-induced unzipping of real DNA molecules

In § II A, § II B and § II C we have discussed the general physical aspects of DNA denaturation and unzipping. In this and the following subsections we discuss how a real DNA molecules behave under the action of an external force. First, we consider the case of long DNA molecules. In § II E we will focus on the kinetics of structural transitions of DNA oligomers.

Essevaz-Roulet, Bockelmann, and Heslot [37–39] performed the following unzipping experiment on single DNA molecules (Fig. 9). The two strands on one end of a phage- λ DNA (48502 base-pairs, total contour length $16.2 \mu\text{m}$) are separately attached to a glass slide and a very soft force lever (a micro-needle with elastic modulus about $1.7 \text{ pN}/\mu\text{m}$), while the other end of the DNA macromolecule was kept free (In Ref. [40], the soft micro-needle

was replaced with optical tweezer to further increase the precision). At room temperature $T = 300$ K and nearly physiological ionic conditions (150 mM NaCl), the two strands of DNA was separated by displacing the glass slide with constant velocity (in the range 20-800 nm/s). During the DNA unzipping process, the distances between the glass slide and the force lever were recorded in real time, as well as the bending extent of the force lever.

Heslot and co-workers [37–39] found that the typical force for DNA unzipping is in the range 10-15 pN. The force needed to unzip an GC-rich DNA segment is larger than that to unzip an AT-rich one, in consistent with the fact that a G-C pair has three hydrogen bonds while an A-T pair has only two. They found that the unzipping force is directly related with the GC contents: peaks of unzipping force correspond to peaks in the DNA GC contents. The experimental curves are highly reproducible, suggesting the unzipping experiment was probing the equilibrium unzipping properties of double-stranded DNA. The irregularities in the force-displacement curves in the experimental curves are not random noises. These irregularities contain information about the DNA sequence.

Trying to understand and reproduce the experimental unzipping force-extension profiles from theoretical calculations, an *equilibrium* slip-stick mechanism was suggested by Heslot and co-workers [38, 39]. Here we recall their model in some detail.

In the DNA-micro-needle system, there the following three energy terms:

1). the energy cost of separating j base-pairs, $E_{\text{DNA}}(j)$. Bockelmann et al. [38, 39] suggested the following form

$$E_{\text{DNA}}(j) = \sum_{\mu=1}^j E_{\text{pair}}(\mu), \quad (38)$$

where $E_{\text{pair}}(\mu)$ is the base-pairing energy of the μ th base-pair:

$$E_{\text{pair}}(\mu) = \begin{cases} E_{\text{G-C}}, & \text{for a G - C base - pair} \\ E_{\text{A-T}}, & \text{for an A - T base - pair} \end{cases} \quad (39)$$

Therefore, E_{DNA} is determined by two phenomenological parameters $E_{\text{G-C}}$ and $E_{\text{A-T}}$, and sequence-dependent interactions between neighboring base-pairs are averaged into these parameters. For the experiment with resolution in the range of about 100 base-pairs [37], a two-parameter approximation is reasonable.

2). The elastic energy stored in the two single-stranded DNA segments. When the external force is F , the average extension of a single-stranded DNA segment of j base-pairs

is

$$x(F, j) = jl_{\text{SS}} \left[\coth \left(\frac{Fb}{k_B T} \right) - \frac{k_B T}{Fb} \right] \left(1 + \frac{F}{S} \right). \quad (40)$$

where l_{SS} is the contour length of one nucleotide unit in the single-stranded DNA, and S is the stretching modulus of a single-stranded DNA [115]. In deriving Eq. (40) the single-stranded DNA chain is modeled as a freely-jointed-chain of Kuhn length b and stretching modulus S [115, 116]. The force needed at keeping the single-stranded DNA's extension fixed at x could be obtained by inversion of Eq. (40). The elastic energy to keep the extension of a single-stranded DNA at x is then

$$E_{\text{elastic}}(x, j) = \int_0^x F(x') dx' = F(x)x - \int_0^{F(x)} x(F', j) dF'. \quad (41)$$

3). The bending energy stored in the force lever,

$$E_{\text{lever}}(x_0) = \frac{k_{\text{lever}}}{2} x_0^2, \quad (42)$$

where x_0 is the total deflection of the tip of the force lever (Fig. 9).

The total energy of the system is

$$E_{\text{total}}(x_0, x_1, x_2, j) = E_{\text{DNA}}(j) + E_{\text{elastic}}(x_1, j) + E_{\text{elastic}}(x_2, j) + E_{\text{lever}}(x_0), \quad (43)$$

with the total displacement of the system fixed at $x = x_0 + x_1 + x_2$. The average force at total displacement x is

$$\langle F \rangle = \frac{\sum_{j=0}^N \int dx_1 \int dx_2 k_{\text{lever}}(x - x_1 - x_2) \exp[-\beta E_{\text{total}}(x - x_1 - x_2, x_1, x_2, j)]}{\sum_{j=0}^N \int dx_1 \int dx_2 \exp[-\beta E_{\text{total}}(x - x_1 - x_2, x_1, x_2, j)]}. \quad (44)$$

The theoretical force-displacement profile obtained with fitting parameter $E_{\text{G-C}} = 2.9 k_B T$, $E_{\text{A-T}} = 1.3 k_B T$, $b = 1.5$ nm, $S = 800$ pN, $Nl_{\text{SS}} = 30$ μm , and $T = 300$ K is in good agreement with experimental data [38, 39]. The fine details of the experimental unzipping curves could be quantitatively reproduced to some extent by this simple model.

A detailed analysis of the end-to-end distance of λ -DNA revealed the existence of many saw-tooth patterns [37]. Typically, when the stretching force increases, at first the distance between the two ends of DNA almost keeps constant (the stick phase). Then, as the increase in force is about 1-2 pN, suddenly the end-to-end distance of DNA is increased by about 100-200nm (the slip phase). As the displacement of the whole system keeps increasing with

a constant velocity, this stick-slip process repeats many times. Bockelmann et al. [38, 39] regarded this phenomenon as an equilibrium stick-slip process.

The stick-slip phenomenon in unzipping λ -phage DNA has its origin from jumping of the DNA “opening fork” between adjacent energy local minima. We explain this insight by an example. In Fig. 10 the minimum value of the total energy Eq. (43) as a function of the opening-fork position j is shown at two fixed displacements. When the total extension of the system is $x = 18.1 \mu\text{m}$, the energy absolute minimum is located at $j_1 \simeq 10050$. This energy absolute minimum is separated from a metastable state at $j_2 \simeq 10150$ by an energy barrier of $10 k_B T$ and another metastable state at $j_3 \simeq 10260$ by a barrier of $20 k_B T$. Between the two metastable states at j_2 and j_3 there is also an energy barrier of $15 k_B T$. At extension $x = 18.2 \mu\text{m}$, position j_3 now corresponds to the absolute energy minimum and position j_1 becomes metastable. As the total extension changes from x_1 to x_2 , the opening fork position will also tries to move from the old best position j_1 to the new best position j_3 . This is achieved by two steps, from j_1 to j_2 , and then from j_2 to j_3 . Thermal activation causes a nucleotide base-pair to switch frequently between the closed state and the opened state, the characteristic time scale is $3.6 \times 10^{-6} \text{s}$ [75, 117]. However, to jump from j_1 to j_2 , the thermal activation energy should be at least $10 k_B T$, the probability of which is $e^{-10} = 4.54 \times 10^{-5}$. Therefore, the typical time needed for the opening fork to jump from j_1 to j_2 is about 0.1s . Similarly, the typical needed time to jump from j_2 to j_3 is about 1.0s . In the experiment with displacement velocity 20 nm/s , it takes five seconds to move a distance of 100 nm . During this period it is very possible that the jumping process $j_1 \rightarrow j_2 \rightarrow j_3$ has happened once. When the $j_1 \rightarrow j_2$ or the $j_2 \rightarrow j_3$ hopping happens, about 100 base-pairs are unzipped cooperatively, leading to a rapidly drop in the pulling force.

For homogeneous DNA chains, the energy landscape is quite smooth and the resulting force-extension profile will be featureless.

In § II E, we will discuss more quantitatively why hopping processes such as that from position j_1 to position j_2 in Fig. 10 have a time scale of several seconds.

A practical issue is the possibility of using DNA unzipping to obtain sequence information of DNA. The analysis in this subsection suggests that, during DNA unzipping, the boundary between unzipped and base-paired DNA moves along DNA polymer by a hopping process, the hopping distance is of the order of 10^1 - 10^2 base-pairs. Therefore, one cannot attain single-nucleotide resolution by mechanical unzipping. On the other hand, as

demonstrated by Quake and co-workers [70], combining fluorescence and moving of DNA polymerase along single-molecule DNA chain could be used to obtain sequence information with single-nucleotide resolution. We expect more experimental and theoretical investigations along this line.

1. *Scaling relationship in DNA unzipping*

The unzipping experiment of Essevaz-Roulet et al. [37] was performed by fixing the total extension of the DNA-micro-needle system. It is also possible to pull the two strands of DNA apart with a constant force, as was done by Danilowicz et al. [41]. For an extremely long DNA under the action of constant external force, § II B 2 showed that there exists a threshold force F_c . At given temperature, when the external force is lower than F_c , most of the base-pairs of DNA are in the closed state; while when the external force is larger than F_c , the two strands of DNA are separated completely and there is no closed base-pairs. It is interesting to study the scaling behavior of the number of opened base-pairs as the external force approaches the transition force F_c from below.

First, let us consider the unzipping of a homogeneous DNA chain. Because of the first-order nature of the unzipping transition, the probability that n base-pairs be opened is proportional to $\exp(-n\delta\epsilon_0)$, where $\delta\epsilon_0 \sim (F_c - F)$ is the free energy difference of a base-pair in the paired state and in the opened state. Therefore, as $F \rightarrow F_c$, the average number of opened base-pairs scales as

$$\langle n \rangle = \frac{\sum_{n=1}^{\infty} n \exp(-n\delta\epsilon_0)}{\sum_{n=1}^{\infty} \exp(-n\delta\epsilon_0)} \sim (F_c - F)^{-1}. \quad (45)$$

When heterogeneity in a real DNA sequence (such as a λ -phage DNA) is considered, will the scaling relationship Eq. (45) still hold? Lubensky and Nelson [118] argued that the average number of opened base-pairs divergences should even faster, with a scaling form of $\langle n \rangle \sim (F_c - F)^{-2}$! This is a surprising conclusion. Does the conclusion of Lubensky and Nelson [118] indicate that, as $F \nearrow F_c$ the free energy difference between a base-pair in the closed state and that in the opened state proportional to $(F_c - F)^2$? However, this should not be true, because the unzipping transition should be *first-order* even in the presence of sequence heterogeneity.

Lubensky and Nelson [118] showed that, for heterogeneous DNA Eq. (45) is still true, but it is just one part of the contribution. The total number of opened base-pairs has another contribution due to sequence heterogeneity. As $F_c - F$ is much smaller than a characteristic force determined by the sequence heterogeneity, this additional contribution, which scales as $(F_c - F)^{-2}$, becomes more important than Eq. (45).

Here, we explain the above-mentioned insight of Lubensky and Nelson in some detail. A more quantitative analysis could be found in the paper of Lubensky and Nelson [118].

When there is sequence randomness, the free energy increase caused by breaking n base-pairs is

$$\delta E = n\delta\epsilon_0 + \sum_{i=1}^n \delta\epsilon_i, \quad (46)$$

where $\delta\epsilon_0 \sim (F_c - F)$ is the average increase in the free energy caused by breaking one base-pair, and $\delta\epsilon_i$ is the difference between the pairing energy of the i -th base-pair with respect to the average pairing energy. For simplicity, we model $\delta\epsilon_i$ as an independent and identically distributed Gaussian variable with standard deviation σ :

$$P(\delta\epsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\delta\epsilon_i)^2}{2\sigma^2}\right]. \quad (47)$$

Figure 11 demonstrates the energy landscape of breaking n base-pairs of a randomly generated DNA sequence according to Eqs. (46) and (47). The first term of Eq. (46) causes a constant slope of $\delta\epsilon_0$ to the free energy profile; while the second term of Eq. (46) causes a deviation from this slope, the magnitude of this deviation scales as $\sqrt{n}\sigma$. Because of this random deviation, the free energy minimum is not located at $n = 0$, but is shifted to n_{optimal} (in Fig. 11 $n_{\text{optimal}} \simeq 500$). This optimal number of base-pairs opened is expressed by the following equation:

$$n_{\text{optimal}} \sim \frac{\sigma^2}{4(\delta\epsilon_0)^2}. \quad (48)$$

Around the n_{optimal} the free energy has a well of order of $-\frac{\sigma^2}{4\delta\epsilon_0}$, which is proportional to σ^2 . When the number of broken base-pairs goes beyond n_{optimal} , the first term of Eq. (46) dominates and the free energy increases linearly with n . The total number of broken base-pairs is then n_{optimal} plus Eq. (45):

$$\langle n \rangle \sim \frac{\sigma^2}{(\delta\epsilon_0)^2} + \frac{1}{\delta\epsilon_0}. \quad (49)$$

When the first term of Eq. (49) is less important than the second term, then Eq. (45) holds; however, when the first term outperforms the second term, then

$$\langle n \rangle \sim \frac{1}{(F_c - F)^2}. \quad (50)$$

To observe the scaling form Eq. (50), one must have strong sequence heterogeneity ($\sigma \gg 0$ in Eq. (49)), or the external force is very close to the critical force F_c . Otherwise, the scaling form Eq. (45) will be observed. For the actual case of random sequence DNA, the binding energy of A-T and G-C pairs differ about $1 k_B T$, which is considerable, so the scaling form Eq. (50) instead of Eq. (45) should be observed. Lubensky and Nelson have performed Monte Carlo simulations for several different heterogeneous DNA molecules [119] and their results confirmed the validity of Eq. (50).

E. Unzipping kinetics of DNA/RNA helix-loops under constant force

We mentioned that the hopping processes in unzipping λ -phage DNA occur with a characteristic time scale of *seconds*. Recently, single-molecule unzipping experiments were also performed on RNA [51, 52]. A helix-loop RNA of 49 bases, P5ab, has a native structure of a hairpin with 22 base-pairs, plus a hairpin loop of four bases and a single-nucleotide bulge-loop. Liphardt et al. [51] revealed that, under a constant force of about 15 pN, P5ab switched between two well-defined “open” and “close” states on a time scale also of *seconds*.

On the other hand, experimental measurement of Altan-Bonnet and co-authors [75] reported that, at room temperature of $T = 300 K$, the closing-opening transition of a single nucleotide base-pair takes place on a time scale of $10^{-6} s$. In § II D we have briefly mentioned that this huge gap in time scales from microseconds to seconds is caused by the existence of high energy barriers. In this subsection we look into this issue more quantitatively.

Cocco and co-authors [117] presented a semimicroscopic model to explain the above-mentioned huge gap in transition time scales. We review the calculations of Ref. [117] by focusing on the following model system: a hairpin-loop made of 12 G-C base-pairs and a 4-base ATAT loop. The two ends of this single-stranded DNA are pulled apart by a constant force of magnitude f .

Model and dynamic rules Under the pulling of a constant force f , the free energy

associated with unzipping n bases is

$$G(n, f) = \sum_{i=1}^n \Delta g(i, f), \quad (51)$$

where $\Delta g(i, f)$ is the free energy change due to opening the i -th base-pair at the boundary between the open and closed region,

$$\Delta g(i, f) = 2g_s(f) - g_0(i). \quad (52)$$

In Eq. (52), $g_0(i)$ is the free energy of opening base pair i . For the poly(GC) hairpin, the following values are used [117]:

$$\begin{aligned} g_0(i) &= g_0 = -3.442 k_B T, \quad (1 \leq i \leq 11) \\ g_0(12) &= g_0 + g_{\text{loop}} = 3.062 k_B T, \\ g_{\text{loop}} &= 6.504 k_B T. \end{aligned}$$

$g_s(f)$ is the free energy of an opened DNA base. For simplicity, single-stranded DNA is regarded as a freely-jointed chain, without considering its deformability:

$$g_s(f) = -k_B T \frac{l_{SS}}{b} \ln \left[\frac{\sinh \beta f b}{\beta f b} \right], \quad (53)$$

where $l_{SS} = 5.6 \text{ \AA}$, $b = 15 \text{ \AA}$ [117].

The following semimicroscopic transition rates are assumed for each elementary step in the structural transition path. At time t , suppose the opening fork is at base-pair position n (the n -th base-pair is opened but the $n + 1$ -th base-pair is closed). The movement of this boundary is governed by an opening rate r_o and a closing rate r_c :

$$r_o(n) = r e^{g_0(n)/k_B T}, \quad (54)$$

$$r_c(f, n) = r e^{2g_s(f, n)/k_B T}. \quad (55)$$

In Eq. (54) and Eq. (55), r is essentially the microscopic rate for a base pair to move together or apart in the absence of tension or base-pairing interactions. Its value can be estimated by considering the inverse self-diffusion time for an object of nanometer scale [117]: $r = k_B T / 2\pi\eta\ell^3 \approx 5 \times 10^6 \text{ s}^{-1}$, with $\ell = 5 \text{ nm}$, $\eta = 0.001 \text{ Pa sec}$ (the viscosity of water). Another way to estimate r is to fit the experimental data of Liphardt et al. [51] on P5ab unzipping. This fit gives $r = 3.6 \times 10^6 \text{ s}^{-1}$ [117].

The probability that at time t , the opening fork is at base-pair n is $\rho_n(t)$. It is governed by the following Markovian equation

$$\begin{aligned}\frac{\partial \rho_n(t)}{\partial t} &= -[r_o(n+1) + r_c(f, n)]\rho_n(t) + r_o(n)\rho_{n-1}(t) + r_c(f, n+1)\rho_{n+1}(t) \\ &= -\sum_{m=0}^N T_{n,m}\rho_m(t),\end{aligned}\tag{56}$$

where the matrix $T_{n,m}$ ($0 \leq n, m \leq 12$) is:

$$\begin{aligned}T_{n,n} &= r_o(n+1) + r_c(f, n), \\ T_{n,n+1} &= -r_c(f, n+1), \\ T_{n,n-1} &= -r_o(n),\end{aligned}$$

and all the other elements are zero.

Switching kinetics For the poly(GC) hairpin, the energy landscape as a function of opening fork position n is saw-tooth shaped. At force $f = 15.95$ pN, the completely closed and the completely opened states have the same free energy. The highest energy state is at $n = 11$ base-pair, with a energy barrier of $10.17 k_B T$.

The structural fluctuation of the poly(GC) hairpin is simulated based on the above-mentioned model. Figure 12 records the opening fork position as a function of time at fixed external force 15.95 pN. Figure 13 and Fig. 14 demonstrate the life time distribution for the polymer to be in the opened state and the closed state, respectively. Both of them obey exponential distribution. The mean life time of a closed state is $\bar{t}_{\text{mean close}} = 0.325$ s and that of an opened state is $\bar{t}_{\text{mean open}} = 0.208$ s; therefore $\bar{t}_{\text{mean close}}/\bar{t}_{\text{mean open}} = 1.561$. These data are in close agreement with data obtained by equilibrium calculations mentioned below.

The eigenvalues of the matrix $T_{m,n}$ in Eq. (56) are listed below:

$$\begin{array}{ll}1. 0 & 2. 2.233 \times 10^{-6}r \\ 3. 1.748 \times 10^{-2}r & 4. 3.175 \times 10^{-2}r \\ 5. 5.251 \times 10^{-2}r & 6. 7.790 \times 10^{-2}r \\ 7. 1.060 \times 10^{-1}r & 8. 1.346 \times 10^{-1}r \\ 9. 1.616 \times 10^{-1}r & 10. 1.850 \times 10^{-1}r \\ 11. 2.032 \times 10^{-1}r & 12. 2.146 \times 10^{-1}r \\ 13. 2.145 \times 10^1r & \end{array}\tag{57}$$

Eigenvalue zero corresponds to equilibrium distribution of the opening fork position; the second eigenvalue corresponds to the transition between open and closed states and it is equal to $8.04s^{-1}$. All the other eigenvalues of the order $10^{-1}r-10^1r$ and correspond to internal structural transitions within the open and the closed states.

In the work of Cocco and co-authors [117], besides poly(GC) hairpins and P5ab RNA, some other real RNA molecules with more complex native structures were studied. The unzipping kinetic of λ -phage DNA under the action of a constant force was simulated [117]. In all these examples, unzipping at constant force proceeds by successive steps, each corresponding to the crossing of a free energy barrier. This is similar with the conclusion of § II D. Therefore, unzipping experiments performed at constant external force are also unlikely to access DNA sequence information with single-nucleotide resolution.

In Ref. [120], Cocco, Monasson, and Marko further considered the effect of torque contributions in the double-helix part and showed this torque could influence the unzipping and re-zipping kinetics when the force increase rate is above certain threshold which is comparable with the relaxation time of torques in double-stranded DNA. In Ref. [121] the force and kinetic barriers of initialization of DNA unzipping were studied detailedly.

In Ref. [122], Ritort et al. used a two-state kinetic model to study the unzipping dynamics of RNA and the relationship between work and load velocity. The dynamics of DNA unzipping was also studied by Marenduzzo et al. [123] by a lattice model. The relation between number of opened base pairs and time was obtained by simulation and was found to obey a scaling law.

In our above discussion, we have not considered the possibility of mis-matched configurations. In RNA hairpins such mis-matched states are likely to occur. The folding-unfolding kinetics of RNA hairpins in the absence of external force was studied by Chen and Dill [124] and by Zhang and Chen [125]. It was found [124, 125] that when the environmental temperature is higher than a glass-transition temperature T_g , the folding-unfolding kinetics is two-state-like, similar as the result shown here. When the temperature is lower than T_g , the molecule is trapped in one of many low-free energy intermediate states, and the folding-unfolding corresponds to trapping-detrapping transition.

A practical application of unzipping DNA hairpins maybe to measure the thermodynamic parameters of DNA base-pairs. The experimental setup is relatively simple, and as demonstrated in this subsection, quantitative fitting of experimental data with theoretical

calculations is easy to perform.

III. SECONDARY STRUCTURES OF RNA POLYMER

In biological bodies RNA usually exists as a single-stranded biopolymer. Similar as DNA, it is also formed by four types of nucleotide bases, G, C, A, and U (Uracil); and two types of canonical Watson-Crick base-pairs, A-U and G-C, can be formed. The non-Watson-Crick base-pairing pattern G-U is also quite common in RNA structures.

Some kinds of RNA polymers, such as transfer-RNA (tRNA), have specific stable three-dimensional native configurations. Such configurations are formed when the RNA chain fold up onto itself to form many base-pairs between its complementary nucleotide bases. Such kind of RNA sequences should have been “designed” to make sure that there is only one native configuration, the free energy of which being much lower than those of other possible configurations. The sequences of tRNA molecules are obviously highly designed by natural selection. On the other hand, the tRNA sequences of different organisms also show some extent of variability. Investigation of this variability may shed some light on the evolutionary history of different organisms.

However, for many other RNA molecules including message-RNA (mRNA) it is undesirable to have stable structures. Message-RNA prefers to be in a random coil conformations with very few base-pairing interactions, so that its information can be easily read by the protein synthesis machine (the ribosome). The coding region of a genome must have been evolved to ensure that, (1) its corresponding mRNA transcript contains very few mutually complementary segments and, (2) its encoding protein can fold into a stable three-dimensional native configuration. This is a very difficult combinatorial optimization problem in biology and nature took an extremely long time to find the right answers. Today, identifying and interpreting correlations and anti-correlations in a DNA sequence becomes a very important subfield of scientific research. To meet this challenge, a structural biology viewpoint will turn out to be essential.

In this section we review recent works on the structural properties of RNA polymers. We focus on the stability of RNA secondary structures. By the secondary structure we mean the following picture [126–128]. The position of each nucleotide base along RNA chain with respect to the 5'-end of RNA is specified by an integer index i . Suppose in a RNA secondary structure, nucleotide base at position i and position j (with $j > i$) form a base-pair denoted as pair (i, j) , and nucleotide base at position k and position l ($l > k$) forms another base-pair

(k, l) . Then it is required that base-pairs (i, j) and (k, l) are either *independent*, with $j < k$, or *nested*, with $i < k < l < j$.

In an actual RNA configuration, some base-pairs may violate these conditions. These base-pairs are classified into the higher-order tertiary structure of RNA. For example, in the three-dimensional yeast tRNA^{Phe} there are base-pairing interactions between two RNA hairpin loops. These additional base-pairs stabilize the twisted L conformation of tRNA [13]. However, for RNA the main contribution to the free energy is from its secondary structures. Energy contributions from RNA tertiary structures could be regarded as perturbations to the free energy of the system. This separation of energy scales makes the study of RNA secondary structure of particular interest.

We discuss in § III A the structure of RNA polymers made of homogeneous nucleotide sequences. In § III B we show the effect of weak sequence design on RNA secondary structure. Glassy transition in RNA structures is discussed in § III C; and the influence of electrostatic interactions to RNA structural stability is considered in § III D. Section III E focuses on the issue of predicting the most stable secondary structure of a given RNA sequence.

A. The homogeneous-sequence approximation

As a starting point, this subsection neglects sequence heterogeneity and regards RNA as a homogeneous chain, with short-ranged attractive interactions between its constructing units.

When a RNA folds back onto itself and nucleotide bases form base-pairs, many double-helical segments appear. For a RNA made of random sequences (referred to as a random-sequence RNA hereafter), because the correlation along the RNA chain is very weak, each such double-helical segment in general is very short (of several base-pairs). According to recent experimental work of Ref. [75], at physiological temperature double-helical segment of 2-10 base-pairs is very easy to be disturbed by thermal noise. Therefore, secondary structures of a random-sequence RNA are transient, with typical life-time of the order of microseconds. The polymer could easily fluctuate from one secondary structure to another when the environmental temperature is not too low. Entropy plays an important role here.

Since mRNA could be well regarded as a random-sequence in some sense, the results of this subsection may have direct relevance to mRNA structural stability.

When one end of the RNA is fixed and the other end is pulled with an external force, it may be energetically unfavorable for the polymer to fold back and form double-helical segments. At some critical force, there is a so-called helix-coil structural transition, where the polymer is stretched considerably straight. It is of interest to estimate the magnitude of this critical force.

As a first step to understand the elastic property of a random-sequence RNA, we average over the sequence randomness and regard RNA as an effective chain with homogeneous sequence. Our treatment will be improved in § III B by considering the effect of weak sequence design; and in § III C we discuss how the homogeneity approximation of this subsection breaks down at low temperatures.

Montanari and Mézard [88] have studied the structural transitions in a random-sequence RNA by analytical means. In the treatment of Ref. [88], sequence randomness was smoothed out and replaced with a (weak) short-ranged average interaction between two nucleotide bases. The RNA chain was modeled as $n + 1$ beads connected sequentially by n deformable chemical bonds. Each unit (a bead plus a bond) of the model corresponds to several nucleotide bases of a real RNA chain. Each bead can form a “base-pair” with at most one another bead at the same time. In the calculation of Ref. [88], the base-pairing interaction was the only interaction considered. The excluded volume effect was neglected, and the possible base-pair stacking interactions were also neglected. It is believed that excluded volume effect is irrelevant in high salt conditions but becomes important at low salt conditions (see § III D). The pairing energy is short-ranged with range a and strength ϵ_0 , as was assumed in § II B. The deformable bond between two beads has the following bond length distribution [see Eq. (10)]:

$$\mu(\mathbf{u}) \propto \exp\left(-\frac{(|\mathbf{u}| - b)^2}{2l_0^2}\right), \quad (58)$$

where the bond length fluctuation l_0 is much shorter than the intrinsic bond length b .

Because the RNA chain is modeled as a homopolymer, the free energy and the partition function are sequence-independent and are functions only of the polymer length. Denote the partition function of a RNA chain of n bonds as $Z_0(n, \mathbf{r})$, where \mathbf{r} is the end-to-end distance vector of the chain. According to Fig. 15 we have the following recursive equation [$Z_0(0, \mathbf{r}) = \delta(\mathbf{r})$]:

$$Z_0(n, \mathbf{r}) = \int d\mathbf{u}_1 d\mathbf{u}_2 \mu(\mathbf{u}_1) Z_0(n-1, \mathbf{u}_2) \delta(\mathbf{r} - \mathbf{u}_1 - \mathbf{u}_2) +$$

$$\begin{aligned}
& + \Theta(a - |\mathbf{r}|)(e^{\beta\epsilon_0} - 1) \int \left[\prod_{i=1}^3 d\mathbf{u}_i \right] \mu(\mathbf{u}_1)\mu(\mathbf{u}_2)Z_0(n-2, \mathbf{u}_3)\delta\left(\mathbf{r} - \sum_{i=1}^3 \mathbf{u}_i\right) + \\
& + \sum_{l=0}^{n-3} \int \left[\prod_{i=1}^3 d\mathbf{u}_i \right] Z_0(n-l-3, \mathbf{u}_1)\mu(\mathbf{u}_2) \times \\
& \times \Theta(a - |\mathbf{r}|)(e^{\beta\epsilon_0} - 1) \int \left[\prod_{j=4}^6 d\mathbf{u}_j \right] \mu(\mathbf{u}_4)\mu(\mathbf{u}_5)Z_0(l, \mathbf{u}_6)\delta\left(\mathbf{u}_3 - \sum_{j=4}^6 \mathbf{u}_j\right), \quad (59)
\end{aligned}$$

where $\Theta(x)$ is the Heaviside step function defined in § II B 1.

With the same procedure as that in § II B 1, we can perform a Fourier transform for the partition function $Z_0(n, \mathbf{r})$ and obtain an iterative expression for the Fourier transformed partition function $\tilde{Z}_0(n, \mathbf{p})$. The generating function of $\tilde{Z}_0(n, \mathbf{p})$ can then be calculated following the method of § II B 1. The result reads [88]

$$G_0(z, \mathbf{p}) = \sum_{n=0}^{\infty} \tilde{Z}_0(n, \mathbf{p})z^n = \frac{1}{z} \frac{\omega(z)}{1 - \tilde{\mu}(\mathbf{p})\omega(z)}. \quad (60)$$

In Eq. (60), $\omega(z)$ is determined by the equation

$$\omega(z) = z + z^2 B(\omega), \quad (61)$$

with

$$B(\omega) = \frac{a^3}{6\pi^2} (e^{\beta\epsilon_0} - 1) \int d\mathbf{q} \tilde{\mu}^2(\mathbf{q}) \frac{\omega}{1 - \tilde{\mu}(\mathbf{q})\omega}, \quad (62)$$

and [see also Eq. (18)]

$$\tilde{\mu}(\mathbf{p}) = \frac{\sin b_0 |\mathbf{p}|}{b_0 |\mathbf{p}|} \exp\left(-\frac{l_0^2 |\mathbf{p}|^2}{2}\right). \quad (63)$$

When an external force \mathbf{F} is pulling the two ends of a RNA polymer apart, the generating function of the total partition function has the same form as Eq. (60), but with \mathbf{p} replaced by $i\beta\mathbf{F}$ (see § II B 2 for an explanation). A careful analysis [88] shows that in this case, Eq. (60) has two singular points, one corresponds to the point z_{coil} at which $\omega(z_{\text{coil}})\tilde{\mu}(i\beta\mathbf{F}) = 1$, and the other singular point is at z_{molten} , which is determined by $z_{\text{molten}} = \max_{\omega} \left[\frac{2\omega}{1 + \sqrt{1 + 4\omega B(\omega)}} \right]$.

At given temperature T , when the external force is less than some threshold value $F_c(T)$, the polymer is in the so-called *molten* phase [91], with the configuration of RNA fluctuating frequently between many transient secondary structures. The average extension of the polymer is zero. When the force is larger than $F_c(T)$, the polymer is in the stretched coil phase, with few base-pairing interactions. The threshold force F_c at given temperature is determined by the equation

$$z_{\text{coil}}(F_c, T) = z_{\text{molten}}(T). \quad (64)$$

When $|\mathbf{F}| > F_c(T)$, the average extension of RNA in units of its total contour length Nb is calculated by

$$\text{Ex} = -\frac{1}{b}k_B T \frac{\partial \ln z_{\text{coil}}}{\partial |\mathbf{F}|}. \quad (65)$$

The experimental and theoretical force-extension profile of a random-sequence RNA are shown in Fig. 16 (the solid line and symbols), where the solid line is obtained through Eq. (65).

In a random-sequence RNA, the main interaction is base-pairing interactions. The solid curve of Fig. 16 suggests that the Montanari-Mézard model could quantitatively explain the elastic response at high salt conditions of single-stranded charomid DNA [2, 34]. At room temperature $T \simeq 25^\circ\text{C}$ and high ionic conditions (for example, 50 mM Na^+ and 5 mM Mg^{2+} buffer), when the external force is less than about 0.3 pN, the average end-to-end distance of RNA approaches zero, and the RNA molecule is in the collapsed molten phase. When the external force is increased beyond this level, the secondary pairing structures in RNA are gradually pulled out and an increasing part of the polymer is in the stretched coil state. This gradual increase of coiled segment length with external force leads to the gradual force-extension profile of Fig. 16, without the appearance of a force plateau.

At the phase boundary $F_c(T)$, there is a continuous (*secondary-order*) molten-coil structural phase transition [88]. This transition is secondary-order, because at the transition point, the free energy and its first-order derivative are both continuous [77, 89, 90, 108]. The continuous nature of this phase transition can also be predicted by a simple two-state model [91]. The argument is as follows.

The RNA configuration could be divided into three parts: two stretched coil segments at the ends and a (complicated) hairpin segment in the middle. The partition function of a hairpin segment of m bases has the following scaling form:

$$Z_0^{\text{hairpin}}(m) \sim m^{-\frac{3}{2}} \exp\left(-n \frac{g_{\text{hairpin}}}{k_B T}\right), \quad (66)$$

where g_{hairpin} is the (model-dependent) free energy density of a RNA hairpin segment. Equation (66) will be derived in § III C. Because of the scaling exponent of 3/2 in Eq. (66), we know from the Lifson argument [89] (see § II C) that the transition between stretched coil state and hairpin state must be continuous. In both the continuous model [88] and the discrete model [91] mentioned here, self-avoidance effect is completely neglected. It is not yet completely clear whether the inclusion of this effect will qualitatively change the above

mentioned picture, as in the case of DNA thermal denaturation (see Sec. II). However, the quantitative agreement between the theoretical result of the model of Ref. [88] and the experimental observations (See Fig. 16) suggests that self-avoidance effect may not be so important here, since the scaling exponent of $3/2$ in Eq. (66) is not caused by the formation of loops but by the constraint of secondary structures.

1. *Elasticity of real homogeneous RNA chains*

The homogeneous chain model studied by Montanari and Mézard has also real correspondence. For example, a chain with sequence poly(GC), poly(AU), or poly(CAG), etc., can be regarded as a homogeneous chain.

When such a “truly” homogeneous RNA molecule is stretched with an external force, the force-extension profile is much more sharper than the solid curve of Fig. 16. For example, Rief and co-workers [47] studied the elastic responses of single poly(AT) and poly(GC) nucleotide chains and found that the force-extension profile has a marked plateau at force 9 ± 3 pN for poly(AT) and 20 ± 3 pN for poly(GC). The existence of a force plateau suggests a discontinuous structural transition process. Pulling a poly(AT) or poly(GC) single-stranded DNA is just like pulling apart the two strands of a double-stranded DNA, indicating that in the absence of external force, the structure of a single-stranded poly(GC) or poly(AT) chain can not fluctuate between many transient configurations. It forms a giant hairpin with an very long stem region and very few single-stranded loop regions. The entropic effect is suppressed.

However, based on the Montanari-Mézard model, it was found [93] that, even when the pairing energy ϵ_0 is very high the resulting force-extension curve is gradual and not abrupt. This suggests that the entropy is suppressed not by high base-pairing potentials and must be attributed to other effects including base-pair stacking interactions [92].

In the Montanari-Mézard model [88] base-pairing stacking interaction is neglected. This may be a good approximation for a random-sequence RNA; but it may not be valid for a highly homogeneous polymer such as poly(GC), since the probability that two base-pairs to be adjacent to each other in a poly(GC) will be much higher than in a random-sequence RNA. In Refs. [92, 93] base-pair stacking interactions were considered to explain the force-extension profile of a homogeneous RNA chain. When two formed base-pairs are adjacent

to each other, i.e., a (i, j) base-pair and a $(i + 1, j - 1)$ base pair, then there is an additional energy gain of magnitude ϵ_s (J in Fig. 16) [93].

The actual theoretical calculation is similar with that outlined in § IIB 1 and is detailedly given in Ref. [93]. Here we just mention the salient results. The effect of increasing base-pair stacking interaction is evidently demonstrated in Fig. 16. When the base-pairing potential ϵ_0 is kept fixed and stacking potential ϵ_s (or J) is increased, the force-extension becomes more and more flat. When $\epsilon_s \simeq 6.0 k_B T$ ($T = 300$ K) corresponding to an average stacking energy of $2 k_B T$ between two adjacent nucleotide base-pairs, a force plateau at 10 pN becomes very evident. Therefore, the cooperativity of hairpin-coil transition in a homogeneous RNA could be explained by base-pair stacking interactions. This conclusion reinforces the insight of § IIB, where we show that the cooperativity of DNA denaturation is also caused by base-pair stacking interactions.

Similar pictures can also be obtained by numerical Monte Carlo calculations with including electrostatic interactions. The quantitative comparisons of the experimental data and the modeling calculations will be displayed at IIID (see Figure 25).

B. structural transitions in a weakly designed RNA polymer

In the preceding subsection, the elastic property of a random-sequence RNA is studied. In this subsection, we study the elastic property of a long RNA chain which has a preferred native hairpin configuration [91]. The polymer is weakly designed in the following sense. First, generate a random RNA sequence of $2n + 1$ nucleotide bases. Second, for index $i = 1$ till n , with probability $p \leq 1$, change the nucleotide base at position $2n + 1 - i$ to be complementary to the base at position i and with probability $1 - p$ keep the nucleotide base at position $2n + 1 - i$ unchanged. Energetically, the RNA polymer has a designed preferred configuration of to form a long hairpin of $2n + 1$ bases. The designability is controlled by the parameter p . When p is much less than 1, the chain is only weakly designed. In this section, we mainly discuss this situation. The pairing energy is smoothed out along the polymer and the following pairing energy is assumed for the polymer chain:

$$\epsilon_{i,j} = -\epsilon_0 - (\epsilon_1 - \epsilon_0)\delta_{i+j}^{2n+1}, \quad (67)$$

with $\epsilon_1 > \epsilon_0$. Of course the difference $\epsilon_1 - \epsilon_0$ is controlled by the parameter p .

The thermal denaturation behavior of such system was first studied by Bundschuh and Hwa [91]. They concluded that when the temperature is high enough, the polymer will be in a molten phase with partition function governed by Eq. (66), and the polymer is not trapped by just one configuration. However, when the temperature is low, the polymer is in the designed hairpin configuration because of the energy preference of this state. There exists a *second-order* structural transition between the hairpin and the molten states at some critical temperature T_c .

In this subsection, the work of Ref. [91] is extended to consider also the effect of external force. The general picture is given in the phase diagram of Fig. 17. When both the temperature and the external force are small, the system is in the designed hairpin phase; when temperature is small but the external force is large, the system is in the stretched coil phase; when the temperature is high but the external force is small, the system is in the molten phase. The hairpin-coil transition is *first-order*; the molten-coil transition is *second-order*, and the hairpin-molten transition is also *second-order*.

Here we show the main steps of the analytical calculations which lead to the above-mentioned conclusions. We use the same continuous model as in the preceding subsection.

First, we define the following notations: (1) $Z_{\text{paired}}(n, \mathbf{r})$ is the total partition of a designed RNA of $2n$ effective bonds, with the bases at the two ends forming a base-pair of distance \mathbf{r} ; (2) $Z_{\text{unpaired}}(n, \mathbf{r})$ is the total partition function of a designed RNA of $2n$ bonds, with the two bases at the ends separated with distance \mathbf{r} ($|\mathbf{r}| > a$, the bonding distance); (3) $Z(n, \mathbf{r})$ is the total partition function of a designed RNA of $2n$ bonds.

Based on Fig. 18, we have the following function for $Z_{\text{paired}}(n, \mathbf{r})$:

$$Z_{\text{paired}}(n, \mathbf{r}_1) = \sum_{m=1}^n \sum_{l_1=1} \dots \sum_{l_m=1} \delta_{l_1+\dots+l_m}^n \int d\mathbf{r}_2 \dots d\mathbf{r}_{m+1} e^{-m\beta\epsilon_1} \times \\ \times \prod_{i=1}^m [\theta(a - |\mathbf{r}_i|) Z_{\text{molten-bubble}}(l_i, \mathbf{r}_i, \mathbf{r}_{i+1})] \delta(\mathbf{r}_{m+1}), \quad (68)$$

where $Z_{\text{molten-bubble}}(l, \mathbf{r}_1, \mathbf{r}_2)$ is the partition function of a loop of $2l$ bonds. From Eq. (68), the partition function $Z_{\text{paired}}(n) = \int d\mathbf{r} Z_{\text{paired}}(n, \mathbf{r})$ can be obtained. Finally, we obtain the following generating function for this partition function:

$$G_{\text{paired}}(z) = \sum_{n=1}^{\infty} z^n Z_{\text{paired}}(n) = \frac{\frac{4\pi}{3} a^3 e^{\beta\epsilon_1} G_{\text{molten-bubble}}(z)}{1 - \frac{4\pi}{3} a^3 e^{\beta\epsilon_1} G_{\text{molten-bubble}}(z)}, \quad (69)$$

where $G_{\text{molten-bubble}}(z) = \sum_{l=1} z^l Z_{\text{molten-bubble}}(l, \mathbf{0}, \mathbf{0})$. Clearly, $Z_{\text{molten-bubble}}(l, \mathbf{0}, \mathbf{0})$ should be independent of the value of ϵ_1 . To find the expression for $G_{\text{molten-bubble}}(z)$ we follow

the idea of Ref. [91] to set ϵ_1 to be identical to ϵ_0 in Eq. (69) and calculate the partition function for a hairpinned configuration of such a “un-designed” polymer. This could be done following the same line as that in Sec. III A and we find that

$$G_{\text{molten-bubble}}(z) = \frac{\frac{z}{(2\pi)^3} \int d\mathbf{q} G_{\text{even}}^0(z, \mathbf{q})}{1 + \frac{4\pi}{3} a^3 e^{\beta\epsilon_0} \frac{z}{(2\pi)^3} \int d\mathbf{q} \tilde{\mu}^2(\mathbf{q}) G_{\text{even}}^0(z, \mathbf{q})}. \quad (70)$$

In Eq. (70), $G_{\text{even}}^0(z, \mathbf{p}) = \sum_{n=0}^{\infty} z^n \tilde{Z}_0(2n, \mathbf{p})$, where $\tilde{Z}_0(2n, \mathbf{p})$ is the Fourier transform of the partition function of an un-designed polymer chain of $2n$ bonds (see Sec. III A). A careful analysis leads to the following expression

$$G_{\text{even}}^0(z, \mathbf{p}) = \frac{1}{\sqrt{z}} \frac{1}{1 - \tilde{\mu}(\mathbf{p})\omega(\sqrt{z})} \frac{\psi(\sqrt{z})}{1 + \tilde{\mu}(\mathbf{p})[2\psi(\sqrt{z}) - \omega(\sqrt{z})]}, \quad (71)$$

where $\omega(\sqrt{z})$ is calculated according to Eq. (61), and

$$\psi(\sqrt{z}) = \sqrt{z} \left[1 + \frac{Az}{(2\pi)^3} \int d\mathbf{q} \tilde{\mu}^2(\mathbf{q}) G_{\text{even}}^0(z, \mathbf{q}) \right]. \quad (72)$$

Similarly, the generating function for $Z_{\text{unpaired}}(n, \mathbf{r})$ can be obtained. Finally we arrive at the generating function for the whole system:

$$G(z, \mathbf{p}) = \sum_{n=1}^{\infty} z^n Z(n, \mathbf{p}) = \frac{G_{\text{even}}^0(z, \mathbf{p})}{1 - \frac{4\pi}{3} a^3 (e^{\beta\epsilon_1} - e^{\beta\epsilon_0}) \frac{z}{(2\pi)^3} \int d\mathbf{q} \tilde{\mu}^2(\mathbf{q}) G_{\text{even}}^0(z, \mathbf{q})} - 1. \quad (73)$$

When an external force \mathbf{F} is acting on the end of the polymer, the total generating function for the partition function has the same form as Eq. (72), but with \mathbf{p} replaced by $i\beta\mathbf{F}$. From Eq. (72) we see that $G(z, i\beta\mathbf{F})$ has three singular points, corresponding respectively to the native hairpin phase, the denatured molten phase and the denatured stretched coil phase. The phase diagram is shown in Fig. 17.

At given temperature and external force, the total relative extension of the polymer along the direction of the external force is obtained by the following expression:

$$E_x = -k_B T \frac{\partial \ln z}{\partial |\mathbf{F}|}. \quad (74)$$

In Fig. 19 the force-extension profiles of a model designed RNA chain at different fixed temperatures are calculated based on Eq. (74). For the parameters chosen in this figure, when environmental temperature is higher than 318.88 K and the external stretching is lower than 0.3 pN , the model polymer is in the molten-globule state, and the force-extension curve above 0.3 pN is continuous. When temperature is lower than 318.88 K the polymer is first

in the designed hairpin state, and unzipping of this hairpin by external force leads to a force plateau.

It is of interest to measure the elastic behavior of weakly designed RNA polymers and compare the experimental results with theoretical calculations mentioned above.

C. Glassy transitions in RNA sequences at low temperature

At low temperatures, a RNA polymer prefers to fold into a configuration that has the lowest possible energy. As discussed in the preceding subsection, if the RNA polymer under study has a designed native structure, it will fold to this structure at low temperatures. However, what will happen to a random-sequence RNA? Will many configurations of comparably low energies appear? Will the RNA be trapped in one of these low-energy configurations? These are very interesting questions that have attracted considerable theoretical thinking [127, 129–134]. Here we will only briefly discuss this issue, following the work of Bundschuh and Hwa [91, 131, 132].

In the analytical calculations of Refs. [91, 131, 132], each RNA monomer is either unpaired or is forming a base-pair with another monomer. A given RNA secondary configuration S is a set of base-pairs, any two of which are either mutually independent or are nested. The energy of a configuration S is assumed to be

$$E[S] = \sum_{(i,j) \in S} \epsilon_{ij}, \quad (75)$$

where ϵ_{ij} is the base-pairing energy between base i and base j . In the analytical work of Bundschuh and Hwa [131] the pairing energies ϵ_{ij} is assumed to be an identically and independently distributed Gaussian random variable:

$$\rho(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\epsilon - \epsilon_0)^2}{2\sigma^2}\right], \quad (76)$$

where ϵ_0 is the average pairing free energy and σ is the standard variation in the pairing free energy. The partition function of a RNA segment from base i to base j is denoted as $Z_{i,j}$:

$$Z_{i,j} = \sum_S \exp(-\beta E[S]). \quad (77)$$

This partition function satisfies the following recursive relationship:

$$Z_{i,j} = Z_{i,j-1} + e^{-\beta\epsilon_{ij}} Z_{i+1,j-1} + \sum_{k=i+1}^{j-2} Z_{i,k-1} e^{-\beta\epsilon_{kj}} Z_{k+1,j-1} + Z_{i,j-2} e^{-\beta\epsilon_{j-1,j}}, \quad (78)$$

with $Z_{k,k} \equiv 1$.

Homogeneous sequence First, let us consider the case of homogeneous RNA sequences [$\sigma = 0$ in Eq. (76)]. In this case the partition function $Z_{i,j}$ depends only on chain length $n = j - i$, so we denote it as $Z(n)$. Its generation function is

$$G(z) = \sum_{n=1}^{+\infty} Z(n)z^n. \quad (79)$$

Combining Eq. (79) and Eq. (78) leads to the following expression [91]

$$G(z) = \frac{1}{2qz^2} \left[1 - z - 2qz^2 - \sqrt{(1-z)^2 - 4qz^2} \right], \quad (80)$$

where $q = e^{-\beta\epsilon_0}$.

The scaling form of the partition function ($n \gg 1$) could then be obtained:

$$\begin{aligned} Z(n) &= \frac{1}{2\pi i} \oint \frac{G(z)}{z^{n+1}} dz \\ &= -\frac{1}{4\pi qi} \oint \frac{\sqrt{[1 - (1 + 2\sqrt{q})z][1 + (2\sqrt{q} - 1)z]}}{z^{n+3}} dz \\ &= -\frac{(1 + 2\sqrt{q})^{n+2}}{4\pi qi} \oint \frac{\sqrt{[1 - \zeta][1 + \kappa\zeta]}}{\zeta^{n+3}} d\zeta \\ &= \text{const} \times \frac{(1 + 2\sqrt{q})^n}{n^{3/2}}, \end{aligned} \quad (81)$$

where $\kappa = \frac{2\sqrt{q}-1}{2\sqrt{q}+1}$. The scaling exponent $\theta = \frac{3}{2}$ is explained in Ref. [131] using a “mountain” picture. This scaling exponent characterizes the RNA molten phase.

Heterogeneous chain For Heterogeneous RNA chain where the pairing free energy has the distribution Eq. (76), the quantity of interest is the average free energy over different realizations of the random sequence, i.e.,

$$[\mathcal{F}]_{\text{av}} = -k_B T [\ln Z]_{\text{av}}. \quad (82)$$

However, it is extremely difficult to calculate the average of the logarithm of the partition function. A trick that is routinely used in studying spin-glass systems is to rewrite this average into another form:

$$[\ln Z]_{\text{av}} = \lim_{n \rightarrow 0} \frac{[Z^n]_{\text{av}} - 1}{n}, \quad (83)$$

where Z^n means the total partition function of n RNA chains with the same sequence [135].

For our present system,

$$\begin{aligned} Z^n &= \sum_{S_1} \dots \sum_{S_n} \exp \left[-\beta \left(\sum_{(i_1, j_1) \in S_1} \epsilon_{i_1 j_1} + \dots + \sum_{(i_n, j_n) \in S_n} \epsilon_{i_n j_n} \right) \right] \\ &= \sum_{S_1} \dots \sum_{S_n} \exp \left[-\beta \sum_{(i, j)} n_{ij} \epsilon_{ij} \right], \end{aligned} \quad (84)$$

where n_{ij} is the total number of (i, j) base-pairs in the n replicas of the same system. Averaging over the random variables ϵ_{ij} in Eq. (84) leads to

$$[Z^n]_{\text{av}} = \sum_{S_1} \dots \sum_{S_n} \exp \left[-\beta \sum_{(i, j)} n_{ij} \epsilon_0 \right] \exp \left[\frac{\beta^2 \sigma^2}{2} \sum_{(i, j)} n_{ij}^2 \right]. \quad (85)$$

Another way to express Eq. (85) is as follows. Define $\theta_{ij}(k) = 1$ if in replica index k the two bases i and j forms a base-pair (i, j) , and $\theta_{ij}(k) = 0$ otherwise. Then $n_{ij} = \sum_k \theta_{ij}(k)$. One then has

$$[Z^n] = \sum_{S_1} \dots \sum_{S_n} \prod_{k=1}^n u^{|S_k|} \prod_{1 \leq k < l \leq n} \tilde{u}^{|S_k \cap S_l|}, \quad (86)$$

where

$$u = \exp[-\beta \epsilon_0 + \frac{1}{2} \beta^2 \sigma^2], \quad (87)$$

$$\tilde{u} = \exp[\beta^2 \sigma^2]. \quad (88)$$

Equation (86) is equivalent to Eq. (85), because $|S_k \cap S_l| = \sum_{(i, j)} \theta_{ij}(k) \theta_{ij}(l)$.

For general integer n , Eq. (86) is still very difficult to calculate. To make some progress, Bundschuh and Hwa focused on the case of $n = 2$ [131]. They found that Eq. (86) for this two-replica system is exactly solvable. When $N \gg 1$, the partition function has the following scaling form [131]

$$[Z^2] \sim N^{-\theta'} \zeta^N, \quad (89)$$

where θ' is a scaling exponent and ζ is related to the average free energy density.

Bundschuh and Hwa [131] demonstrated that at given level of sequence randomness, when the temperature is decreased below some threshold, there is a phase transition in the two-replica system. The main conclusions are as follows.

1. There exists a critical temperature T_c determined by the equation $\tilde{u} = \tilde{u}_c$, where \tilde{u}_c could be calculated by calculating the partition function of a double-stranded RNA system made of two independent RNA chains. When the sequence randomness is weak, i.e., $\sigma \simeq 0$, this critical temperature T_c also approaches zero.

2. When the temperature $T > T_c$, the system is in the molten phase, with $\theta' = 3 = 2 \times \frac{3}{2}$. This indicates that the two replicas are un-coupled in the thermodynamic limit. Therefore the sequence randomness is irrelevant at this temperature range.
3. When $T < T_c$, the system is in a *weak* glassy phase, with $\theta' = \frac{3}{2}$ in Eq. (89). The two replicas are locked together and the two-replica system behaves as a single RNA chain. This glassy phase is weak, because the exciting energy scales *logarithmically* with the domain size.

The work of Bundschuh and Hwa [131, 132] therefore proves the existence of a glassy phase in random-sequence RNA chains at low-enough temperature. Related issues are also been discussed in the literature [127, 129, 130, 133, 134].

Complementary to these analytical investigations, in Ref. [125] the temperature-induced folding-unfolding kinetics of a hairpin-forming RNA chain was followed. Zhang and Chen [125] found that when the environment temperature T is higher than a glass-transition T_g , the folding-unfolding process is a two-state process, between a pre-equilibrated quasi-equilibrium macro-state and the unfolded coil macro-state. While when $T < T_g$, the transition is not two-state-like but has many transition intermediates. This is consistent with the qualitative results of Bundschuh and Hwa [131]. At low temperature, the RNA chain is trapped in one of many free energy local minima states, and there are many comparable time scales corresponding to transitions between these intermediate states.

Hyeon and Thirumalai [136] demonstrated that the typical energy scale ϵ characteristic of the roughness of the energy landscape could be measured by performing force-extension experiments on RNA over different ambient temperatures.

D. Single-stranded DNA/RNA at low salt solution

In our previous discussion of RNA or single-stranded (ss) DNA elastic properties, the intramolecular and polymer-solution electrostatic interactions are completely ignored. However, as native single-stranded (ss) DNA/RNA molecules can achieve their biological activities only at their physiological salt milieu, it is of practical importance to understand the elasticity of ssDNA or RNA at different salt conditions. This subsection focuses on the influence of electrostatic interactions to ssDNA/RNA elasticity.

The electrostatic interactions of biopolymers in salt solution are usually theoretically approached by Debye-Huckel method [137–140]. On the experimental side, during recent years, there are several experimental groups who have pulled ssDNA or RNA molecules and reported the force-extension data in a variety of salt environments [2, 18, 34, 36, 47, 141]. These data are critical for checking these theoretical approaches to biopolymers in solution and meanwhile determine some important parameters of the molecules.

1. Debye-Huckel theory of biopolymers

In principle, the electrostatic potential $\psi(\mathbf{r})$ of an electrolyte solution at position \mathbf{r} is governed by the Poisson-Boltzmann equation [142]

$$\nabla^2\psi(\mathbf{r}) = -\frac{4\pi}{D} \sum_{i=1}^n v_i e c_i \exp(-v_i e \psi(\mathbf{r})/k_B T). \quad (90)$$

Here the solution is assumed to contain n different types of ions. The i th species has valence v_i , and the total number of ions of the i th type is N_i , with c_i ($= N_i/V$, V is the volume) being the bulk concentration of this species. D is the dielectric constant of the solution, and e denotes the protonic charge.

Equation (90) can be solved numerically under proper boundary conditions according to the series expansion method that was first used by Pierce [143]. In Figure 20 are illustrative examples of the electrostatic potentials of an ssDNA/RNA cylinder in 2 mM NaCl and 5 mM MgCl₂ solutions. Here a cylinder radius of $r_0 = 0.5$ nm and surface charge density of $e/0.6$ nm are assumed; and potential function $\psi(\mathbf{r})$ is expanded up to the 17th and 14th order for NaCl and MgCl₂, respectively [140].

In case the electrostatic potential is small relative to $k_B T$, the terms higher than square in Pierce-expansion can be neglected. The electrostatic potential around a point charge q can be written implicitly in the Debye-Huckel form

$$\psi_{\text{DH}}(r) = \frac{q}{D|r|} \exp(-\kappa|r|), \quad (91)$$

where r is the radial distance from the charge, and κ is the inverse Debye length, equaling to $(8\pi c_0 e^2 / D k_B T)^{1/2}$ (for NaCl solution) or $(24\pi c_0 e^2 / D k_B T)^{1/2}$ (for MgCl₂).

To count the influence of higher expansion terms of the Poisson-Boltzmann equation, one can phenomenologically change the amplitude of the Debye-Huckel potential of Eq. (91) to

fit the precise solution of the Poisson-Boltzmann equation [144, 145]. According to Eq. (91), the electrostatic potential of a uniformly charged straight cylinder of infinite length can be written as

$$\psi_{\text{DH}}(r) = \int_{-\infty}^{\infty} \frac{\nu d\lambda}{D} \frac{\exp(-\kappa\sqrt{\lambda^2 + r^2})}{\sqrt{\lambda^2 + r^2}} = \frac{2\nu}{D} K_0(\kappa r), \quad (92)$$

where λ is arclength along the cylinder axis, ν is effective linear charge density, and K_0 is the first-order modified Bessel function [146]. By comparing Eq. (92) with the numerical Poisson-Boltzmann solution in the tail region as shown in Fig. 20, we can determine the effective linear charge density ν in different bulk ionic concentration c for both the NaCl and the MgCl₂ solution (see Table I). In Table I we also show the effective charge density of double-stranded (ds) DNA, as well as ssDNA. All the data of ν can be very well fitted by a formula of

$$\nu = \exp(\alpha + \beta c^{2/5}), \quad (93)$$

with the fitting parameters α and β listed also in Table I.

2. Elastic freely-jointed chain model and interplay with electrostatic force

Because of high degree of flexibility, ssDNA/RNA molecules are usually modeled by a completely flexible “freely-jointed chain” (FJC) model: A chain consists of N rigid and independently oriented links, each of length b [147]. Since the contour length of the molecules can be elongated under high extended force, the modification of elastic links has been introduced into the FJC model [18]. Under the external force F , the free energy of this so-called elastic freely jointed chain model (eFJC) in salt solution can be written as

$$F_{\text{eFJC}} = \frac{Y}{2} \sum_{i=1}^N (|\mathbf{r}_i - \mathbf{r}_{i-1}| - b)^2 + \frac{\nu^2}{D} \int d\lambda_1 d\lambda_2 \frac{\exp(-\kappa|\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} - F z_N, \quad (94)$$

where $\mathbf{r}_i = (x_i, y_i, z_i)$ is the position of link i , and Young’s modulus Y characterizes the stretch stiffness of backbone. The integration is performed over the molecule’s curvilinear coordinates (with $i < j$).

In Figure 21 is a comparison of force-extension data with Monte Carlo calculations of the eFJC model [140, 141]. In the experimental setup of Dessinges and co-authors [141], ssDNA molecules were treated physicochemically to suppress the base-pairing interaction. In this condition, ssDNA extension was observed to increase almost logarithmically with force,

which deviates strongly from the behavior of a pure elastic FJC (Fig. 21). According to the Monte Carlo simulation, the electrostatic interaction tends to swell the volume occupied by the chain and make the segments more easily aligned along the force direction. This is equivalent to enlarging the Kuhn length of the molecule. In principle, the lower the ionic concentration becomes, the larger the effective Kuhn length is and the more rigid the molecule looks. As shown in Fig. 21, well agreement with experimental data up to 70 pN can be obtained by the model of Eq. 94 without adjustable parameters. Beyond the external force of 70 pN, however, the extremely large deformation of the ssDNA backbone could not be described by the commonly-used quadratic elastic approximation [88, 93, 140] but requires the introduction of a quartic elastic term with elastic module $Y_4 = 400 k_B T / nm^4$ in the model [141]. This nonlinear elastic behavior may result from a structural transition in ssDNA as it is pulled above 70 pN [18].

The logarithmic increasing of extension with external stretching in ssDNA and RNA revealed in Fig. 21 is very striking. Recently, it was observed that similar force-extension behavior could happen in the stretching of spider-silk proteins [148]. It is yet to see whether this phenomenon is also caused by electrostatic interactions or not.

3. Base-pairing interactions of ssDNA/RNA in solution

Long-range base-pairs can be formed between the complementary bases of ssDNA/RNA, which tend to bend the molecule into specific secondary structures. Since hydrogen-bonds of the base-pairs themselves are short-ranged ($\sim 0.3-0.4$ nm of donor-acceptor distance [149]), this secondary structures are strongly influenced by the excluded volume interactions that can be dramatically enhanced by the electrostatic repulsion of phospho-diester backbone.

To include the self-pairing interaction in the eFJC model, we approximate the base-pairing potential by a node-pairing energy, i.e.

$$E_{\text{pair}} = \sum_{i=1}^{N_P} V_P, \quad (95)$$

where N_P is the number of node pairs and V_P is a sequence-dependent parameter denoting the intensity of a certain pairing mode.

As shown in Fig. 22, in low-salt solution, the electrostatic repulsive interaction along ssDNA backbone is weakly screened. For example, in 1mM phosphate buffer (mainly

NaH₂PO₄), the Debye length $l_D = 5.95$ nm which is much larger than the hydrogen-bond interaction range. In this case, there are few base pairs being able to form because the dominant electrostatic repulsive potential excludes the bases from getting close into the Watson-Crick base-pair range. In high-salt solution, the electrostatic self-avoiding interaction is much weaker because of the strong screen effect. The complementary bases can therefore come spatially close and form base-pairs. In this case, it need slightly larger force to elongate the molecule than a pure FJC chain to overcome additional base-pairing interaction in low force range.

4. Sequence sensitivity of ssDNA/RNA elasticity

There are three hydrogen-bonds formed in a G-C pair-pair and two hydrogen-bonds in a A-T base-pair. This difference of hydrogen-bonding strength leads to sequence sensitivity of ssDNA/RNA secondary structure and elasticity in high salt solutions.

In Fig. 23, two different sequences of charomid and pXΔII, which have 50% and 30% GC contents, respectively, are pulled under external forces [141]. The MC simulation of Eq. 94 plus pairing interaction Eq. 95 fits the data well with $V_P = 4.6 k_B T$ for the charomid and $V_P = 3.8 k_B T$ for the GC-poorer pXΔII construction. As expected, the higher the GC content, the higher the average pairing energy V_P , the more stable are the hairpins and the higher the force required to stretch the ssDNA to a given length. This sensitivity to ionic conditions and nucleotide content disappears at force large enough to unzip DNA (>10 pN). As hairpins are less likely, the elasticity of ssDNA is less sensitive to variations of sequence and buffers. Furthermore, recent analytical results [88, 93] are in good agreement with experimental observations and Monte Carlo simulations.

When the pair-bases in RNA hairpin structures are neighboring along the sequence, there exists van der Waals interaction between consecutive base-pairs. This stacking interaction most likely happens in homogeneous sequences, such as poly(G-C) and poly(A-T).

To include the stacking interaction in the eFJC model, we count the stacking energy as

$$E_{\text{stacking}} = \sum_{i=1}^{N_S} V_S, \quad (96)$$

where N_S is the number of stacked node pairs and V_S is the interactive potential between two neighboring node pairs. Two node pairs are considered as stacked only when they are

nearest neighbors to each other.

As shown in Fig. 24, the stacking potential can dramatically change the conformation of ssDNA at 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 (left column of Fig. 24). When the stacking interaction exists, on the other hand, the base pairs are encouraged energetically to be neighboring and therefore lead to the formation of bulk hairpin structure (right column of Fig. 24). This kind of abrupt transition from hairpin to coil was indeed observed in recent experiments by Rief and co-authors [47], in which two designed poly(A-T) and poly(G-C) ssDNA sequences are pulled (see Fig. 25). The value of threshold force, i.e., the height of the plateau in the force-extension curve, is dependent on the pairing and stacking potentials of the nodes. The best fit of MC calculations [140] to the data are $V_P = 4.1 k_B T$ and $V_S = 4 k_B T$ in poly(A-T) sequence, $V_P = 10.4 k_B T$ and $V_S = 6 k_B T$ in poly(G-C) sequence. Bearing in mind that each Kuhn length (~ 1.6 nm) in our simulation contains about three nucleotide bases, we can infer that the pairing energy of each A-T base pair is $1.37 k_B T$ and that for each G-C base-pair is $3.47 k_B T$. These values are comparable with the measurements of Bockelmann et al. [38] by unzipping the two strands of a dsDNA helix.

E. RNA secondary structure prediction through mechanical stretching

In preceding several subsections we have discussed the general physics of RNA secondary structures. Recently, the biological significance of RNA molecules have been acknowledged by more and more researchers. Besides the conventional functions of RNA, the mechanisms of small-RNA and RNA binding and RNA interference were discovered. Therefore, an issue of great biological interest is to predict faithfully the structure of a given RNA sequence. In this subsection we mention some recent investigations on this respect.

RNA structure prediction is certainly not a new question. Many algorithms have already been existed for quite a long time, see, for example, Refs. [150–155] and references cited therein. These algorithms work mainly for the secondary structures of RNA. They try to identify the structures with the minimum free energy by calculating the partition function iteratively [see Eq. (78)] and by assigning different empirical statistical weights to

different structural elements [150–152]. For short RNA sequences which contain hundreds of nucleotides, these algorithms usually predict a pairing pattern for a given RNA that is in reasonable agreement with experimental measurements.

However, these iterative algorithms are unreliable for very long RNA chains. Part of the reason is that these algorithms excluded the possibility of tertiary structures such as *pseudo-knots*. A pseudo-knot is formed when the nucleotide bases in the loop regions of two hairpin segments pairing with each other (see Fig. 26).

With the development of single-molecule manipulation techniques, the suggestion has been raised to detect the secondary structure of RNA by mechanical stretching. For long RNA sequences, however, experimental measurements [2, 34, 141] revealed that the force-extension curves of real RNA chains are very smooth and featureless. This situation is dramatically different from the unzipping curves of double-stranded DNA molecules [37].

To find out why the force-extension profiles of stretching RNA can not reveal its structure, Gerland and co-workers [156] performed numerical stretching simulations on several real RNA molecules with well-defined native structures. They also found that the equilibrium force-extension curves for natural or random RNA chains are considerably smooth and do not contain many features to enable the detection of the underlying secondary structure of RNA (Figure 27a).

Two reasons may contributed to this observed smoothness. First is a compensation effect [156]. As demonstrated in Fig. 28a, as the total extension of a RNA polymer is increased, new structural motifs may be opened, causing a transient decrease in the stretching force; however, some structural motifs that were opened previously may reform rapidly, leading to a reverse increase of the force. When pulling velocity is not so quick, the resulting force-extension will then be smooth. Lower the temperature or increasing the pulling rate may help improving the detection of RNA secondary structure.

The second reason is the contribution of suboptimal structures [156]. According to the calculation in § III A, fluctuations of RNA structures over different suboptimal configurations could result in a very smooth stretching curve.

To delete the compensation effect and to probe the native secondary structure of a RNA molecule by mechanical stretching, Gerland and co-workers [157] suggested an elegant way. The idea is to let the RNA molecule to pass through a nano-pore (see Fig. 28b). The radius of the nano-pore is so small that only single-stranded RNA segments can pass through

and double-stranded segments could not. By pulling the single-stranded part, the double-stranded hairpin segment nearest to the nano-pore will be unzipped while other double-stranded segments will not be destroyed. Therefore, the hairpin secondary structures in a RNA polymer will be pulled out in a discrete manner, starting from one end of the polymer. Gerland and co-workers performed simulations with this idea and their simulation results suggested that by this means, the native secondary structures of RNA can indeed be largely probed. Figure 29 shows the simulated force-extension curves of pulling a real RNA polymer through a nano-pore. We see that saw-tooth patterns which signify the breaking of individual RNA structural elements appear with considerable reproducibility. For more quantitative discussion of this issue, please refer to Ref. [157].

IV. DNA ENTROPIC ELASTICITY

We have discussed in the preceding two sections structural transitions in DNA and RNA that are associated with breaking of nucleotide base-pairs. In this and the next two sections, we focus on the mechanical property of DNA double-helix when the base-pairing pattern is kept intact.

Double-stranded DNA (dsDNA) is a long polymer of nucleotides. In solution its spatial configuration fluctuates over time because of thermal activations which are characterized by energy scale $k_B T$. What is the best model to describe the elastic behavior of solution dsDNA?

Here we first outline the main predictions of three polymer models: the Gaussian model, the freely-jointed-chain (FJC) model, and the wormlike-chain (WLC) model. The WLC model is found to be most suitable in describing the elastic response of dsDNA under the action of low external forces (less than 10 piconewton).

A. The Gaussian model and the freely-jointed-chain model

In the Gaussian model, a polymer is assumed to be formed by tandemly connecting N bonds. The end-to-end extension is

$$\mathbf{R} = \sum_{i=1}^N \mathbf{r}_i. \quad (97)$$

Each bond \mathbf{r}_i has the following bond length distribution

$$P(\mathbf{r}_i) = \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{r}_i^2}{2b^2}\right). \quad (98)$$

The characteristic length scale b is called the Kuhn length [147], $\langle \mathbf{r}_i^2 \rangle = b^2$. Between any two bonds there is no other interactions.

Under the action of an external stretching force of magnitude f , the average extension of the polymer along the force direction is

$$X = \frac{Nb^2}{3k_B T} f. \quad (99)$$

The average extension is proportional to force f .

The Gaussian model works well when the external force is small. However, under moderate or high stretching conditions, the assumption Eq. (98) becomes invalid, since each bond

could have only finite extensibility. The freely-jointed-chain model replace this assumption with that of a rigid bond,

$$P(\mathbf{r}_i) = (4\pi b^2)^{-1} \delta(|\mathbf{r}_i| - b). \quad (100)$$

However, the orientation of nearest-neighboring bonds are completely independent. Under this assumption, the force-induced total average extension is

$$X = Nb \left[\frac{\cosh(\beta fb)}{\sinh(\beta fb)} - \frac{1}{\beta fb} \right] = L\mathcal{L}(\beta fb), \quad (101)$$

where $L = Nb$ is the total contour length of the polymer and $\mathcal{L}(x) = \coth(x) - 1/x$ is known as the Langevin function.

When the force is small, Eq. (101) reduces to the linear relationship Eq. (99). However, for large force, Eq. (101) predicts that

$$1 - \frac{X}{L} = \frac{k_B T}{fb}. \quad (102)$$

This scaling law characterizes the FJC model.

B. The wormlike-chain model

In 1992, Smith, Finzi, and Bustamante [1] made the first measurement on the elastic response curve of λ -phage DNA (48,502 base-pair in length) under the action of an external force. These authors found that, neither the Gaussian model Eq. (99) nor the freely-jointed-chain model Eq. (101) could explain their experimental data [1]. Later it turned out that dsDNA is a semiflexible polymer, and to properly account the elasticity of dsDNA one has to use the wormlike-chain model [16, 158].

In the WLC model, a polymer is regarded as an *inextensible* chain of contour length L , and its configuration is described by the position vectors $\mathbf{r}(s)$ at each arclength point $0 \leq s \leq L$ along the chain. The chain is *semiflexible*, so there is an energy penalty associated with bending of the chain. The configurational energy is expressed as follows:

$$\mathcal{H}[\mathbf{r}] = \int_0^L ds \frac{k_B T \ell_p}{2} \left(\frac{d\mathbf{r}(s)}{ds} \right)^2 = \int_0^L ds \frac{k_B T \ell_p}{2} \left(\frac{d\mathbf{t}(s)}{ds} \right)^2. \quad (103)$$

In Eq. (103), $\mathbf{t}(s) = \frac{d\mathbf{r}(s)}{ds}$ is the unit tangent vector at arclength s . The physical meaning of the *bending persistence length* ℓ_p will be explained shortly.

The probability distribution of the tangent vector \mathbf{t} at arclength s is determined by the following Green equation:

$$\frac{\partial P(\mathbf{t}, s)}{\partial s} = \frac{1}{2l_p} \nabla_{\mathbf{t}}^2 P(\mathbf{t}, s). \quad (104)$$

The boundary condition of Eq. (104) is $P(\mathbf{t}, 0) = \delta(\mathbf{t} - \mathbf{t}_0)$, where \mathbf{t}_0 is the tangent of the chain at arclength $s = 0$. The derivation of Eq. (104) is given in the Appendix.

Solution of Eq. (104) leads to

$$P(\mathbf{t}, s) = \sum_{l=0}^{+\infty} \exp\left[-\frac{l(l+1)}{2l_p} s\right] \frac{2l+1}{4\pi} P_l(\mathbf{t} \cdot \mathbf{t}_0), \quad (105)$$

where $P_l(\mathbf{t} \cdot \mathbf{t}_0)$ is l -th order Legendre function. Therefore, the tangent-tangent correlation is

$$\langle \mathbf{t}(s) \cdot \mathbf{t}_0 \rangle = \exp\left(-\frac{s}{l_p}\right). \quad (106)$$

From Eq. (106) we know that l_p is the characteristic correlation length of the polymer's direction. For dsDNA, $l_p \simeq 53$ nm or 150 base-pairs [16, 158].

Under the action of an external stretching force F , the total average end-to-end extension is determined by finding the ground-state eigenvalue and eigenvector of the following Green function (see Appendix):

$$\frac{\partial \Psi(\mathbf{t}, s)}{\partial s} = \left[\frac{1}{2l_p} \nabla_{\mathbf{t}}^2 + \beta F \mathbf{z}_0 \cdot \mathbf{t} \right] \Psi(\mathbf{t}, s), \quad (107)$$

where \mathbf{z}_0 denotes the direction of the external force. The ground-state eigenvalue and eigenvector of Eq. (107) could be obtained exactly (see, for example, Ref. [159]), and the resulting force-extension relation is shown in Fig. 30. Here, we also give an approximate variational solution, which turns out to be very accurate.

We choose the following variational ground-state eigenvector

$$\Psi_0 = \exp[-g_0 s + (\eta/2) \mathbf{t} \cdot \mathbf{t}_0], \quad (108)$$

and the ground-state “energy” is

$$g_0 = \frac{\int \Psi_0^* [-(2l_p)^{-1} \nabla_{\mathbf{t}}^2 - \beta F \mathbf{t} \cdot \mathbf{z}_0] \Psi_0 d\mathbf{t}}{\int d\mathbf{t} \Psi_0^* \Psi_0} = \left(\frac{\eta}{4l_p} - \beta F \right) \mathcal{L}(\eta). \quad (109)$$

The variational parameter η should be chosen such that g_0 attains its minimal value. Therefore, η is determined by

$$4\beta F l_p = \eta + \left(\frac{d \ln \mathcal{L}(\eta)}{d\eta} \right)^{-1}, \quad (110)$$

and the average end-to-end extension along the force direction \mathbf{z}_0 is

$$X = -Lk_B T \frac{\partial g_0}{\partial F} = L\mathcal{L}(\eta). \quad (111)$$

For many applications, the following famous interpolation formula [16, 158] for the force-extension relation is often used:

$$F = \frac{k_B T}{4\ell_p} \left[\frac{1}{\left(1 - \frac{X}{L}\right)^2} - 1 + \frac{4X}{L} \right]. \quad (112)$$

To understand Eq. (112), first we notice that when the external force is very weak, the average extension X is related to the force through the linear-response theorem

$$X = \frac{F}{k_B T} \left[\langle ([\mathbf{r}(L) - \mathbf{r}(0)] \cdot \mathbf{z}_0)^2 \rangle_{F=0} - \langle [\mathbf{r}(L) - \mathbf{r}(0)] \cdot \mathbf{z}_0 \rangle_{F=0}^2 \right] = \frac{2\ell_p L}{3k_B T} F. \quad (113)$$

On the other hand, when the external force is very large, the configuration of the worm-like chain will be aligned along the direction of the external force, the tangent vector can only perform very small fluctuations around the direction of the external force. We can expand the tangent vector along the force direction to quadratic term and obtain the following equation

$$1 - \frac{X}{L} = \sqrt{\frac{k_B T}{4\ell_p F}}. \quad (114)$$

Combining Eqs. (113) and (114) then leads to the interpolation formula Eq. (112).

Comparing Eq. (114) with Eq. (102), we know that as the external force F is large, the difference between a semiflexible WLC chain's contour length and its extension scales as $F^{-1/2}$ rather than F^{-1} . This $-\frac{1}{2}$ scaling law was observed in dsDNA [1, 16].

The results mentioned in this section are summarized in Fig. 30.

The most interesting behavior of a semiflexible polymer is when the total contour length L is of the same order as the bending persistence length ℓ_p . On such kind of systems there has recently been many work (see, for example, Refs. [160–163]).

V. DNA OVER-STRETCHING TRANSITION

The worm-like chain model of the preceding section regards DNA as an inextensible chain. When DNA is under the action of a relative large external force ≥ 10 pN, the molecule becomes extensible. In the force range $10 \text{ pN} \leq F \leq 70 \text{ pN}$, the contour length of DNA linearly increases with external stretching, with a stretching modulus $\simeq 10^3$ pN [18]. Surprisingly, at external force $F \simeq 70$ pN, the contour length of DNA suddenly increases to 1.7 times its B-form length [18, 19, 30, 164]. This abrupt structural transition is referred to as DNA over-stretching transition in the literature. When $F < 70$ pN, DNA is a double-helical molecule, with folded sugar-phosphate backbones; when $F > 70$ pN, DNA is over-stretched, suggesting that the hidden length in the double-helix is pulled out. This section reviews recent efforts in understanding the nature of this DNA over-stretching transition.

In the literature there are two kinds of opinions concerning DNA over-stretching transition. In one opinion, the transition is between B-form DNA and an over-stretched S-form DNA. S-form DNA is also a double-helical molecule. In the other opinion, which was first suggested by Smith, Cui, and Bustamante [18] and later more thoroughly investigated by Bloomfield and co-workers [164–166], is that the over-stretching transition is actually a force-induced DNA melting process.

A. Co-operative over-stretching transition viewed from a discrete model

Many theoretical and computational efforts have been made to understand the origin of DNA over-stretching transition (see, for example, Refs. [19, 28, 167–173]). The first theoretical understanding of DNA over-stretching was undertaken by Cluzel and co-workers [19] based on a discrete Zimm-Bragg model (and similar models were also considered in Refs. [168–170, 173]). Here we show the main points of this model.

Under the action of an external force f , each structural unit of DNA (i.e., a nucleotide pair) has two states: a B-form state with extension $l_B(f)$ and a over-stretched S-form state with extension $l_S(f)$. There is an energy difference ΔE between these two states. Furthermore, there is an addition energy cost ω with each junction between an S-form segment and a B-form segment. A DNA chain of length N base-pairs may have a configuration characterized by the following set of integers $\{m_0, n_1, m_1, \dots, n_s, m_s, n_0\}$, where $m_0 \geq 0$ and

$m_i \geq 1$ ($i = 1, \dots, s$) are, respectively, the number of nucleotide pairs in the first, second, ..., $s + 1$ -th B-form segment; and $n_i \geq 1$ ($i = 1, \dots, s$) and $n_0 \geq 0$, are the corresponding S-form segments between any two consecutive B-form segments. The configuration energy is

$$\mathcal{H}[\{m_i, n_i\}] = - \sum_{i=0}^s m_i f l_B - \sum_{j=0}^s n_j (f l_S - \Delta E) + (2s + 1)\omega - \omega \delta_{m_0}^0 - \omega \delta_{n_0}^0. \quad (115)$$

An application of the generating function method of § II C leads to the following expression for the average extension X of a nucleotide pair:

$$X = \frac{[\alpha_1 l_S + \alpha_2 l_B] \sqrt{(\alpha_1 - \alpha_2)^2 + 4\alpha_3} + (\alpha_1 - \alpha_2)[\alpha_2 l_B - \alpha_1 l_S] - 2\alpha_3[l_B + l_S]}{(\alpha_1 + \alpha_2) \sqrt{(\alpha_1 - \alpha_2)^2 + 4\alpha_3} - (\alpha_1 - \alpha_2)^2 - 4\alpha_3}, \quad (116)$$

where $\alpha_1 = \exp(\beta f l_B)$, $\alpha_2 = \exp(\beta f l_S - \beta \Delta E)$, and $\alpha_3 = \alpha_1 \alpha_2 \exp(-2\beta \omega)$.

When external force is larger than 10 pN, the B-form DNA could be well modeled by an extensible worm-like chain with bending persistence length $l_p = 53.0$ nm and stretching modulus $S_B = 1000$ pN [2]. Therefore,

$$l_B(f) = d_B \mathcal{L}(\eta_B)(1 + f/S_B), \quad (117)$$

where $d_B = 3.4 \text{ \AA}$ is the distance between two nucleotide pairs in B-form DNA, and η_B is determined by Eq. (110) with the bending persistence length of B-form DNA. According to the work of Storm and Nelson [173], S-form DNA could also be regarded as a wormlike chain with bending persistence length $l'_p = 12.0$ nm. So we could write down

$$l_S(f) = d_S \mathcal{L}(\eta_S)(1 + f/S_S), \quad (118)$$

where d_S is set to $d_S = 5.8 \text{ \AA}$ and η_S is also determined by Eq. (110) with the bending persistence length l'_p . The stretching modulus of S-DNA is set to $S_S = 2000$ pN.

Figure 31 demonstrated the predicted force-extension relationship for DNA according to this Zimm-Bragg model. We see that the threshold force 70 pN is determined by the energy difference ΔE between S-form DNA and B-form DNA, while the cooperativity of transition is determined by the junction energy cost ω . The higher the junction energy cost, the abrupt the B-S transition is. So this simple model is able to reproduce the experimental observations [18, 19].

What is the physical origin of the cooperation factor ω ? There are some possible alternative explanations. Cluzel et al. [19] performed molecular modeling of the DNA stretching

process using the program JUMNA and suggested that in the over-stretched DNA the nucleotide bases are no longer perpendicular to the molecular axis. Therefore, it is possible that the discontinuities in inclination of the nucleotide bases in the B-form and S-form boundary will cause large energy penalty. However, the molecular modeling of Cluzel et al. [19] was performed by assuming the total twist of DNA to be fixed unchanged, which was in conflict with experimental situations. In the experimental situation, there were many single-stranded breaks (nicks) in the DNA back-bone, which allow the twisting stress in DNA to be released by rotating the DNA molecule around such nicks. Another possibility is that B-S transition was initiated only at these nicks. The existence of just a limited number of single-strand nicks can lead to a phenomenological energy penalty ω . However, this explanation is also not very convincing since (1) the same force-extension behavior was observed in many types of DNA molecules which differ in nucleotide composition and number of single-strand nicks and (2), the over-stretching transition is not sensitive to pulling velocity.

A semi-microscopic model of DNA will be introduced in the next subsection. This model suggests that the short-ranged nature of base-pair stacking interactions might be the main reason for the highly cooperative B-S transition.

B. Double-stranded polymer model of DNA and the over-stretching transition

This subsection investigates theoretically the possible effects of DNA short-ranged base-pair stacking interaction to the co-cooperativity of its over-stretching transition. To incorporate base-pair stacking interactions, Zhou and co-authors [28, 159] introduced a double-stranded semiflexible polymer model for DNA and studied its property detailedly. Here we recall the main points of this model.

In the model of Zhou et al. [28], DNA is modeled as a double-stranded polymer with continuous degrees of freedom (Fig. 32). The two inextensible backbones of DNA [14] are characterized by the same bending rigidity $\kappa = k_B T \ell_p$, where $\ell_p \simeq 1.5$ nm is the bending persistence length of ssDNA. The position vector along ssDNA is $\mathbf{r}_i = \int^s \mathbf{t}_i(s') ds'$, where \mathbf{t}_i ($i = 1, 2$) is the unit tangential vector of the i th backbone, and s its arc length (Fig. 32). The nucleotide basepairs between the backbones are viewed as rigid planar structures with finite area and volume.

Bending energy of DNA First consider the bending energy of the backbones,

$$E_b = \int_0^L \frac{\kappa}{2} \left[\left(\frac{d\mathbf{t}_1}{ds} \right)^2 + \left(\frac{d\mathbf{t}_2}{ds} \right)^2 \right] ds, \quad (119)$$

where L is the contour length of each ssDNA backbone and s is its arc length parameter. To proceed, each base-pair is regarded for the moment as a thin rigid rod of length $2R$, with a unit vector \mathbf{b} pointing along it from \mathbf{r}_1 to \mathbf{r}_2 , i.e., $\mathbf{r}_2(s) - \mathbf{r}_1(s) = 2R\mathbf{b}(s)$. Relative sliding of the backbones is prohibited, the basepair planes are assumed to lie perpendicular to the DNA central axis and $\mathbf{b} \cdot \mathbf{t}_1 = \mathbf{b} \cdot \mathbf{t}_2 \equiv 0$. The central axis of dsDNA can be defined as $\mathbf{r}(s) = \mathbf{r}_1(s) + R\mathbf{b}(s)$, and its tangential vector is denoted by \mathbf{t} , with $\mathbf{t} \cdot \mathbf{b} = 0$. Since both \mathbf{t}_1 and \mathbf{t}_2 lie on the same plane perpendicular to \mathbf{b} , we obtain that $\mathbf{t}_1 = \mathbf{t} \cos \varphi + \mathbf{n} \sin \varphi$ and $\mathbf{t}_2 = \mathbf{t} \cos \varphi - \mathbf{n} \sin \varphi$, where $\mathbf{n} = \mathbf{b} \times \mathbf{t}$ and φ is half the rotational angle from \mathbf{t}_2 to \mathbf{t}_1 (\mathbf{b} being the rotational axis). We call φ the folding angle, it is in the range between $-\pi/2$ and $+\pi/2$ ($\varphi > 0$ for right-handed rotations and < 0 for left-handed ones). It is not difficult to verify that

$$\frac{d\mathbf{b}}{ds} = \frac{(\mathbf{t}_2 - \mathbf{t}_1)}{2R} = -\mathbf{n} \frac{\sin \varphi}{R}. \quad (120)$$

With Eq. (120) and the definition of \mathbf{r} we know that

$$\frac{d\mathbf{r}}{ds} = \frac{(\mathbf{t}_1 + \mathbf{t}_2)}{2} = \mathbf{t} \cos \varphi. \quad (121)$$

With the help of Eqs. (120) and (121), the total bending energy of the backbones Eq. (119) can be rewritten as

$$E_b = \int_0^L \left[\kappa \left(\frac{d\mathbf{t}}{ds} \right)^2 + \kappa \left(\frac{d\varphi}{ds} \right)^2 + \frac{\kappa}{R^2} \sin^4 \varphi \right] ds. \quad (122)$$

The second and the third terms in Eq.(122) is deformation energy caused by folding of the backbones with respect to the central axis, and the first term, $\kappa(dt/ds)^2$, is the bending energy of DNA central axis contributed by the backbone bending rigidity κ . So far, base-pairs are viewed as thin rods and their contribution to the bending rigidity of DNA chain is not considered. Because of steric effects caused by finite volume and area, base-pairs will certainly increase the bending rigidity of DNA chain (the bending persistence length of double-stranded DNA is about 50 nm, quite larger than that of a single-stranded DNA chain). The simplest way to consider such effects is to replace κ in the first term of Eq. (122) with a phenomenological parameter κ^* , with $\kappa^* > \kappa$ [28].

Base-pair stacking energy Besides steric effects, nucleotide base-pairs contribute also base-pair stacking energy. This energy mainly originates from non-covalent van der Waals interactions between adjacent base-pairs [14]. Base-pair stacking interaction is short-ranged and is characterized by an attraction potential proportional to $1/d^6$ and a strong repulsion potential proportional to $1/d^{12}$ (here d is the axial distance between adjacent base-pairs). In our continuous model, the line density of such Lennard-Jones type potential can be written as

$$\rho(\varphi) = \begin{cases} \frac{\epsilon}{r_0} \left[\left(\frac{\cos \varphi_0}{\cos \varphi} \right)^{12} - 2 \left(\frac{\cos \varphi_0}{\cos \varphi} \right)^6 \right] & \text{for } (\varphi \geq 0), \\ \frac{\epsilon}{r_0} [\cos^{12} \varphi_0 - 2 \cos^6 \varphi_0] & \text{for } (\varphi < 0). \end{cases} \quad (123)$$

In Eq. (123), r_0 is the backbone arc length between adjacent bases; φ_0 is a parameter related to the equilibrium distance between a DNA dimer; ϵ is the base-pair stacking intensity which is generally base-sequence specific. Here we focus on macroscopic properties of DNA, so we consider ϵ in the average sense by taking its averaged value over ten different DNA base-pair dimers (see Table II) [14, 174]. In this work, ϵ is set to $14.0 k_B T$, and $T = 300 K$.

The asymmetric base-pair stacking potential Eq. (123) ensures a relaxed DNA to take on a right-handed double-helix configuration with its folding angle $\varphi \sim \varphi_0$. However, if adjacent base-pairs are pulled apart slightly from the equilibrium distance by external forces or thermal stretching fluctuations, the base-pair stacking interaction intensity quickly decreases because of its short-range nature. In other words, the base-pair stacking potential can endure only a limited pulling force.

The total base-pair stacking energy is

$$E_{\text{LJ}} = \int_0^L \rho(s) ds. \quad (124)$$

External stretching energy When an external force is pulling at one end of DNA, the DNA will try to align along the force direction \mathbf{z}_0 . The energy associated with external force is

$$E_f = - \int_0^L f \mathbf{t} \cdot \mathbf{z}_0 \cos \varphi ds. \quad (125)$$

The total energy of a particular configuration is the sum of the bending energy Eq. (122), the base-pair stacking energy Eq. (124), and the alignment energy Eq. (125):

$$E = E_{\text{bend}} + E_{\text{LJ}} + E_f. \quad (126)$$

According to the derivation given in the Appendix, the Green function $G(\mathbf{t}, \varphi; \mathbf{t}', \varphi'; s)$ which determines the probability distribution of \mathbf{t} and φ along DNA chain is governed by

$$\frac{\partial G}{\partial s} = \left[\frac{\partial^2}{4\ell_p^* \partial \mathbf{t}^2} + \frac{\partial^2}{4\ell_p \partial \varphi^2} - \frac{f \mathbf{t} \cdot \mathbf{z}_0}{k_B T} \cos \varphi - \frac{\rho(\psi)}{k_B T} - \frac{\ell_p \sin^4 \varphi}{R^2} \right] G, \quad (127)$$

where $\ell_p^* = \kappa^*/k_B T$. The spectrum of the above Green equation is discrete and hence for long chains, the average extension can be obtained either by differentiation of the ground-state eigenvalue, g , of Eq. (127) with respect to f :

$$\frac{X}{L} = (1/L) \int_0^L \langle \mathbf{t} \cdot \mathbf{z}_0 \cos \varphi \rangle ds = k_B T \frac{\partial g}{\partial f}, \quad (128)$$

or by a direct integration with the normalized ground-state eigenfunction, $\Phi(\mathbf{t}, \varphi)$, of Eq. (127):

$$\frac{X}{L} = \int \int |\Phi|^2 \mathbf{t} \cdot \mathbf{z}_0 \cos \varphi d\mathbf{t} d\varphi. \quad (129)$$

Both g and $\Phi(\mathbf{t}, \varphi)$ can be obtained numerically through standard diagonalization methods and identical results are obtained by Eqs. (128) and (129). The theoretical force-extension of dsDNA is calculated numerically and compared with experimental data of Ref. [19]. Figure 33 shows the result of this comparison. The theoretical curve in Fig. 33 is obtained with just one adjustable parameter. The agreement with experiment is remarkable.

According to the continuous-degree model mentioned above, the onset of co-operative extension of DNA axial length at forces about 65-70 pN is mainly caused by the yielding of the short-range base-pair stacking interaction. After the collapse of base-pair stacking, the model assumes double-stranded DNA will take a planar ladder structure. In reality more complex configurational transition might happen after the yielding of stacking interaction. For example, the un-stacked base-pair may become unstable and break into two unpaired bases. Alternatively, the vertically stretched base-pairs may tilt with respect to the DNA central axis to gain some residual base-pair stacking. To take into account these and other kind of possibilities, one need to extend the basic model to include the possibility of base-pair breaking and tilting. Related investigations could be found in Refs. [175–180].

The continuous model of Zhou et al. [28] is also able to describe the elastic response of DNA at low external forces. Below the onset of co-operative elongation, DNA seems to be very stiff and calculations based on this model show that at $f = 50$ pN the total extension of DNA is only 4.1% longer than its B-form contour length, in close accordance with the value of 4.6% reported by Smith et al. [18]. This is related to the fact that the base-pair stacking

intensity ϵ is very strong. At low forces ($f < 10$ pN), because the fluctuation of the folding angle φ is extremely small, it can just be neglected and DNA elasticity is caused by thermal fluctuations of the axial direction \mathbf{t} (entropic elasticity). The entropic elasticity (wormlike chain) model of § IV B with contour length $L\langle\cos\varphi\rangle_{f=0}$ and persistence length $2\ell_p^*\langle\cos\varphi\rangle_{f=0}$ is an excellent approximation of the present theory (here $\langle\cos\varphi\rangle_{f=0}$ is the average of $\cos\varphi$ at zero force). This point is demonstrated in Fig. 34.

In § VI A we will show that the model mentioned in this subsection can also quantitatively explain the supercoiling property of stretched dsDNA without any fitting parameter.

VI. ELASTICITY OF SUPERCOILED DNA

The number of times the two strands of DNA double helix are interwound, i.e., the linking number Lk , is a topologically invariant quantity for closed DNA molecules [181]. It is also topologically invariant for a linear DNA polymer in case that the orientations of the two extremities of the linear polymer are fixed and no any part of the polymer is allowed to go round the extremities of the polymer [20]. For an unstressed B-form DNA segment, it has one right-handed twist about per 3.4 nm along its length [14], i.e., $Lk_0 = L_B/3.4$, where L_B is the DNA length. In real biological processes, however, DNA molecules are often overwound or underwound from its relaxed state in order to facilitate specific biological activities. For example, in human cells, the nucleus is just a few micrometers in diameter while it contains DNA with total length three meters within 23 chromosomes. If these chromosomes were in the relaxed coil form, it would be hard for them to fit inside the nucleus. It was observed via the reconstitution of nucleosome cores by urea/salt dialysis that a negatively twisted molecule can be packed more efficiently compared with linearized molecules [182]. The reciprocal relations between the twisting topology of chromatin and the genetic processes *in vivo* are also extensively acknowledged, and the twisting stress was shown to influence the efficiency of transcription [183] and to modulate the binding of trans-acting proteins to DNA elements [184, 185].

Conceptionally, in all cases when the linking number of DNA deviates from its relaxed value Lk_0 , the DNA polymer is called “supercoiled”. Since the same linking number deviation can induce different twisting stress in molecules of different size, it is more appropriate to measure extent of supercoiling by a properly normalized quantity. The supercoiling degree σ is measured by the normalized linking number deviation, i.e.,

$$\sigma = \frac{Lk - Lk_0}{Lk_0}, \quad (130)$$

where as mentioned before, Lk_0 is the linking number of a relaxed B-form DNA of the same length. In living organisms, DNA molecules are usually negatively supercoiled, with supercoiling degree $\sigma \simeq -0.06$.

In this section we investigate the elasticity of the supercoiled DNA molecules by analytical means and Monte Carlo simulations, based on the double-stranded model mentioned in § V B. The main purpose is to understand the experimental data reported by Strick and

co-authors [20, 22], who have pulled individual supercoiled DNA molecules and measured their elastic responses under various external forces and torques.

This section does not discuss the supercoiling property of circular DNA molecules, since studies on this topic have already well documented, see for example, Ref. [186].

A. Analytical approximation

In the experiment of Ref. [20], the two strands of a λ -DNA at one end are bound to a glass cover slip and the two strands at the other end to a magnetic bead. The linking number of λ -DNA was then changed by rotating the magnetic bead with a magnetic field. At the same time, an external force in the order of piconewton was applied to the DNA molecule to make it straight enough.

To model this situation theoretically, Zhou and co-authors [28, 159] introduced a torque energy in the double-stranded DNA model (§ VB). This torque energy is proportional to the total linking number of the DNA double-helix. Here we recall their calculations in some detail.

The total number of topological turns one DNA strand winds around the other, Lk , can be expressed as the sum of the twisting number, $Tw(\mathbf{r}_1, \mathbf{r})$, of backbone \mathbf{r}_1 around the central axis \mathbf{r} and the writhing number, $Wr(\mathbf{r})$, of the central axis [181, 187, 188]:

$$Lk = Tw + Wr. \quad (131)$$

Based on the model of § VB, the twisting number is calculated by the following expression [181, 189, 190]

$$Tw(\mathbf{r}_1, \mathbf{r}) = \frac{1}{2\pi} \int_0^L \mathbf{t} \times \mathbf{b} \cdot \frac{d\mathbf{b}}{ds} ds = \frac{1}{2\pi} \int_0^L \frac{\sin \varphi}{R} ds, \quad (132)$$

where L is the total contour length of a DNA strand and R is the cross-sectional radius of DNA double-helix. The writhing number of the central axis is much more difficult to calculate. It is expressed as the following Gauss integral over the central axis [187, 188]:

$$Wr(\mathbf{r}) = \frac{1}{4\pi} \int \int \frac{d\mathbf{r} \times d\mathbf{r}' \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (133)$$

In the case of linear chains, provided that some fixed direction (for example the direction of the external force, \mathbf{z}_0) can be specified and that the tangent vector \mathbf{t} never points to

$-\mathbf{z}_0$, it was proved by Fuller [191] that the writhing number Eq. (133) can be calculated alternatively according to the following formula,

$$Wr(\mathbf{r}) = \frac{1}{2\pi} \int_0^L \frac{\mathbf{z}_0 \times \mathbf{t} \cdot d\mathbf{t}/ds}{1 + \mathbf{z}_0 \cdot \mathbf{t}} ds. \quad (134)$$

Equation (134) can be further simplified for highly extended linear DNA chains whose tangent \mathbf{t} fluctuates only slightly around \mathbf{z}_0 . In this case, Eq. (134) is approximately expressed as

$$Wr(\mathbf{r}) \simeq \frac{1}{4\pi} \int_0^L \left[-t_y \frac{dt_x}{ds} + t_x \frac{dt_y}{ds} \right] ds, \quad (135)$$

where t_x and t_y are, respectively, the two components of \mathbf{t} with respect to two arbitrarily chosen orthonormal directions (\mathbf{x}_0 and \mathbf{y}_0) on the plane perpendicular to \mathbf{z}_0 .

The energy caused by the external torque of magnitude Γ is then equal to

$$E_{\text{torque}} = -2\pi\Gamma Lk = -2\pi\Gamma(Tw + Wr). \quad (136)$$

This energy term is added to Eq. (126) of the preceding section.

Based on Eq. (155) of the Appendix, the Green equation for a supercoiled and highly stretched dsDNA is expressed as

$$\begin{aligned} \frac{\partial \Psi(\mathbf{t}, \varphi; s)}{\partial s} = & \left[\frac{\partial^2}{4\ell_p^* \partial \mathbf{t}^2} + \frac{\partial^2}{4\ell_p \partial \varphi^2} + \frac{f \cos \varphi}{k_B T} \mathbf{t} \cdot \mathbf{z}_0 - \frac{\rho(\varphi)}{k_B T} - \frac{\ell_p}{R^2} \sin^4 \varphi \right. \\ & \left. + \frac{\Gamma}{Rk_B T} \sin \varphi - \frac{\Gamma}{4k_B T \ell_p^*} \frac{\partial}{\partial \phi} + \frac{\Gamma^2}{16\ell_p^* (k_B T)^2} \sin^2 \theta \right] \Psi(\mathbf{t}, \varphi; s), \end{aligned} \quad (137)$$

where (θ, ϕ) are the two directional angles of \mathbf{t} [159]. Similar to what we have done in § VB, we can now express the above Green equation in matrix form [159]. The ground-state eigenvalue and eigenfunction Φ_0 of Eq. (137) can be obtained numerically for given applied force and torque. The average extension is calculated through by the following formula

$$X = L \int \chi_0(\mathbf{t}, \varphi) \mathbf{t} \cdot \mathbf{z}_0 \cos \varphi \Phi_0(\mathbf{t}, \varphi) d\mathbf{t} d\varphi, \quad (138)$$

where $\chi_0(\mathbf{t}, \varphi)$ is the ground-state left-eigenfunction of Eq. (137). The writhing number Eq. (135) is calculated according to Eq. (164) as

$$\langle Wr \rangle = L \frac{\Gamma}{16\pi \ell_p^* k_B T} \int \chi_0(\mathbf{t}, \varphi) \sin^2 \theta \Phi_0(\mathbf{t}, \varphi) d\mathbf{t} d\varphi, \quad (139)$$

and average linking number is then expressed as

$$\langle Lk \rangle = \langle Tw \rangle + \langle Wr \rangle = \frac{L}{2\pi R} \int \chi_0 \sin \varphi \Phi_0 d\mathbf{t} d\varphi + \frac{L\Gamma}{16\pi \ell_p^* k_B T} \int \chi_0 \sin^2 \theta \Phi_0 d\mathbf{t} d\varphi. \quad (140)$$

The theoretical relationship between extension and supercoiling degree is shown in Fig. 35 and compared with the experimental result of Strick et al. [20]. In obtaining these curves, the values of the parameters are the same as those used in Fig. 33 and no adjustment has been done to fit the experimental data. We find that in the case of negatively supercoiled DNA, the theoretical and experimental results are in quantitative agreement, indicating that the present model is capable of explaining the elasticity of negatively supercoiled DNA; in the case of positively supercoiled DNA, the agreement between theory and experiment is not so good, especially when the external force is relatively large. In this simple model, we have not considered the possible deformations of the nucleotide base-pairs. While this assumption might be reasonable in the negatively supercoiled case, it may fail for positively supercoiled DNA chain, especially at large stretching forces. The work done by Allemand et al. [23] suggested that positive supercoiled and highly extended DNA molecule can take on Pauling-like configurations with exposed bases.

For negatively supercoiled DNA molecule, both theory and experiment reveal the following elastic aspects: (a) When external force is small, DNA molecule can shake off its torsional stress by writhing its central axis, which can lead to an increase in the negative writhing number and hence restore the local folding manner of DNA strands to that of B-form DNA; (b) However, writhing of the central axis causes shortening of DNA end-to-end extension, which becomes more and more unfavorable as the external force is increased. Therefore, at large forces, the torsional stress caused by negative torque (supercoiling degree) begins to unwind the B-form double-helix and triggers the transition of DNA internal structure, where a continuously increasing portion of DNA takes on some certain new configuration as supercoiling increases, while its total extension keeps almost invariant.

What is the new configuration? Strick and co-authors [20] suggested this new configuration corresponds to denatured DNA segments, i.e., negative torque leads to breakage of hydrogen bonds between the complementary DNA bases and consequently to strand-separation. To check this picture, Strick and co-authors [21] used short single-stranded homologous DNA segments to detect possible denaturation bubbles. They found that these homologous DNA probes indeed bind onto negatively supercoiled dsDNA molecules.

As seen in Fig. 35, although the simple theoretical model of this section does not taken into account the possibility of strand-separation, it can quantitatively explain the behavior of negatively supercoiled DNA. It is interesting to see what is the prediction of the new

configuration according to this model. The calculations performed in Ref. [159] suggested that, in negatively supercoiled and highly extended DNA, the two strands of DNA might be in a left-handed double-helix configuration. The structural parameters of the left-handed configuration suggested by the theory are listed in Table III and compared with those of Z-form DNA [12]. The strong similarity in these parameters suggests that the torque-induced left-handed configurations may resemble the Z-form of double-stranded DNA [12, 30]. It may be possible that in this left-handed DNA configuration the base-pairing is not stable. This can explain why short homologous DNA segments can bind to the DNA polymer [21].

The supercoiling property of stretched DNA were also studied theoretically using isotropic models by many others, see, for example, Refs. [170, 192–200]

The anisotropic twisting elasticity of double-stranded DNA was investigated by considering the coupling between DNA spontaneous curvature and twist, see, for example, Refs. [171, 201–204]. Garrivier and Fourcade considered the twist-bend coupling due to the intrinsic curvature in DNA and very interesting numerical results were reported [202]. Panyukov and Rabin also demonstrated by analytical and numerical calculations that, under external stretching, a supercoiled helical ribbon polymer has anisotropic force-extension relationship [171, 203]. Theoretical work on this respect may have direct biological significance, since real DNA molecules has local sequence-dependent intrinsic curvature.

B. DNA supercoiling studied by Monte Carlo simulation

The analytical calculations mentioned in § VIA apply to highly stretched DNA. To go beyond this limit, a powerful method is to use Monte Carlo simulation. Vologodskii and Marko in Ref. [205] studied the supercoiling property of stretched DNA by using an achiral model of dsDNA. To better understand the experimental observations of Strick et al. [20], the chiral model of dsDNA mentioned in § VB and § VIA was studied by Zhang and co-authors [206] following the procedure of Ref. [205]. Here we recall the work of Ref. [206] in some detail to demonstrate the main points in simulating the configurational property of topologically constrained DNA.

1. Discrete DNA 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 $2R$ (Figure 36). 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, θ_i , $i = 1, 2, \dots, N$. Each segment is assigned the same amount of basepairs, n_{bp} , so that the length of the i th segment satisfies

$$\Delta s_i = |\mathbf{r}_i - \mathbf{r}_{i-1}| = 0.34n_{\text{bp}} \frac{\cos \theta_i}{\langle \cos \theta \rangle_0}, \quad (141)$$

where $\langle \dots \rangle_0$ means the thermal average for a relaxed DNA molecule.

According to Eq. (126), the total energy of dsDNA molecule with N segments in our discrete computer model is expressed as

$$E = \alpha \sum_{i=1}^{N-1} \gamma_i^2 + \alpha' \sum_{i=1}^{N-1} (\theta_{i+1} - \theta_i)^2 + \frac{\kappa}{R^2} \sum_{i=1}^N \Delta s_i \sin^3 \theta_i \tan \theta_i + \sum_{j=1}^{N_{\text{bp}}} U(\theta_j) - Fz_N, \quad (142)$$

where γ_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 α corresponds to the persistence length $p = 53$ nm of dsDNA according to the direct discretization of Eq. (126), i.e.,

$$\alpha = \frac{p}{2b} k_B T, \quad (143)$$

where b is the average length of segments.

In earlier approaches [205, 206], the bending rigidity constant α of a discrete chain was determined according to the Kuhn statistic length of wormlike chain, which is twice of the persistence length p . In fact, the Kuhn length l_{kuhn} of a discrete wormlike chain with rigidity α is written as (see, for example, Ref. [206])

$$l_{\text{kuhn}} = b \frac{1 + \langle \cos \gamma \rangle}{1 - \langle \cos \gamma \rangle}, \quad (144)$$

where

$$\langle \cos \gamma \rangle = \frac{\int_0^\pi \cos \gamma \exp(-\alpha \gamma^2 / k_B T) \sin \gamma d\gamma}{\int_0^\pi \exp(-\alpha \gamma^2 / k_B T) \sin \gamma d\gamma}. \quad (145)$$

The rigidity constant α defined in this way is only a function of $m = l_{\text{kuhn}}/b$, the number of links within one Kuhn length. The dependence of α versus m obtained from Eqs. (144) and

(145) is shown numerically in Fig. 37. The rigidity constant α 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. (143), 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. (126) is much more directly determined, especially to the models with complicated elasticity.

The constant α' in the second term of Eq. (142) should be associated with stiffness of the DNA single strand. As an approximation, we have taken here $\alpha' = \alpha$. The unpublished data show that, the second term of Eq. (142) is quite small compared with the other four terms in small torsional deviations. The result of the simulation is not sensitive to α' in these cases.

2. Monte Carlo simulation procedure

The equilibrium sets of dsDNA conformations are constructed using the Metropolis MC procedure [207]. The conformational space is sampled through a Markov process. Three kinds of movements are considered [186, 205, 206]: (1) a randomly chosen segment is under-twisted or overtwisted by an angle λ_1 (Fig. 38a); (2) a portion of the chain is rotated by λ_2 around the axis connecting the two ends of rotated chain (Fig. 38b); (3) the segments from a randomly chosen vertex to the free end are rotated by λ_3 around an arbitrary orientation axis that passes the chosen vertex (Fig. 38c). All these three types of movements satisfy the basic requirement of the Metropolis procedure of microscopic reversibility, and each of them is performed with a probability of 1/3. The value of λ_1 , λ_2 , λ_3 are uniformly distributed over interval $(-\lambda_1^0, \lambda_1^0)$, $(-\lambda_2^0, \lambda_2^0)$ and $(-\lambda_3^0, \lambda_3^0)$, respectively, and λ_1^0 , λ_2^0 and λ_3^0 are chosen to guarantee an appropriate acceptance rate of trial movements.

A trial move from a conformation i to a conformation j is accepted on the basis of the probability $p_{i \rightarrow j} = \min(1, \rho_j/\rho_i)$, where ρ_i is the probability density of conformation i . Energetic importance sampling is realized in the Metropolis MC method by choosing the probability density ρ_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. (142).

The starting conformation of the chain is unknotted. To avoid knotted configuration in the Markov process, we calculate the Alexander polynomial of each trial conformation

[208, 209]. 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 from the points on the axis of a segment to the points on the axis of another nonadjacent segment. If the minimum distance for any two chosen segments is less than the DNA diameter $2R$, the energy of trial conformation is set to be infinite and the movement is then rejected.

During the configurational evolution of DNA, the supercoiling degree σ may distribute around all the possible values. In order to avoid the waste of computational samples and also for the propose of comparison with the supercoiling experiments [20, 22], we bound the supercoiling σ of DNA chain inside the region of $-0.12 \geq \sigma \geq 0.12$. 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 Lk of each conformation is calculated according to Eq. (131), where the twist number Tw [Eq. (132)] can be re-expressed by

$$Tw = \frac{1}{2\pi R} \sum_{i=1}^N \Delta s_i \tan \theta_i. \quad (146)$$

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 (Fig. 39). The writhing number Wr of the linear DNA is equal to that of the closed circular configuration and can be thereby calculated by Eq. (133).

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 and supercoiling degree σ of each accepted conformation of DNA chain are calculated. When the trial movement is rejected, the current conformation is counted up twice [207].

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 σ . For each subsample, we calculate the averaged extension

$$\langle z_N \rangle_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{z_N(i)}{L_B}, \quad j = 1, \dots, 15 \quad (147)$$

and the averaged torsion

$$\langle \sigma \rangle = \frac{1}{N_j} \sum_{i=1}^{N_j} \sigma_i, \quad j = 1, \dots, 15, \quad (148)$$

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

We display in Fig. 40 the averaged extension as a function of supercoiling degree for three 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 the Monte Carlo simulations:

- i). At a low force, the elastic behavior of DNA is symmetric under positive or negative supercoiling. This is understandable, since the DNA torsion is the co-operative result of hydrogen-bond constrained bending of DNA backbones and the base-pair stacking interaction in our model. At very low force, the contribution from applied force and the thermodynamic fluctuation perturbs the folding angle θ of base-pair to derive very little from the equilibrium position, θ_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 between the 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 [196, 205].
- ii). At intermediate force, the folding angle of basepairs are pulled slightly further away from equilibrium value θ_0 where van der Waals potential is not symmetric around θ_0 . So the chiral nature of elasticity of the DNA molecule appears. In the negatively supercoiled region, i.e., $\theta < \theta_0$, the contribution of applied force to the total elastic energy dominates that of the stacking potential. So the extension is insensitive to negative supercoiling degree. On the other hand, the positive supercoiling still tends to contract the molecule.
- iii). 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. DNA is in the extended configuration, and its writhing number is small. The polymer is highly twisted.

The biological significance of DNA supercoiling has been discussed intensively in the literature (see, for example, Ref. [186]). Benham and co-workers [210, 211] quantitatively studied the base-pair opening probability as a function of supercoiling degree based on some

simple models. The approach of Benham et al. [210, 211] was recently extended by Li and Ou-Yang [212] and very interesting predictions were obtained concerning the transcriptional initiation of several genes. It is anticipated that in coming years, the link between DNA supercoiling and gene transcription will be understood more thoroughly.

VII. CONCLUSION

In this chapter we have reviewed recent theoretical work on the statistical mechanics of single-molecule DNA and RNA polymers. The topics covered in this chapter include DNA double-helix denaturation, DNA unzipping, DNA entropic elasticity, DNA over-stretching, DNA supercoiling, RNA secondary structure formation and prediction. The theoretical methods used in these studies include generating functions, path integrals, Green equations, and Monte Carlo simulations. These methods were detailedly demonstrated. Here we summarize the main results of this chapter.

DNA thermal denaturation and force-induced unzipping were discussed in § II. Three different kinds of models were mentioned, namely the de Gennes-Peyrard-Bishop model, the Montanari-Mézard model, and the Poland-Scheraga model. One main conclusion is that the cooperativity of the DNA denaturation transition is largely due to the short-ranged base-pair stacking potential. It was also demonstrated that, measuring the force-extension response curves of DNA unzipping is unlikely to probe DNA sequence information down to the single nucleotide resolution.

RNA secondary structure transition was studied in § III. The elastic properties of random-sequence RNA and weakly-designed RNA were investigated, as well as the glassy phase transition in RNA secondary structures. It was demonstrated that the base-pair stacking interactions could stabilize the structure of RNA and make the RNA hairpin-coil transition a highly co-operative process.

Electrostatic interactions in RNA polymers were quantitatively accounted. The idea of using single-molecule force-extension measurement to identify the secondary structure of RNA was also reviewed in § III.

Following a short section (§ IV) on the entropic elasticity of DNA, § V investigated the over-stretching transition of double-stranded DNA. A double-stranded polymer model was mentioned. The highly co-operative over-stretching transition in DNA is closely related the short-ranged nature of DNA base-pair stacking interactions.

The last section (§ VI) focused on the behavior of a supercoiled and stretched DNA polymer. It was demonstrated that the double-stranded polymer model of § V is able to quantitatively explain the asymmetric elasticity of negatively and positively supercoiled DNA chains. It was also suggested that, under the action of an external force, a negatively

supercoiled DNA might be in a left-handed Z-form configuration.

The mechanical property of DNA directly influences its biological functions. This chapter demonstrated that many of the elastic properties of DNA could be understood by taking into account the particular structure of DNA as a double-stranded, base-pair stacked, helical macromolecule. We should emphasize here that the short-ranged base-pair stacking interaction is very important to understand the mechanical property of DNA. On the one hand, the base-pair stacking interaction makes DNA double-helix considerably stable to small environmental variations; on the other hand, due to this short-ranged interaction, dramatic, abrupt, and reversible structural transitions of DNA could be achieved. This later property may be quite important for a biological organism to survive dramatic environmental changes.

The structure and elasticity of DNA also influences its interaction with proteins. Hopefully in the near future the interactions between DNA and various proteins will also be thoroughly interpreted.

Acknowledgments

We are grateful to Dr. Ulrich Gerland for permitting us to reproduce Fig. 27 and Fig. 29, to Mr. LI Ming for reading and correcting an earlier version of the manuscript, to Mr. Doug Lundberg for permitting us to reproduce Fig. 1. We thank Professor Reinhard Lipowsky and Professor YU Lu for support. H.Z. acknowledges the hospitality of the Institute of Theoretical Physics (ITP) and the Interdisciplinary Center of Theoretical Studies (ICTS) in Beijing, where the main part of this manuscript was written.

Appendix: Path Integral method in Polymer statistical physics

In § IV we have used path integral and green equation technique in studying the elastic property of DNA. To make this review self-contained, in this appendix we review some basic ideas on the application of path integral method to the study of polymeric systems [147, 159, 213–215].

Consider a polymeric string, and suppose its total “arclength” is L , and along each arclength point s one can define a n -dimensional “vector” $\mathbf{r}(s)$ to describe the polymer’s local state at this point. (For example, in the case of a flexible Gaussian chain, \mathbf{r} is a three-dimensional position vector; in the case of a semiflexible chain such as the wormlike chain [158, 161], \mathbf{r} is the unit tangent vector of the polymer and is two-dimensional.) We further assume that the energy density (per unit arclength) of the polymer can be written as the following general form:

$$\rho_e(\mathbf{r}, s) = \frac{m}{2} \left(\frac{d\mathbf{r}}{ds} \right)^2 + \mathbf{A}(\mathbf{r}) \cdot \frac{d\mathbf{r}}{ds} + V(\mathbf{r}), \quad (149)$$

where $V(\mathbf{r})$ is a scalar field and $\mathbf{A}(\mathbf{r})$ is a vectorial field. The total partition function of the system is expressed by the following integration:

$$\Xi(L) = \int \int d\mathbf{r}_f \phi_f(\mathbf{r}_f) G(\mathbf{r}_f, L; \mathbf{r}_i, 0) \phi_i(\mathbf{r}_i) d\mathbf{r}_i, \quad (150)$$

where $\phi_i(\mathbf{r})$ and $\phi_f(\mathbf{r})$ are, respectively, the probability distributions of the vector \mathbf{r} at the initial ($s = 0$) and final ($s = L$) arclength point; $G(\mathbf{r}, s; \mathbf{r}', s')$ is called the Green function, it is defined in the following way:

$$G(\mathbf{r}, s; \mathbf{r}', s') = \int_{\mathbf{r}'}^{\mathbf{r}} \mathcal{D}[\mathbf{r}''(s)] \exp[-\beta \int_{s'}^s ds'' \rho_e(\mathbf{r}'', s'')], \quad (151)$$

where integration is carried over all possible configurations of \mathbf{r}'' , and $\beta = 1/k_B T$ is the Boltzmann coefficient. It can be verified that the Green function defined above satisfies the following relation:

$$G(\mathbf{r}, s; \mathbf{r}', s') = \int d\mathbf{r}'' G(\mathbf{r}, s; \mathbf{r}'', s'') G(\mathbf{r}'', s''; \mathbf{r}', s'), \quad (s' < s'' < s). \quad (152)$$

The total free energy of the system is then expressed as

$$\mathcal{F} = -k_B T \ln \Xi. \quad (153)$$

To calculate the total partition function Ξ , we define an auxiliary function $\Psi(\mathbf{r}, s)$ and call it the wave function because of its similarity with the true wave function of quantum systems. Suppose the value of Ψ at arclength point s is related to its value at s' through the following formula such that

$$\Psi(\mathbf{r}, s) = \int d\mathbf{r}' G(\mathbf{r}, s; \mathbf{r}', s') \Psi(\mathbf{r}', s'), \quad (s > s'). \quad (154)$$

(In fact, the choice of the wave function $\Psi(\mathbf{r}, s)$ is not limited. Any function determined by an integration of the form of Eq. (154) can be viewed as a wave function.)

We can derive from Eqs. (149), (151), and (154) that:

$$\frac{\partial \Psi(\mathbf{r}, s)}{\partial s} = \left[\frac{\nabla_{\mathbf{r}}^2}{2m\beta} - \beta V(\mathbf{r}) + \frac{\mathbf{A}(\mathbf{r}) \cdot \nabla_{\mathbf{r}}}{m} + \frac{\nabla_{\mathbf{r}} \cdot \mathbf{A}(\mathbf{r})}{2m} + \frac{\beta \mathbf{A}^2(\mathbf{r})}{2m} \right] \Psi(\mathbf{r}, s) = \hat{H} \Psi(\mathbf{r}, s). \quad (155)$$

Equation (155) is called the Green equation, it is very similar with the Schrödinger equation of quantum mechanics. However, there is an important difference. In the case of $\mathbf{A}(\mathbf{r}) \neq 0$, the operator \hat{H} in Eq. (155) is not Hermitian. Therefore, in this case the matrix form of the operator \hat{H} may not be diagonalized by unitary matrix.

Denote the eigenvalues and the right-eigenfunctions of Eq. (155) as $-g_i$ and $|i\rangle = \Phi_i(\mathbf{r})$ ($i = 0, 1, \dots$), respectively. Then it is easy to know, from the approach of quantum mechanics, that

$$\Psi(s) = \sum_i e^{-g_i(s-s')} |i(s)\rangle \langle i(s') | \Psi(s')\rangle, \quad (156)$$

where $\langle i| = \chi_i(\mathbf{r})$ ($i = 0, 1, \dots$) denote the left-eigenfunctions of Eq. (155), which satisfy the following relation:

$$\langle i|i'\rangle = \int d\mathbf{r} \chi_i(\mathbf{r}) \Phi_{i'}(\mathbf{r}) = \delta_i^{i'}.$$

In the case where \hat{H} is Hermitian (i.e., $\mathbf{A}(\mathbf{r}) = 0$), then we can conclude that

$$\chi_i(\mathbf{r}) = \Phi_i^*(\mathbf{r}).$$

From Eqs. (150), (154), (155), and (156) we know that

$$\begin{aligned} \Xi(L) &= \int d\mathbf{r} \int d\mathbf{r}' G(\mathbf{r}, L; \mathbf{r}', s') \phi_f(\mathbf{r}) \phi_i(\mathbf{r}') = \sum_i \langle \phi_f | i \rangle \langle i | \phi_i \rangle e^{-g_i L} \\ &= e^{-g_0 L} \langle \phi_f | 0 \rangle \langle 0 | \phi_i \rangle \quad (\text{for } L \gg 1/(g_1 - g_0)). \end{aligned} \quad (157)$$

Consequently, for long polymer chains the total free energy density is just expressed as

$$\mathcal{F}/L = k_B T g_0, \quad (158)$$

and any quantity of interest can then be calculated by differentiation of \mathcal{F} . For example, the average extension of a polymer under external force field f can be calculated as $\langle Z \rangle = \partial\mathcal{F}/\partial f = Lk_B T \partial g_0 / \partial f$.

We continue to discuss another very important quantity, the distribution probability of \mathbf{r} at arclength s , $P(\mathbf{r}, s)$. This probability is calculated from the following expression:

$$P(\mathbf{r}, s) = \frac{\int d\mathbf{r}_f \int d\mathbf{r}_i \phi_f(\mathbf{r}_f) G(\mathbf{r}_f, L; \mathbf{r}, s) G(\mathbf{r}, s; \mathbf{r}_i, 0) \phi_i(\mathbf{r}_i)}{\int d\mathbf{r}_f \int d\mathbf{r}_i \phi_f(\mathbf{r}_f) G(\mathbf{r}_f, L; \mathbf{r}_i, 0) \phi_i(\mathbf{r}_i)}. \quad (159)$$

Based on Eqs. (154) and (156) we can rewrite Eq. (159) in the following form:

$$\begin{aligned} P(\mathbf{r}, s) &= \frac{\int d\mathbf{r}_f \int d\mathbf{r}_i \int d\mathbf{r}' \phi_f(\mathbf{r}_f) G(\mathbf{r}_f, L; \mathbf{r}', s) \delta(\mathbf{r}' - \mathbf{r}) G(\mathbf{r}, s; \mathbf{r}_i, 0) \phi_i(\mathbf{r}_i)}{\int d\mathbf{r}_f \int d\mathbf{r}_i \phi_f(\mathbf{r}_f) G(\mathbf{r}_f, L; \mathbf{r}_i, 0) \phi_i(\mathbf{r}_i)} \\ &= \frac{\sum_m \sum_n \langle \phi_f | m \rangle \langle n | \phi_i \rangle \chi_m(\mathbf{r}) \Phi_n(\mathbf{r}) \exp[-g_m(L-s) - g_n s]}{\sum_m \langle \phi_f | m \rangle \langle m | \phi_i \rangle \exp(-g_m L)} \end{aligned} \quad (160)$$

For the most important case of $0 \ll s \ll L$, Eq. (160) then gives that the probability distribution of \mathbf{r} is independent of arclength s , i.e.,

$$P(\mathbf{r}, s) = \chi_0(\mathbf{r}) \Phi_0(\mathbf{r}) \quad (\text{for } 0 \ll s \ll L). \quad (161)$$

With the help of Eq. (161), the average value of a quantity which is a function of \mathbf{r} can be obtained. For example,

$$\langle Q(s) \rangle = \int d\mathbf{r} Q(\mathbf{r}) P(\mathbf{r}, s) = \int d\mathbf{r} \chi_0(\mathbf{r}) Q(\mathbf{r}) \Phi_0(\mathbf{r}) = \langle 0 | Q | 0 \rangle, \quad (162)$$

and

$$\left\langle \int_0^L Q(\mathbf{r}(s)) ds \right\rangle = \int_0^L \langle Q(s) \rangle ds = L \langle 0 | Q | 0 \rangle \quad (\text{for } L \gg 1/(g_1 - g_0)). \quad (163)$$

Finally, we list the formula for calculating $\langle \mathbf{B}(\mathbf{r}) \cdot d\mathbf{r}/ds \rangle$, here $\mathbf{B}(\mathbf{r})$ is a given vectorial field. The formula reads:

$$\left\langle \mathbf{B}(\mathbf{r}) \cdot \frac{d\mathbf{r}}{ds} \right\rangle = \frac{1}{2} \langle 0 | \left[\mathbf{r} \cdot (\hat{H}\mathbf{B}) - \mathbf{B} \cdot (\hat{H}\mathbf{r}) \right] | 0 \rangle \quad (\text{for } L \gg 1/(g_1 - g_0)). \quad (164)$$

-
- [1] S. B. Smith, L. Finzi, and C. Bustamante, *Science* **258**, 1122 (1992).
- [2] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, *Curr. Opin. Struct. Biol.* **10**, 279 (2000).
- [3] H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H. Gaub, *Curr. Opin. Chem. Biol.* **4**, 524 (2000).
- [4] T. Strick, J.-F. Allemand, V. Croquette, and D. Bensimon, *Physics Today* pp. 46–51 (2001).
- [5] M. C. Williams and I. Rouzina, *Curr. Opin. Struct. Biol.* **12**, 330 (2002).
- [6] M. Hegner and W. Grange, *J. Muscle Res. Cell. Motility* **23**, 367 (2002).
- [7] J. F. Allemand, D. Bensimon, and V. Croquette, *Curr. Opin. Struct. Biol.* **13**, 266 (2003).
- [8] C. Bustamante, Z. Bryant, and S. B. Smith, *Nature (London)* **421**, 423 (2003).
- [9] J. D. Watson and F. H. C. Crick, *Nature (London)* **171**, 737 (1953).
- [10] M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson, *Nature (London)* **171**, 738 (1953).
- [11] R. E. Franklin and R. G. Gosling, *Nature (London)* **171**, 740 (1953).
- [12] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner, *Molecular Biology of the Gene* (Benjamin/Cummings, 1987), 4th ed.
- [13] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry* (Worth Publishers, New York, 2000), 3rd ed.
- [14] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
- [15] J. C. Wang, *Proc. Natl. Acad. Sci. USA* **76**, 200 (1979).
- [16] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, *Science* **265**, 1599 (1994).
- [17] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, *Nature (London)* **424**, 338 (2003).
- [18] S. B. Smith, Y. Cui, and C. Bustamante, *Science* **271**, 795 (1996).
- [19] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. L. Viovy, D. Chatenay, and F. Caron, *Science* **271**, 792 (1996).
- [20] T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, *Science* **271**, 1835 (1996).
- [21] T. R. Strick, V. Croquette, and D. Bensimon, *Proc. Natl. Acad. Sci. USA* **95**, 10579 (1998).
- [22] T. R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, *Biophys. J.* **74**, 2016 (1998).

- [23] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette, Proc. Natl. Acad. Sci. USA **95**, 14152 (1998).
- [24] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
- [25] V. A. Bloomfield, D. M. Crothers, and I. Tinoco Jr, *Physical Chemistry of Nucleic Acids* (Harper & Row, New York, 1974).
- [26] D. Bensimon, A. J. Simon, V. Croquette, and A. Bensimon, Phys. Rev. Lett. **74**, 4754 (1995).
- [27] S. B. Smith, Y. Cui, A. C. Hausrath, and C. Bustamante, Biophys. J. **68**, A250 (1995).
- [28] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **82**, 4560 (1999).
- [29] H. Zhou and Z.-C. Ou-Yang, Mod. Phys. Lett. B **13**, 999 (1999).
- [30] J. F. Leger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, Phys. Rev. Lett. **83**, 1066 (1999).
- [31] T. T. Perkins, D. E. Smith, and S. Chu, Science **276**, 2016 (1997).
- [32] D. E. Smith and S. Chu, Science **281**, 1335 (1998).
- [33] D. E. Smith, H. P. Babcock, and S. Chu, Science **283**, 1724 (1999).
- [34] B. Maier, D. Bensimon, and V. Croquette, Proc. Natl. Acad. Sci. USA **97**, 12002 (2000).
- [35] M. Hegner, S. B. Smith, and C. Bustamante, Proc. Natl. Acad. Sci. USA **96**, 10109 (1999).
- [36] G. J. L. Wuite, S. B. Smith, M. Young, D. Keller, and C. Bustamante, Nature (London) **404**, 103 (2000).
- [37] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. USA **94**, 11935 (1997).
- [38] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).
- [39] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. E **58**, 2386 (1998).
- [40] U. Bockelmann, P. Thomen, B. Essevaz-Roulet, V. Viasnoff, and F. Heslot, Biophys. J. **82**, 1537 (2002).
- [41] C. Danilowicz, V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss, Proc. Natl. Acad. Sci. U.S.A. **100**, 1694 (2003).
- [42] G. U. Lee, L. A. Chrisey, and R. Colton, Science **266**, 771 (1994).
- [43] A. D. MacKerrel and G. U. Lee, Eur. Biophys. J. **28**, 415 (1999).
- [44] A. Noy, D. V. Vezenov, J. F. Kayyem, T. J. Meade, and C. M. Liebe, Chem. Biol. **4**, 519 (1997).
- [45] T. Strunz, K. Oroszlan, R. Schäfer, and H.-J. Guntherodt, Proc. Natl. Acad. Sci. USA **96**,

- 11277 (1999).
- [46] L. H. Pope, M. C. Davies, C. A. Laughton, C. J. Roberts, S. J. B. Tendler, and P. M. Williams, *Eur. Biophys. J.* **30**, 53 (2001).
 - [47] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, *Nature Struct. Biol.* **6**, 346 (1999).
 - [48] H. Clausen-Schaumann, M. Rief, C. Tolksdorf, and H. E. Gaub, *Biophys. J.* **78**, 1997 (2000).
 - [49] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton, *Phys. Rev. Lett.* **90**, 238101 (2003).
 - [50] T. Boland and B. D. Ratner, *Proc. Natl. Acad. Sci. USA* **92**, 5297 (1995).
 - [51] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr, and C. Bustamante, *Science* **292**, 733 (2001).
 - [52] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr, and C. Bustamante, *Science* **296**, 1832 (2002).
 - [53] B. Onoa, S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco Jr, and C. Bustamante, *Science* **299**, 1892 (2003).
 - [54] X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu, *Science* **296**, 1473 (2002).
 - [55] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
 - [56] C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
 - [57] H. Yin, R. Landick, and J. Gelles, *Biophys. J.* **67**, 2468 (1994).
 - [58] H. Yin, M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles, *Science* **270**, 1653 (1995).
 - [59] M. D. Wang, M. J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S. M. Block, *Science* **282**, 902 (1998).
 - [60] H. Yin, I. Artsimovitch, R. Landick, and J. Gelles, *Proc. Natl. Acad. Sci. USA* **96**, 13124 (1999).
 - [61] J. R. Davenport, G. L. Wuite, R. Landick, and C. Bustamante, *Science* **287**, 2497 (2000).
 - [62] G. Charvin, D. Bensimon, and V. Croquette, *Proc. Natl. Acad. Sci. USA* **100**, 9820 (2003).
 - [63] Y. Cui and C. Bustamante, *Proc. Natl. Acad. Sci. USA* **97**, 127 (2000).
 - [64] B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang, *Proc. Natl. Acad. Sci. USA* **99**, 1960 (2002).
 - [65] B. Ladoux, J.-P. Quivy, P. Doyle, O. duRoure, G. Almouzni, and J.-L. Viovy, *Proc. Natl. Acad. Sci. USA* **97**, 14251 (2000).

- [66] M. G. Poirier and J. F. Marko, Proc. Natl. Acad. Sci. USA **99**, 15393 (2002).
- [67] J. F. Leger, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, Proc. Natl. Acad. Sci. USA **95**, 12295 (1998).
- [68] G. V. Shivashankar, M. Feingold, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. USA **96**, 7916 (1999).
- [69] R. Bar-Ziv and A. Libchaber, Proc. Natl. Acad. Sci. USA **98**, 9068 (2001).
- [70] I. Braslavsky, B. Hebert, E. Kartalov, and S. R. Quake, Proc. Natl. Acad. Sci. USA **100**, 3960 (2003).
- [71] S. Cocco, J. F. Marko, and R. Monasson, C. R. Physique **3**, 569 (2002).
- [72] J. F. Marko and S. Cocco, Physics World pp. 37–41 (2003).
- [73] E. Yeramian, Gene **255**, 139 (2000).
- [74] T. Lipniacki, Phys. Rev. E **64**, 051919 (2001).
- [75] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
- [76] P.-G. de Gennes, C. R. Physique **2**, 1505 (2001).
- [77] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers: Statistical Mechanical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules* (Academic Press, New York, 1970).
- [78] P.-G. de Gennes, Rep. Prog. Phys. **32**, 187 (1969).
- [79] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
- [80] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
- [81] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
- [82] J. Xiao, J. Lin, and B. Tian, Phys. Rev. E **50**, 5039 (1994).
- [83] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
- [84] Y. L. Zhang, W. M. Zheng, J. X. Liu, and Y. Z. Chen, Phys. Rev. E **56**, 7100 (1997).
- [85] M. Barbi, S. Cocco, M. Peyrard, and S. Ruffo, J. Biol. Phys. **24**, 97 (1999).
- [86] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, Phys. Rev. Lett. **85**, 6 (2000).
- [87] H. Zhou, *Force-induced melting and thermal melting of a double-stranded biopolymer* (2000), e-print: cond-mat/0007015.
- [88] A. Montanari and M. Mézard, Phys. Rev. Lett. **86**, 2178 (2001).
- [89] S. Lifson, J. Chem. Phys. **40**, 3705 (1964).
- [90] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1464 (1966).

- [91] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
- [92] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **86**, 356 (2001).
- [93] H. Zhou and Y. Zhang, J. Chem. Phys. **114**, 8694 (2001).
- [94] V. M. Pavlov, J. L. Lyubchenko, A. S. Borovik, and Y. Lazurkin, Nucl. Acids. Res. **4**, 4052 (1977).
- [95] A. S. Borovik, Y. A. Kalambet, Y. L. Lyubchenko, V. T. Shitov, and E. Golovanov, Nucl. Acids. Res. **8**, 4165 (1980).
- [96] A. Montrichok, G. Gruner, and G. Zocchi, Europhys. Lett. **62**, 452 (2003).
- [97] S. M. Bhattacharjee, J. Phys. A **33**, L423 (2000).
- [98] D. Marenduzzo, A. Trovato, and A. Maritan, Phys. Rev. E **64**, 031901 (2001).
- [99] I. Rouzina and V. A. Bloomfield, Biophys. J. **77**, 3242 (1999).
- [100] I. Rouzina and V. A. Bloomfield, Biophys. J. **77**, 3252 (1999).
- [101] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
- [102] M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000).
- [103] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).
- [104] T. Garel, C. Monthus, and H. Orland, Eurphys. Lett. **55**, 132 (2001).
- [105] S. M. Bhattacharjee, Europhys. Lett. **57**, 772 (2002).
- [106] M. Baiesi, E. Carlon, E. Orlandini, and A. L. Stella, Eur. Phys. J. B **29**, 129 (2002).
- [107] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
- [108] Y. Kafri, D. Mukamel, and L. Peliti, Eur. Phys. J. B **27**, 135 (2002).
- [109] B. Duplantier, Phys. Rev. Lett. **57**, 941 (1986).
- [110] B. Duplantier, J. Stat. Phys. **54**, 581 (1989).
- [111] L. Schäfer, C. von Ferber, U. Lehr, and B. Duplantier, Nucl. Phys. B **374**, 473 (1992).
- [112] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- [113] A. Hanke and R. Metzler, Phys. Rev. Lett. **90**, 159801 (2003).
- [114] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **90**, 159802 (2003).
- [115] T. Odijk, Macromolecules **28**, 7016 (1995).
- [116] F. Bueche, *Physical Properties of Polymers* (Interscience, New York, 1962).
- [117] S. Cocco, J. F. Marko, and R. Monasson, Eur. Phys. J. E **10**, 153 (2003).
- [118] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000).

- [119] D. K. Lubensky and D. R. Nelson, Phys. Rev. E **65**, 031917 (2002).
- [120] S. Cocco, R. Monasson, and J. F. Marko, Phys. Rev. E **66**, 051914 (2002).
- [121] S. Cocco, R. Monasson, and J. F. Marko, Phys. Rev. E **65**, 041907 (2002).
- [122] F. Ritort, C. Bustamante, and I. Tinoco Jr, Proc. Natl. Acad. Sci. USA **99**, 13544 (2002).
- [123] D. Marenduzzo, S. M. Bhattacharjee, A. Maritan, E. Orlandini, and F. Seno, Phys. Rev. Lett. **88**, 028102 (2002).
- [124] S.-J. Chen and K. A. Dill, Proc. Natl. Acad. Sci. USA **97**, 646 (2000).
- [125] W. Zhang and S.-J. Chen, Proc. Natl. Acad. Sci. USA **99**, 1931 (2002).
- [126] Z. Wang and K. Zhang, in *Current Topics in Computational Molecular Biology*, edited by T. Jiang, Y. Xu, and M. Q. Zhang (Tsinghua University Press, Beijing, China, 2002), chap. 14, pp. 345–363.
- [127] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
- [128] I. Tinoco Jr and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).
- [129] F. Krzakala, M. Mézard, and M. Müller, Europhys. Lett. **57**, 752 (2002).
- [130] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. pp. 2026–2029 (2000).
- [131] R. Bundschuh and T. Hwa, Phys. Rev. E **65**, 031903 (2002).
- [132] R. Bundschuh and T. Hwa, Europhys. Lett. **59**, 903 (2002).
- [133] E. Marinari, A. Pagnani, and F. Ricci-Tersenghi, Phys. Rev. E **65**, 041919 (2002).
- [134] M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2002).
- [135] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [136] C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. USA **100**, 10249 (2003).
- [137] S. Sugai and K. Nitta, Biopolymers **12**, 1363 (1973).
- [138] J. J. Delrow, J. A. Gebe, and J. M. Schurr, Biopolymers **42**, 455 (1997).
- [139] K. Klenin, H. Merlitz, and J. Langowski, Biophys. J. **74**, 780 (1998).
- [140] Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **81**, 1133 (2001).
- [141] M. N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **89**, 248102 (2002).
- [142] S. A. Rice and M. Nagasawa, *Polyelectrolyte solutions* (Academic Press, New York, 1961).
- [143] P. Pierce, Ph.D. thesis, Department of Chemistry, Yale University (1958).
- [144] S. L. Brenner and V. A. Parsegia, Biophys. J. **14**, 327 (1974).

- [145] D. Stigter, *Biophys. J.* **16**, 1435 (1977).
- [146] I. S. Gradshteyn and I. M. Ryzhik, *Table of integral, series, and products* (Academic Press, New York, 1980).
- [147] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, UK, 1986).
- [148] N. Becker, E. Oroudjev, S. Mutz, J. P. Cleveland, P. K. Hansma, C. Y. Hayashi, D. E. Makarov, and H. G. Hansma, *Nature Materials* **2**, 278 (2003).
- [149] G. A. Jeffrey, *An Introduction to hydrogen bonding* (Oxford University Press, New York, 1997).
- [150] M. Zuker and P. Stiegler, *Nucleic Acids Res.* **9**, 133 (1981).
- [151] J. S. McCaskill, *Biopolymers* **29**, 1105 (1990).
- [152] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker, *Proc. Natl. Acad. Sci. USA* **91**, 9218 (1994).
- [153] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, *Monatshefte für Chemie* **125**, 167 (1994).
- [154] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, *J. Mol. Biol.* **288**, 911 (1999).
- [155] I. L. Hofacker, *Nucleic Acids Res.* **31**, 3429 (2003).
- [156] U. Gerland, R. Bundschuh, and T. Hwa, *Biophys. J.* **81**, 1324 (2001).
- [157] U. Gerland, R. Bundschuh, and T. Hwa, *Mechanical pulling through a nanopore can reveal the secondary structure of single rna molecules*, e-print: cond-at/0306126 (2003).
- [158] J. F. Marko and E. D. Siggia, *Macromolecules* **28**, 8759 (1995).
- [159] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, *Phys. Rev. E* **62**, 1045 (2000).
- [160] J. Wilhelm and E. Frey, *Phys. Rev. Lett.* **77**, 2581 (1996).
- [161] K. Kroy and E. Frey, *Phys. Rev. Lett.* **77**, 306 (1996).
- [162] J. Samuel and S. Sinha, *Phys. Rev. E* **66**, 050801 (2002).
- [163] J. Kierfeld and R. Lipowsky, *Europhys. J.* **62**, 285 (2003).
- [164] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, *Biophys. J.* **80**, 874 (2001).
- [165] I. Rouzina and V. A. Bloomfield, *Biophys. J.* **80**, 882 (2001).
- [166] I. Rouzina and V. A. Bloomfield, *Biophys. J.* **80**, 894 (2001).
- [167] A. Lebrun and R. Lavery, *Nucleic Acids Res.* **24**, 2260 (1996).
- [168] P. Cizeau and J. L. Viovy, *Biopolymers* **42**, 383 (1997).

- [169] A. Ahsan, J. Rudnick, and R. Bruinsma, *Biophys. J.* **74**, 132 (1998).
- [170] J. F. Marko, *Phys. Rev. E* **57**, 2134 (1998).
- [171] S. Panyukov and Y. Rabin, *Europhys. Lett.* **57**, 512 (2002).
- [172] P. Y. Lai and Z. C. Zhou, *Physica A* **321**, 170 (2003).
- [173] C. Storm and P. C. Nelson, *Phys. Rev. E* **67**, 051896 (2003).
- [174] R. L. Ornstein, R. Rein, D. L. Breen, and R. D. MacElroy, *Biopolymers* **17**, 2341 (1978).
- [175] T. B. Liverpool, R. Golestanian, and K. Kremer, *Phys. Rev. Lett.* **80**, 405 (1998).
- [176] R. Golestanian and T. B. Liverpool, *Phys. Rev. E* **62**, 5488 (2000).
- [177] Z. Zhou and P. Y. Lai, *Chem. Phys. Lett.* **346**, 449 (2001).
- [178] B. Mergell, M. R. Ejtehadi, and R. Everaers, *Phys. Rev. E* **66**, 011903 (2002).
- [179] B. Mergell, M. R. Ejtehadi, and R. Everaers, *Phys. Rev. E* **68**, 021911 (2003).
- [180] P. Y. Lai and Z. C. Zhou, *J. Chem. Phys.* **118**, 11189 (2003).
- [181] F. B. Fuller, *Proc. Natl. Acad. Sci. USA* **68**, 815 (1971).
- [182] H. G. Patterton and T. von Holt, *J. Mol. Biol.* **229**, 623 (1993).
- [183] R. M. Harland, H. Weintraub, and S. L. Mcknight, *Nature (London)* **302**, 38 (1983).
- [184] H. Weintraub, *Cell* **32**, 1191 (1983).
- [185] E. K. Hoffman, S. P. Trusko, M. Murphy, and D. L. George, *Proc. Natl. Acad. Sci. USA* **87**, 2705 (1990).
- [186] A. V. Vologodskii, *Topology and Physics of Circular DNA* (CRC Press, Boca Roton, 1992).
- [187] G. Caqlugareanu, *Rev. Math. Pure Appl.* **4**, 5 (1959).
- [188] J. H. White, *American J. Math.* **9**, 693 (1969).
- [189] F. H. C. Crick, *Proc. Natl. Acad. Sci. USA* **73**, 2639 (1976).
- [190] J. H. White and J. M. Bauer, *Proc. Natl. Acad. Sci. USA* **85**, 772 (1988).
- [191] F. B. Fuller, *Proc. Natl. Acad. Sci. USA* **75**, 3557 (1978).
- [192] J. F. Marko and E. D. Siggia, *Phys. Rev. E* **52**, 2912 (1995).
- [193] J. F. Marko, *Phys. Rev. E* **55**, 1758 (1997).
- [194] J. F. Marko, *Europhys. Lett.* **38**, 183 (1997).
- [195] B. Fain, R. Rudnick, and S. Östlund, *Phys. Rev. E* **55**, 7364 (1997).
- [196] C. Bouchiat and M. Mézard, *Phys. Rev. Lett.* **80**, 1556 (1998).
- [197] C. Bouchiat and M. Mézard, *Eur. Phys. J. E* **2**, 377 (2000).
- [198] R. D. Kamien, T. C. Lubensky, P. Nelson, and C. S. O'hern, *Europhys. Lett.* **38**, 237 (1997).

- [199] J. D. Moroz and P. Nelson, Proc. Natl. Acad. Sci. USA **94**, 14418 (1997).
- [200] J. D. Moroz and P. Nelson, Macromolecules **31**, 6333 (1998).
- [201] H. Zhou and Z.-C. Ou-Yang, J. Chem. Phys. **110**, 1247 (1999).
- [202] D. Garrivier and B. Fourcade, Europhys. Lett. **49**, 390 (2000).
- [203] S. Panyukov and Y. Rabin, Phys. Rev. Lett. **85**, 2404 (2000).
- [204] S. Panyukov and Y. Rabin, Phys. Rev. E **62**, 7135 (2000).
- [205] A. V. Vologodskii and J. F. Marko, Biophys. J. **73**, 123 (1997).
- [206] Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **78**, 1979 (2000).
- [207] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
- [208] A. V. Vologodskii, A. V. Lukashin, M. D. Frank-Kamenetskii, and V. V. Anshelevich, Sov. Phys. JETP **39**, 1059 (1974).
- [209] B. A. Harris and S. C. Harvey, J. Comput. Chem. **20**, 813 (1999).
- [210] W. R. Bauer and C. J. Benham, J. Mol. Biol. **234**, 1184 (1993).
- [211] W. R. Bauer, H. Ohtsubo, E. Ohtsubo, and C. J. Benham, J. Mol. Biol. **253**, 438 (1995).
- [212] M. Li and Z.-C. Ou-Yang, unpublished.
- [213] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw Hill, New York, 1965).
- [214] F. W. Wiegand, *Introduction to Path-Integral Methods in Physics and Polymer Sciences* (World Scientific, Singapore, 1986).
- [215] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics* (World Scientific, Singapore, 1990).

TABLE I: The effective linear charge density ν (in unit of e/nm) of DNA molecules, calculated from the comparison of the Poisson-Boltzmann solution and the modified Bessel function (see text). α and β are the parameters of Eq. (93) fitted to the data of ν . [Y. Zhang, H. Zhou, and Z.-c. Ou-Yang, Biophys. J. **81**, 1133-1143 (2001). Copyright (2001) by the Biophysical Society of USA.]

Ionic Concentration c_0 (M)	ssDNA		dsDNA	
	NaCl	MgCl ₂	NaCl	MgCl ₂
1.	4.18	9.50	91.85	993.16
0.75	3.50	6.74	56.15	410.67
0.5	2.84	4.51	31.22	144.10
0.2	2.04	2.31	11.73	24.52
0.15	1.89	1.97	9.29	16.22
0.1	1.73	1.64	7.02	9.82
0.05	1.53	1.27	4.78	4.98
0.02	1.37	0.99	3.29	2.66
0.01	1.29	0.86	2.66	1.91
0.005	1.23	0.78	2.26	1.45
0.002	1.17	0.71	1.93	1.13
0.001	1.14	0.67	1.76	1.00
α	0.0338	-0.577	0.300	-0.505
β	1.36	2.80	4.18	7.33

TABLE II: The base-pair stacking energy of ten possible DNA base-pair stacking patterns based on quantum chemical calculations [174]. In the table, the arrows are pointing from the 5'-end to 3'-end of DNA. The stacking energy is in units of $k_B T$ with $T = 300 K$.

DNA dimer	stacking energy	DNA dimer	stacking energy
$\begin{array}{c} \uparrow \\ \text{C} \cdot \text{G} \\ \downarrow \\ \text{G} \cdot \text{C} \end{array}$	-24.09	$\begin{array}{c} \uparrow \\ \text{C} \cdot \text{G} \\ \downarrow \\ \text{A} \cdot \text{T} \\ \downarrow \\ \text{G} \cdot \text{C} \end{array}$	-17.35
$\begin{array}{c} \uparrow \\ \text{C} \cdot \text{G} \\ \downarrow \\ \text{T} \cdot \text{A} \\ \downarrow \\ \text{G} \cdot \text{C} \end{array}$	-16.20	$\begin{array}{c} \uparrow \\ \text{G} \cdot \text{C} \\ \downarrow \\ \text{C} \cdot \text{G} \\ \downarrow \\ \text{T} \cdot \text{A} \end{array}$	-16.00
$\begin{array}{c} \uparrow \\ \text{G} \cdot \text{C} \\ \downarrow \\ \text{G} \cdot \text{C} \\ \downarrow \\ \text{G} \cdot \text{C} \end{array}$	-13.64	$\begin{array}{c} \uparrow \\ \text{A} \cdot \text{T} \\ \downarrow \\ \text{A} \cdot \text{T} \\ \downarrow \\ \text{G} \cdot \text{C} \end{array}$	-10.85
$\begin{array}{c} \uparrow \\ \text{G} \cdot \text{C} \\ \downarrow \\ \text{T} \cdot \text{A} \\ \downarrow \\ \text{A} \cdot \text{T} \end{array}$	-10.85	$\begin{array}{c} \uparrow \\ \text{G} \cdot \text{C} \\ \downarrow \\ \text{A} \cdot \text{T} \\ \downarrow \\ \text{A} \cdot \text{T} \end{array}$	-11.20
$\begin{array}{c} \uparrow \\ \text{A} \cdot \text{T} \\ \downarrow \\ \text{A} \cdot \text{T} \end{array}$	-8.87	$\begin{array}{c} \uparrow \\ \text{T} \cdot \text{A} \\ \downarrow \\ \text{C} \cdot \text{G} \\ \downarrow \\ \text{T} \cdot \text{A} \end{array}$	-6.31

TABLE III: Structural parameters for the suggested torque-induced left-handed DNA configurations at different external force. d , the average rise per base-pair; num, the number of base-pairs per turn of helix. The last row contains the corresponding values for Z-form DNA. [H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. E **62**, 1045-1058 (2000). Copyright (2000) by the American Physical Society.]

force (pN)	torque ($k_B T$)	d (Å)	pitch (Å)	num
1.3	-5.0	3.59	41.20	11.48
1.0	-5.0	3.57	40.93	11.44
1.3	-4.0	3.83	46.76	12.19
1.0	-4.0	3.82	46.38	12.15
Z-form:		3.8	45.6	12

FIG. 1: Schematic representation of DNA. The two strands of DNA runs in opposite direction (indicated by the arrows along the phosphate-sugar back-bones), and are bound together by many G-C and A-T base-pairs. The two strands fold into a double-helix because of base-pair stacking interactions. In the relaxed state, each pitch of the double-helix contains 10.5 base-pairs and the vertical distance between two neighboring base-pairs is 0.34 nm. [Adapted from <http://ns2.d20.co.edu/kadets/lundberg/dna.html>].

FIG. 2: Clover leaf model of the base-pairing pattern of a tRNA molecule. Filled dots denote nucleotides and thin lines denote base-pairs.

FIG. 3: Phase diagram for a double-stranded biopolymer with an asymmetric potential. This system shows second order thermal melting behavior at zero force. Inset shows how the threshold force for melting changes with temperature T .

FIG. 4: DNA denaturation and unzipping based on the Montanari-Mézard model. Each DNA strand is modeled as a deformable freely-jointed chain, there is on-site short-range attractive interaction between the i -th pair of bases if their separation is less than distance a . (Left) DNA thermal melting (denaturation), with the formation of denaturation bubbles. (Right) An external force may act on the last base-pairs to pull the the two strands apart (unzipping).

FIG. 5: The melting curves of a hypothetical DNA made of 50% G-C base-pairs and 50% A-T base-pairs. The experimental melting temperature is about 90 °C. The parameters used in the calculations are: $a = 4.194 \text{ \AA}$, $b = 1.7 \text{ nm}$, $l_0 = 1.105 \text{ \AA}$. In these different curves, the base-pairing energy ϵ_0 and base-pair stacking energy ϵ_s are set to $(\epsilon_0, \epsilon_s) = (1.247, 10.040)$, $(2.494, 8.768)$, $(3.742, 7.483)$, $(4.989, 6.178)$, $(7.483, 3.455)$, and $(9.977, 0.361)$ from top to bottom (here the energy is in units of kJ/mol).

FIG. 6: The phase-diagrams of a model DNA molecule made of 50% G-C base-pairs and 50% A-T base-pairs. In the calculation, $\epsilon_0 = 3.742 \text{ kJ/mol}$ and $\epsilon_s = 7.483 \text{ kJ/mol}$ (for the dashed curve) and $\epsilon_s = \delta H_s - T\delta S_s$ (for the solid curve), where $\delta H_s = 72.893 \text{ kJ/mol}$ and $\delta S_s = 0.180 \text{ kJ/(mol K)}$. All the other parameters are the same as those listed in the caption of Fig. (5).

FIG. 7: A schematic configuration of a DNA polymer in the Poland-Scheraga model. In this example, there are $s = 3$ denaturation bubbles (length j_1 , j_2 , and j_3) and $s + 1 = 4$ double-helical segments (length i_0 , i_1 , i_2 , and i_3).

FIG. 8: A dsDNA of $2(N + n)$ bases, with a loop of $2n$ bases embedded in the middle. When both N and n are very large, the stem segments AB and CD and the two strands in the loop BCB could be regarded as flexible chains.

FIG. 9: Schematic representation of the basic experimental setup of Essevaz-Roulet et al. [37–39]. The two strands on one end of a λ -DNA was attached to a movable glass slide and to a micro-needle. The strand separation process was followed by moving the glass slide with constant velocity and recording the time-dependent distance x between the glass slide and the micro-needle and the bending extent x_0 of the micro-needle.

FIG. 10: The energy landscape when the total displacement of the system is fixed at $18.1 \mu\text{m}$ (solid curve) and at $18.2 \mu\text{m}$ (dotted curve).

FIG. 11: The free energy cost (in units of $k_B T$) due to unzipping n base-pairs of a random-sequence DNA. In drawing the figure, $\delta\epsilon_0$ of Eq. (46) is set to $0.05 k_B T$ and σ of Eq. (47) is set to $0.5 k_B T$.

FIG. 12: Typical structural fluctuations of a poly(GC) hairpin of 28 nucleotide bases. The external force is fixed at $f = 15.95\text{pN}$. The position of the opening fork as a function of time is shown.

FIG. 13: The life time distribution for a poly(GC) hairpin in the open configurations. The distribution could be well fitted by an exponential distribution (the dotted line) with mean reciprocal life time of $4.94 \pm 0.10 \text{ s}^{-1}$.

FIG. 14: The life time distribution for a poly(GC) hairpin in the closed configurations. The distribution is fitted with an exponential distribution with mean reciprocal life time of $3.19 \pm 0.07 \text{ s}^{-1}$.

FIG. 15: Hartree diagram on the secondary structure of RNA polymer. The thick lines such as $\{i, j\}$ denote the partition function of the RNA segment from base i to base j , the thin segments denote a single nucleotide monomer, and the dotted lines denote base-pairing interactions.

FIG. 16: Force extension curve of random-sequence RNA. The experimental data come from Ref. [2] (pluses) and Ref. [34] (squares), and the ionic concentration is 5 mM Mg^{2+} and 50 mM Na^+ . The theoretical parameters are $a = 4.194 \text{ \AA}$, $b = 1.7 \text{ nm}$, $l_0 = 1.105 \text{ \AA}$, $T = 300 \text{ K}$, $\epsilon_0 = 7.482 \text{ kJ/mol}$. In the figure, $\gamma = (4\pi a^3/3b^3)(e^{\beta\epsilon_0} - 1)$ and J characterize the average base-pair stacking interaction strength. [H. Zhou and Y. Zhang, J. Chem. Phys. **114**, 8694 (2001). Copyright (2001) by the American Institute of Physics.]

FIG. 17: Phase-diagram of a weakly-designed RNA polymer. The system has three phases, hairpin, molten, and stretched-coil. The hairpin-coil phase-transition is first-order (solid line), while hairpin-molten and molten-coil transitions are both second-order (dotted line and dashed line). The theoretical parameters are set to: $a = 0.4194 \text{ nm}$, $b = 1.7 \text{ nm}$, $T_0 = 300 \text{ K}$, $\epsilon_0 = 3.0 k_B T_0$, $\epsilon_1 = 3.6 k_B T_0$.

FIG. 18: (A) Schematic representation of a weakly designed RNA with the last two bases forming a base-pair. (B) Schematic representation of a weakly designed RNA with the last two bases does not form a base-pair.

FIG. 19: Force-extension relationship of a designed RNA polymer at different temperatures. When temperature is lower than $T_c = 318.881 \text{ K}$, stretching RNA cause a first-order hairpin-stretched coil phase-transition. Therefore there is a force-plateau in each force curve. When temperature is higher than T_c , stretching cause a second-order The values of the theoretical parameters are listed in the caption of Fig. 17.

FIG. 20: Electrostatic potential of ssDNA cylinder versus the radial distance from the cylinder axis in the solutions of 2 mM NaCl and 5 mM MgCl_2 . The solid and dashed curves are the numerical solutions of Poisson-Boltzmann equation (P-B); the dotted and dashed-dotted denote corresponding Debye-Huckel approximations (G-H) with effective linear charge density ν along the axis listed in Table I. [Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **81**, 1133-1143 (2001). Copyright (2001) by the Biophysical Society of USA.]

FIG. 21: Extension of ssDNA molecules normalized by the contour length of an equivalent dsDNA molecule observed in 10 mM PB. Solid circles are from Dessinges et al. [141], where a physico-chemical treatment has been employed on the molecule for the propose of suppressing the pairing interactions of the complementary bases; open circles are from Rief et al., which was pulled with an AFM at high force [47]. The continuous line is the result of a numerical simulation of the model described in Eq. (94) with $Y = 120 k_B T / nm^2$. To fit the results at high forces ($F > 70$ pN), a nonlinear (quartic elastic term with $Y_2 = 400 k_B T / nm^4$ was added. The dashed line is the predictions of the eFJC model without electrostatic force included. [Adapted from M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **89**, 248102 (2002). Copyright (2002) by the American Physical Society.]

FIG. 22: Stretching of a single ssDNA charomid of extension $5.7 \mu\text{m}$ in 1mM phosphate buffer (PB) (triangle), 10mM PB (square), and 0.5mM Mg^{++} (circle). Data are taken from [141]. The continuous curves are the results of Monte Carlo simulations of eFJC model [140] with a node-pairing strength of $V_P = 4.6 k_B T$. [Adapted from M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **89**, 248102 (2002). Copyright (2002) by the American Physical Society.]

FIG. 23: Stretching of two different ssDNA molecules: A 50% GC-rich charomid (solid circle) and a 30% GC-rich pX Δ II (open circle) in 10 mM Tris buffer [141]. The continuous curves are the results of MC simulations of eFJC model plus a pair potential with $V_P = 4.6 k_B T$ for the charomid and $V_P = 3.8 k_B T$ for pX Δ II. [Adapted from M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **89**, 248102 (2002). Copyright (2002) by the American Physical Society.]

FIG. 24: Typical conformations of ssDNA chains with random sequence (left) and designed poly(A-T) sequence (right) in 150 mM NaCl solution stretched at different external forces, which are produced in the Monte Carlo simulation of 60 nodes. The fictitious bases and backbone are expressed by green nodes and lines, with the nodes of two ends denoted by two bigger red globes. The hydrogen bonds of base pairs are denoted by magenta short lines. [Y. Zhang, H. Zhou, and Z.-c. Ou-Yang, Biophys. J. **81**, 1133-1143 (2001). Copyright (2001) by the Biophysical Society of USA.]

FIG. 25: Stretching of designed poly(A-T) sequence (a) and poly(dG-dC) sequence (b). The data are taken from [47]; the lines are the Monte Carlo simulations of an eFJC model with base-pairing and base-stacking potentials. [Y. Zhang, H. Zhou, and Z.-c. Ou-Yang, *Biophys. J.* **81**, 1133-1143 (2001). Copyright (2001) by the Biophysical Society of USA.]

FIG. 26: An schematic RNA configuration with two helical segments and one pseudo-knot.

FIG. 27: (a) Simulated force-extension curve for a group I intron (solid line) and that for a homogeneous RNA chain (dashed line). The native structure of the intron sequence is also shown. (b) The force-extension curve for a DNA hairpin of random sequence (solid line) and that of a homogeneous DNA hairpin (dashed line). (c) Mean number of external stems for the group I intron. [U. Gerland, R. Bundschuh, and T. Hwa, *Biophys. J.* **81**, 1324-1332 (2001). Copyright (2001) by the Biophysical Society of USA.]

FIG. 28: (A) Compensation effect in RNA unfolding. One end of a RNA chain is fixed and the other end is pulled with a stiff force lever. At extension X , hairpin with index 1 is opened; at a slightly increased extension $X + \delta X$, hairpin with index 5 is also destroyed, which leads to a transient drop in the pulling force. This drop in pulling force makes it favorable for hairpin 1 to be closed again, resulting to an increase in the pulling force. This compensation effect partially accounts for the smooth behavior of force-extension curves in RNA pulling experiment. (B) To use pulling experiment to explore the native structure of a RNA chain, Gerland and co-workers [157] suggested the use of a nano-pore. Only single-stranded RNA segment could pass through the nano-pore. When applied with a external force F , hairpin 1 will first be opened, followed by hairpin 2, and then hairpin 3 and so on. Because the hairpins break in a sequential order, from the recorded force-extension curve, the secondary structure of RNA could be inferred.

FIG. 29: Typical simulated force-extension curves for pulling a RNA polymer through a nano-pore. [Figure reproduced from U. Gerland, R. Bundschuh, and T. Hwa, e-print: cond-mat/0306126 (2003), with permission.]

FIG. 30: The force-extension curve for a wormlike-chain model and the comparison with those of the Gaussian model and the freely-jointed-chain model. The extension is in units of the total contour length (for the case of Gaussian polymer, it is in units of Nb , where N is the number of bonds). The model parameters is $T = 300 K$ and $\ell_p = 53.0 \text{ nm}$ (correspondingly, $b = 106.0 \text{ nm}$). The variational curve is calculated from Eq. (104) and the interpolation curve is calculated from Eq. (112). (A) linear-logarithmic scale and (B) linear-linear scale.

FIG. 31: Theoretical force-extension profile of DNA based on Eq. (116). In the calculation, ΔE is set to 8.4 kJ/mol per nucleotide pair as in Ref. [19] and the other parameters are given in the main text. The cooperation factor ω is set to $\omega = 15.0 \text{ kJ/mol}$ (solid line), $\omega = 10.0 \text{ kJ/mol}$ (dashed line), $\omega = 5.0 \text{ kJ/mol}$ (dotted line), and $\omega = 1.0 \text{ kJ/mol}$ (dot-dashed line).

FIG. 32: The double-stranded DNA model of Zhou et al. [28]. DNA is modeled as two inextensible worm-like chain winding along a common central axis. The two strands are bound together by many rigid rods which mimic base-pairs.

FIG. 33: Force-extension relation of DNA. Experimental data is from Fig. 2A of Ref. [19]. Theoretical curve is obtained by the following considerations: (i) $\ell_p = 1.5 \text{ nm}$ and $\epsilon = 14.0k_B T$; (ii) $\ell_p^* = 53.0/2\langle\cos\varphi\rangle_{f=0} \text{ nm}$, $r_0 = 0.34/\langle\cos\varphi\rangle_{f=0} \text{ nm}$ and $R = (0.34 \times 10.5/2\pi)\langle\tan\varphi\rangle_{f=0} \text{ nm}$; (iii) adjust the value of φ_0 to fit the data. For each φ_0 , the value of $\langle\cos\varphi\rangle_{f=0}$ is obtained self-consistently. The present curve is drawn with $\varphi_0 = 62.0^\circ$ (in close consistence with the structural property of DNA), and $\langle\cos\varphi\rangle_{f=0}$ is determined to be 0.573840 . DNA extension is scaled with its B-form contour length $L\langle\cos\varphi\rangle_{f=0}$. [H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **82**, 4560-4563 (1999). Copyright (1999) by The American Physical Society.]

FIG. 34: Low-force elastic behavior of DNA. Experimental data is from Fig. 5B of Ref. [1], the dotted curve is obtained for a wormlike chain with bending persistence length of 53.0 nm and the parameters for the solid curve are the same as those in Fig. 33. [H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **82**, 4560-4563 (1999). Copyright (1999) by The American Physical Society.]

FIG. 35: Extension vs supercoiling relations at fixed pulling forces for torsionally constrained DNA. The parameters for the curves are the same as Fig. 33 and experimental data is from Fig. 3 of Ref. [20] (symbols). [H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **82**, 4560-4563 (1999). Copyright (1999) by the American Physical Society.]

FIG. 36: A 3-segment configuration of discrete DNA chain in Monte Carlo simulation. [Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **78**, 1979-1987 (2000). Copyright (2000) by the Biophysical Society of USA.]

FIG. 37: The bending rigidity constant α of a discrete wormlike chain as function of the number of segments within one Kuhn length $m = l_{\text{kuhn}}/b$. The circle points denote the result calculated from Eqs. (144) and (145); the solid line from direct discretization of Eq. (126).

FIG. 38: Movements of dsDNA chain during Monte Carlo simulations. The current conformation of the DNA central axis is shown by the solid lines and the trial conformations by dashed lines. (a) The folding angle θ_i in the i th segment is changed into $\theta_i + \lambda_1$. All segments between i th vertex and the free end are translated by the distance of $|\Delta s_i - \Delta s'_i|$. (b) A portion of the chain is rotated by an angle of λ_2 around the axis connecting the two ends of rotated chain. (c) The segments from a randomly chosen vertex to the free end are rotated by an angle λ_3 around an arbitrary orientation axis that passes the chosen vertex. [Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, Biophys. J. **78**, 1979-1987 (2000). Copyright (2000) by the Biophysical Society of USA.]

FIG. 39: The schematic diagram to calculate link number in our simulations. (a). For a linear supercoiled DNA chain with one end attached to a microscope slide and with the other end attached to a magnetic bead, when the orientation of the bead is fixed and the DNA chain is forbidden to go round the bead, the number of times for two strands to interwind each other, the linking number of the linear DNA (Lk_l), is a topological constant. (b). The DNA double helix is stretched to a fully extended form while the orientation of bead keeps unchanged. The link number of linear DNA chain is equal to the twist number, i.e. $Lk_l = Tw_l$. (c). Three long flat ribbons are connected to the two ends of the linear twisted DNA of (b). The link number of the new double helix circle is equal to that of linear DNA chain, i.e. $Lk_c = Tw_c = Tw_l = Lk_l$ since the writhe of the rectangle loop is 0. (d). The DNA circle in (c) can be deformed into a new circle, one part of which has the same steric structure as the linear supercoiled DNA chain in (a). So by adding three straight ribbons, the link number of linear double helix DNA can be obtained by calculating the link number of the new DNA circle, i.e. $Lk_l = Lk_c = Tw + Wr$. [Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, *Biophys. J.* **78**, 1979-1987 (2000). Copyright (2000) by the Biophysical Society of USA.]

FIG. 40: Relative extension versus supercoiling degree of DNA polymer for three typical stretch forces. Open points denote the experimental data [22], and solid points the results of Monte Carlo simulation of the discrete dsDNA model [206]. 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. [Y. Zhang, H. Zhou, and Z.-C. Ou-Yang, *Biophys. J.* **78**, 1979-1987 (2000). Copyright (2000) by the Biophysical Society of USA.]

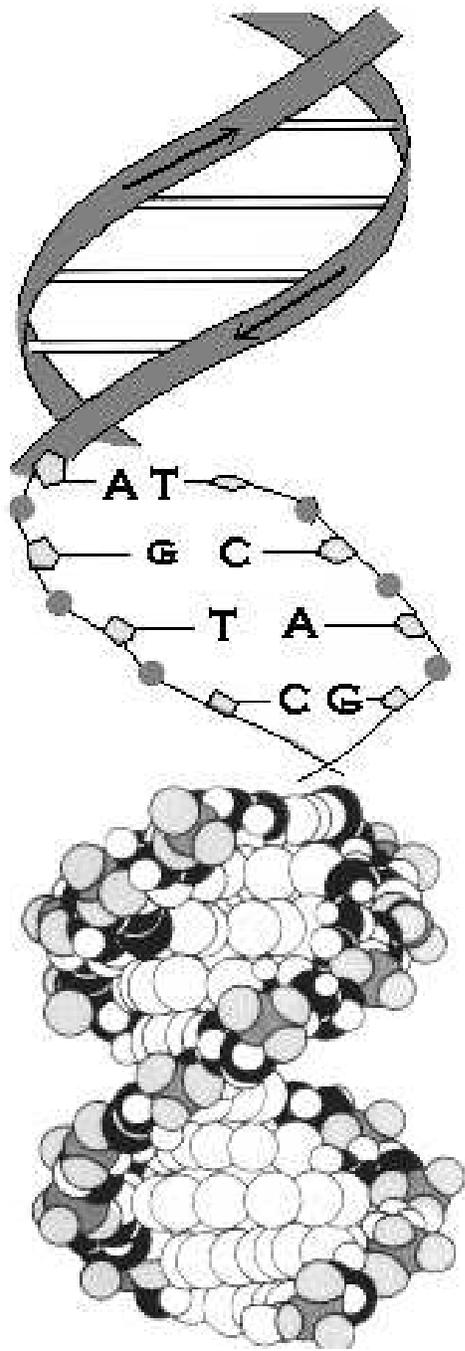


Figure 01

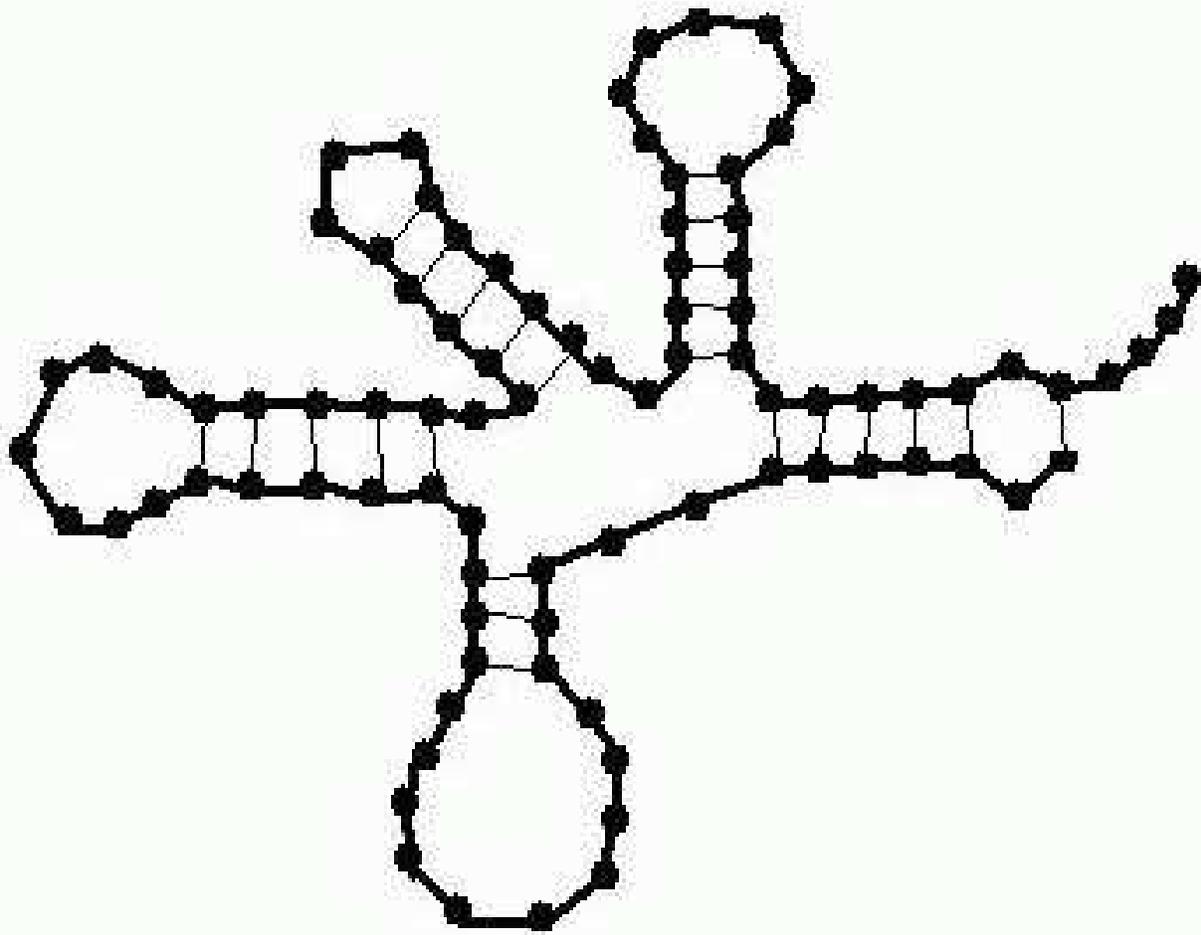


Figure 02

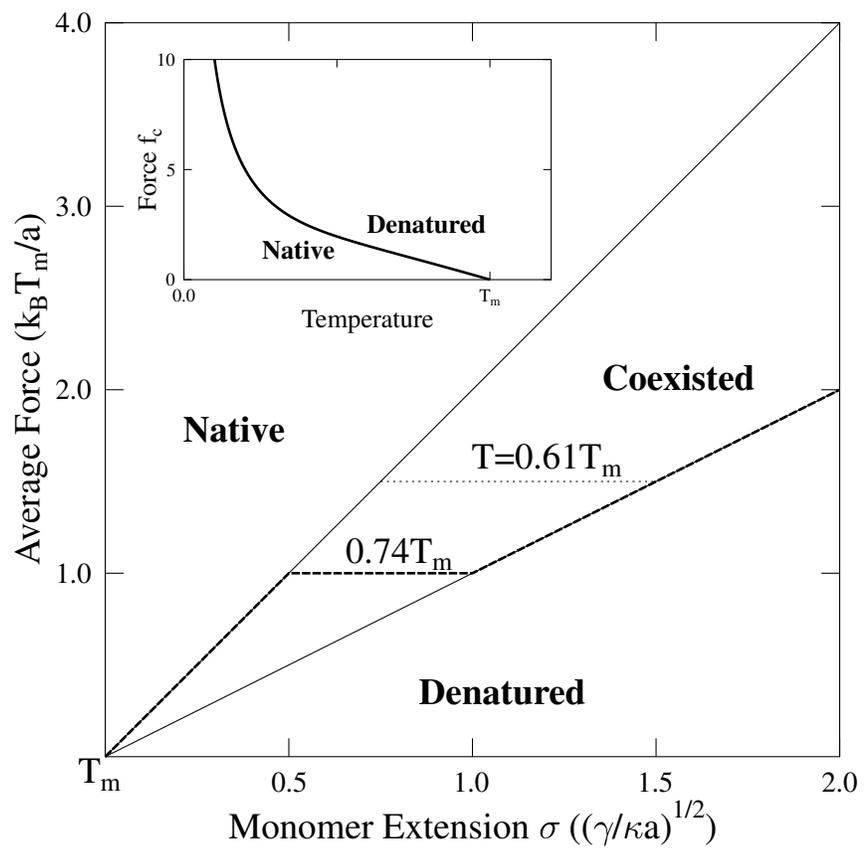


Figure 03

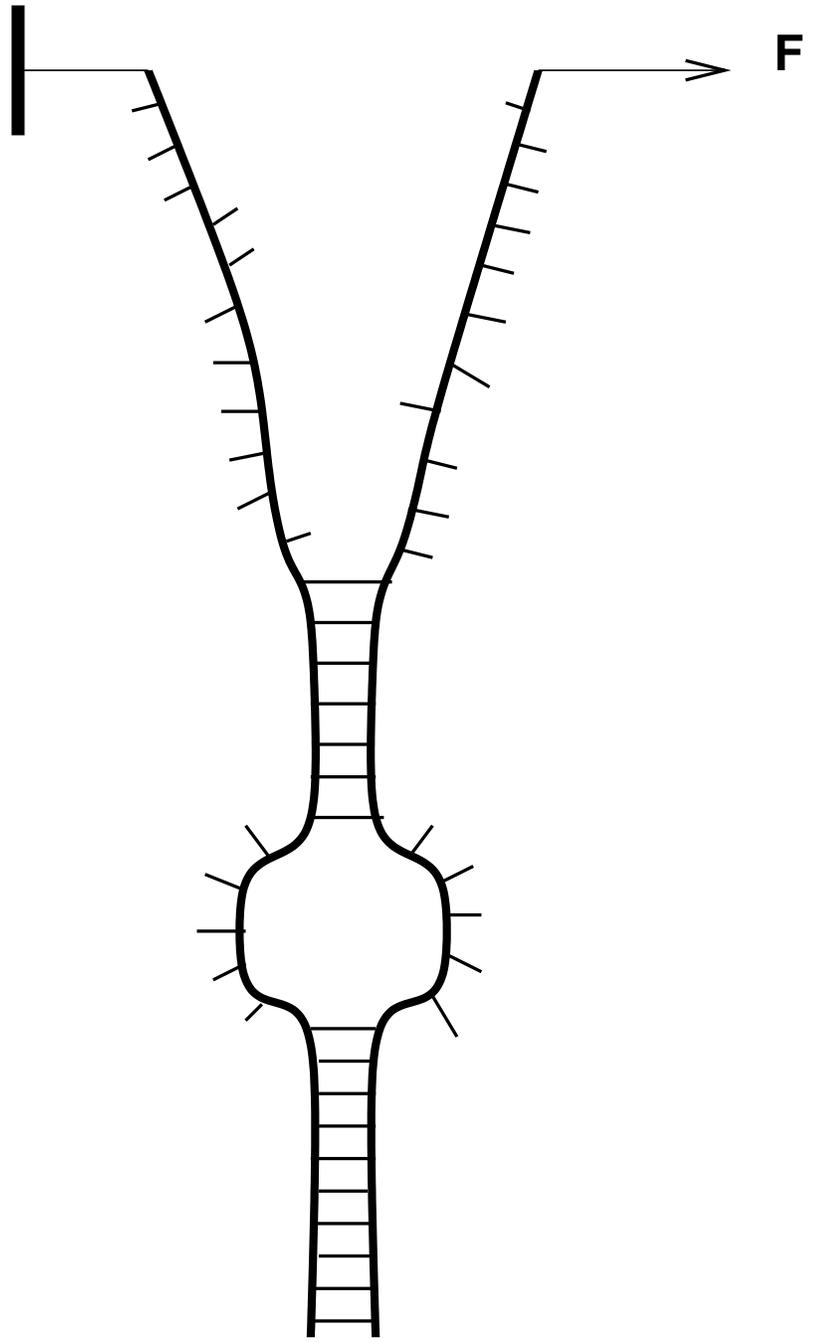
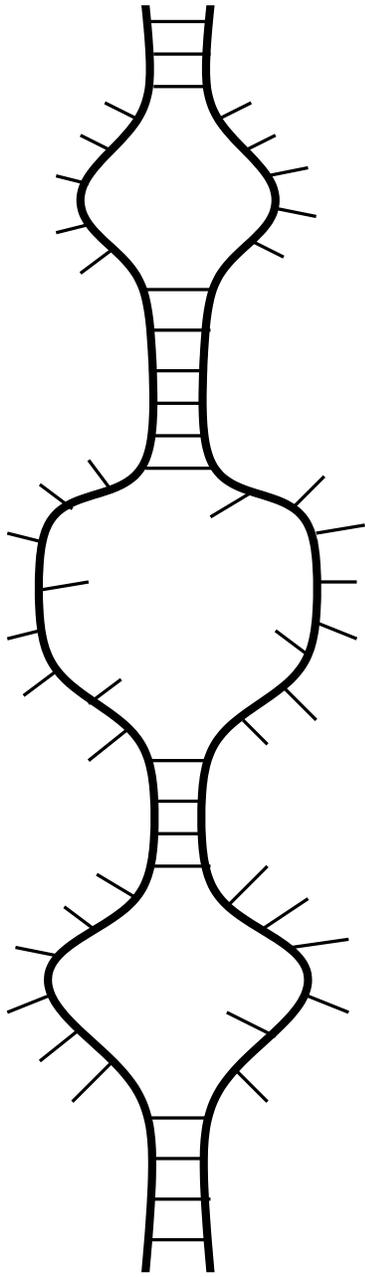


Figure 04

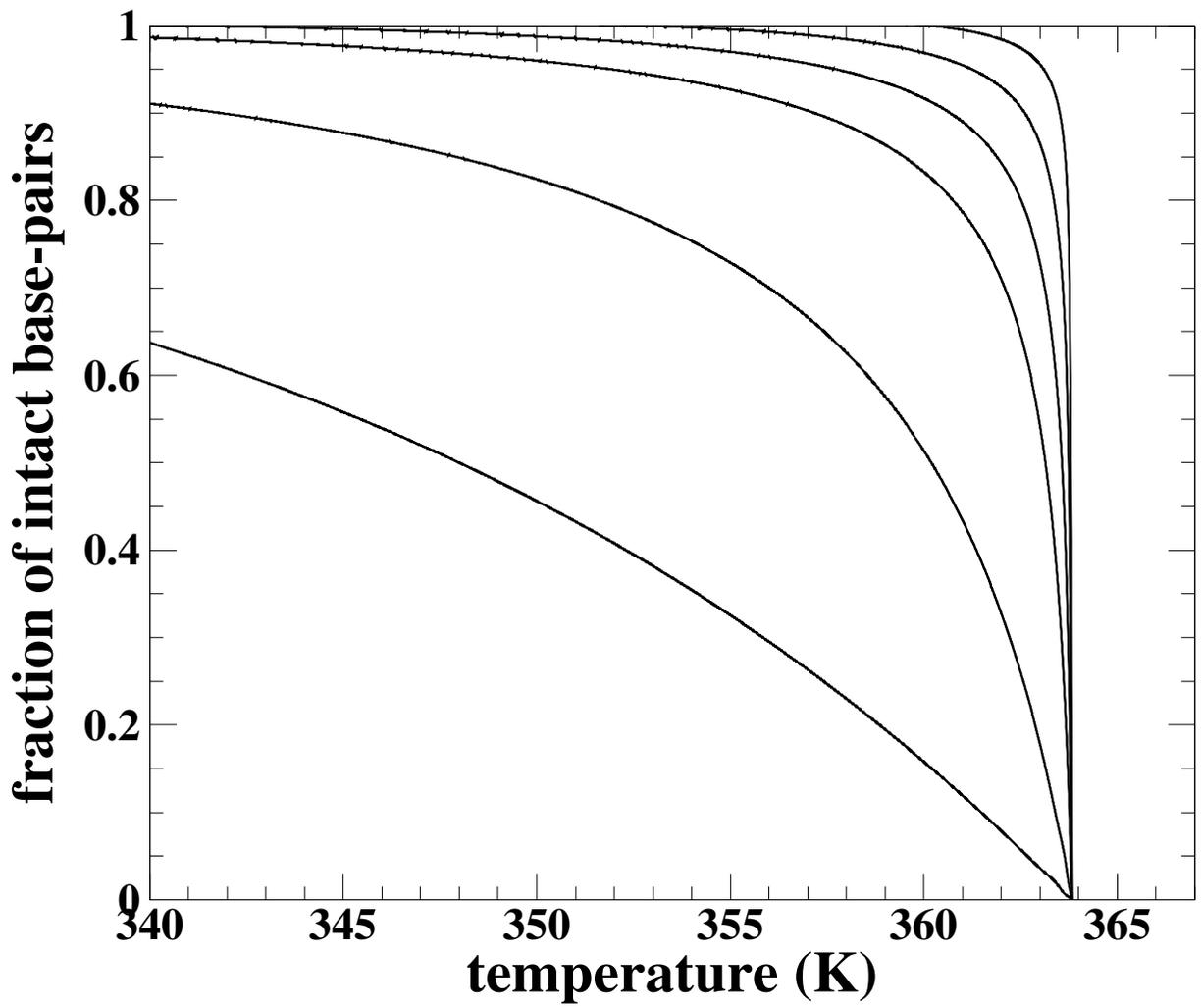


Figure 05

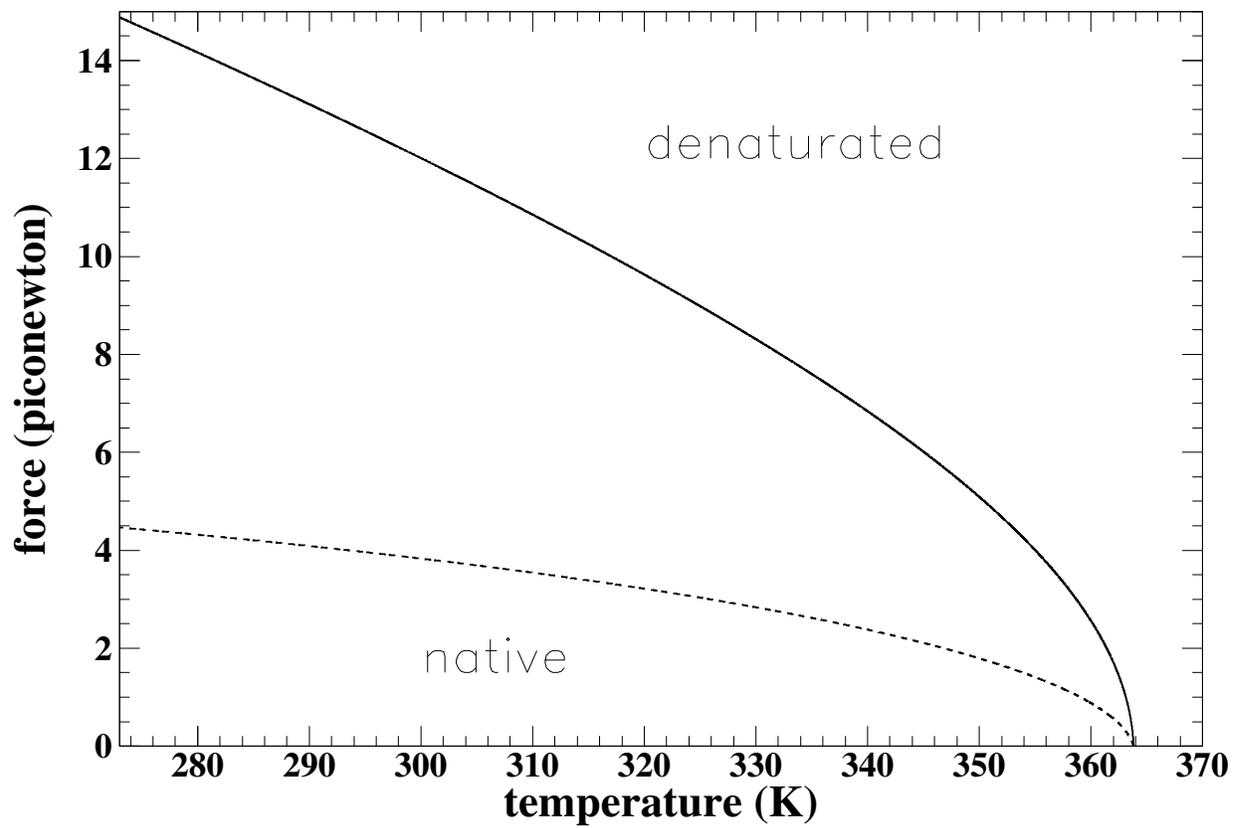


Figure 06

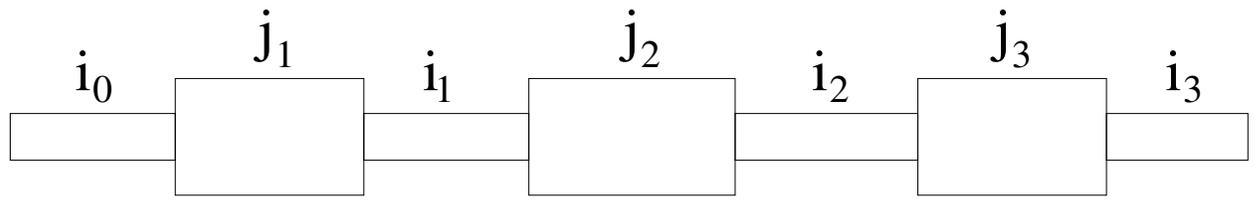


Figure 07

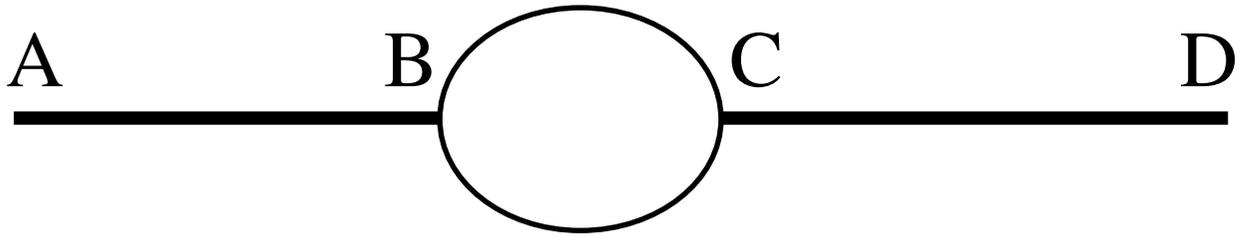


Figure 08

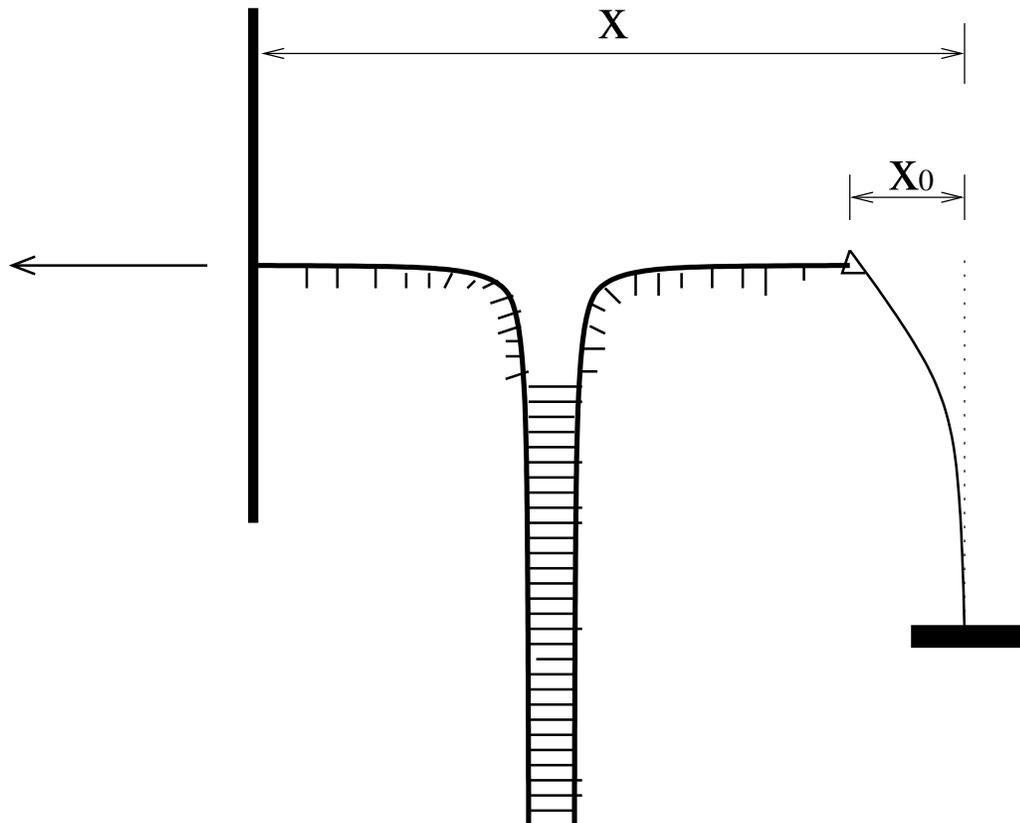


Figure 09

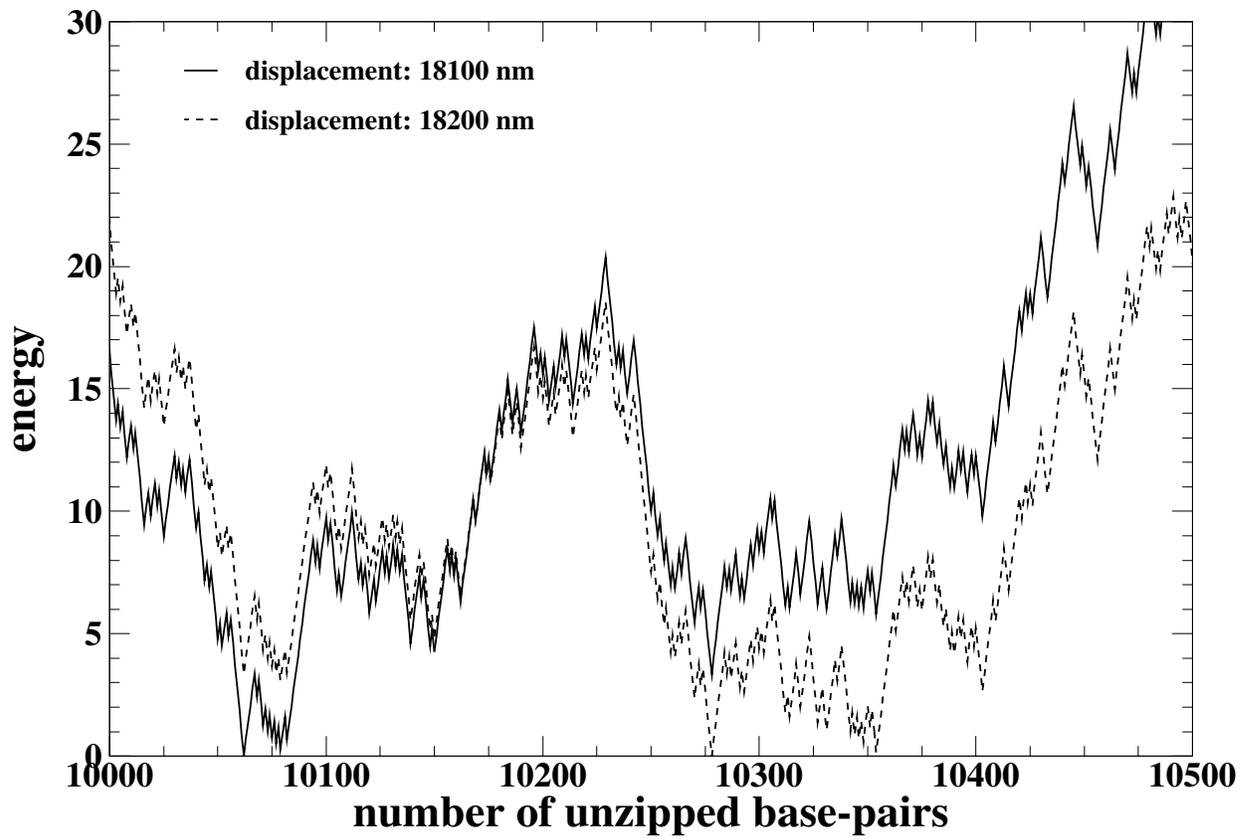


Figure 10

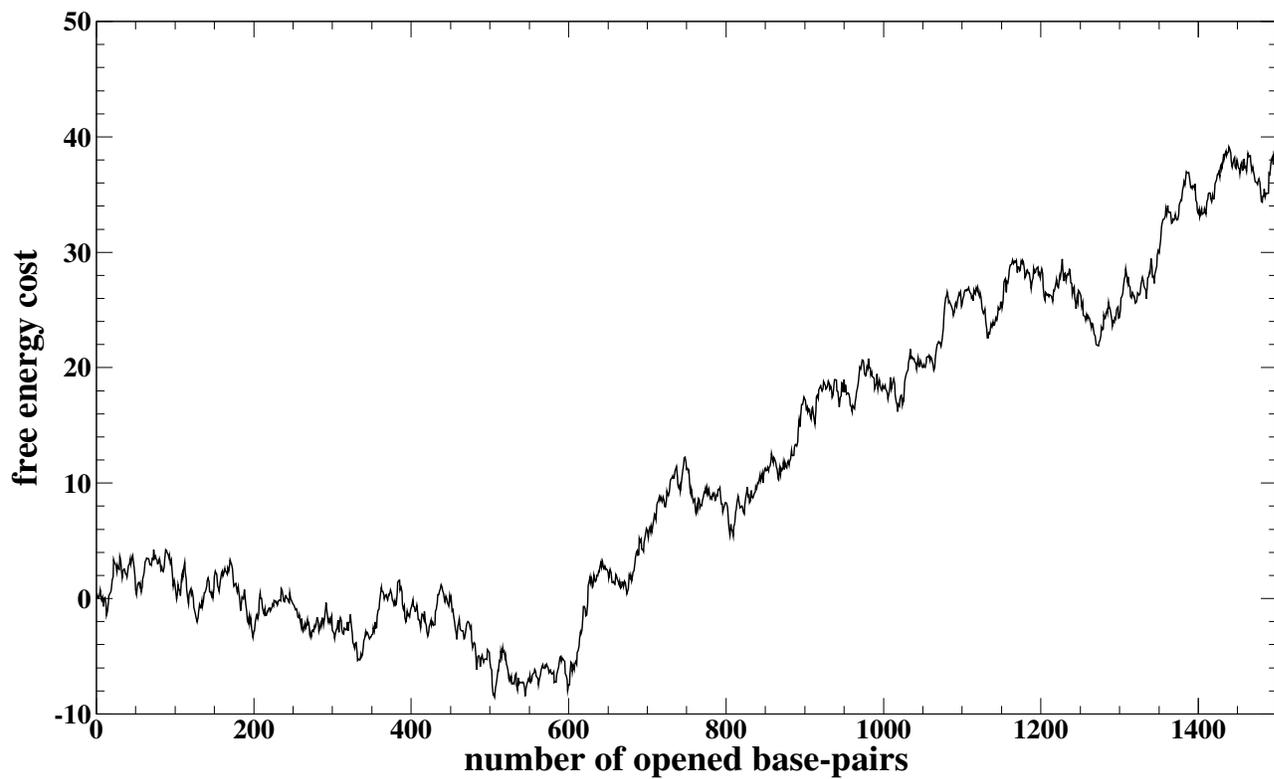


Figure 11

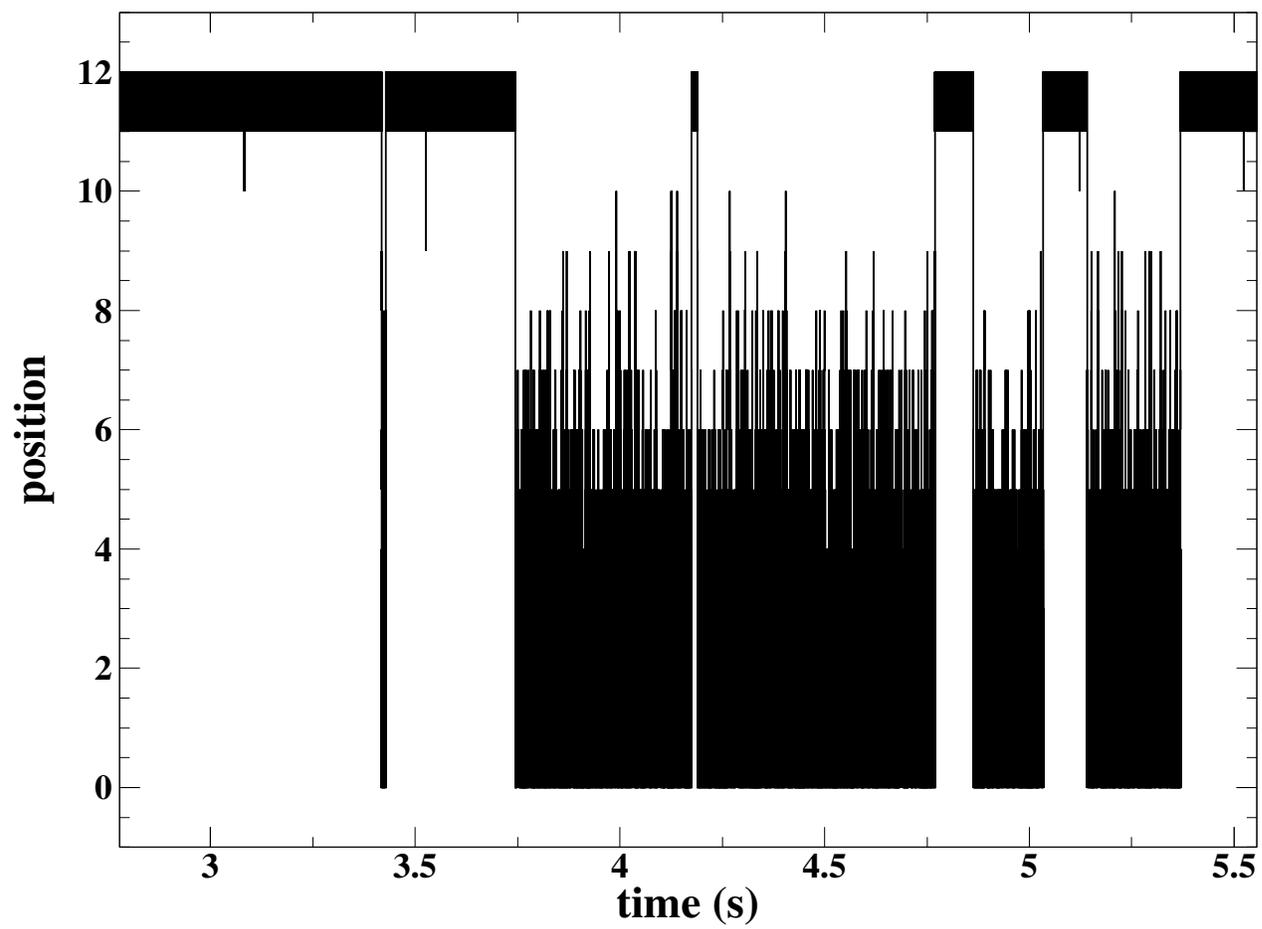


Figure 12

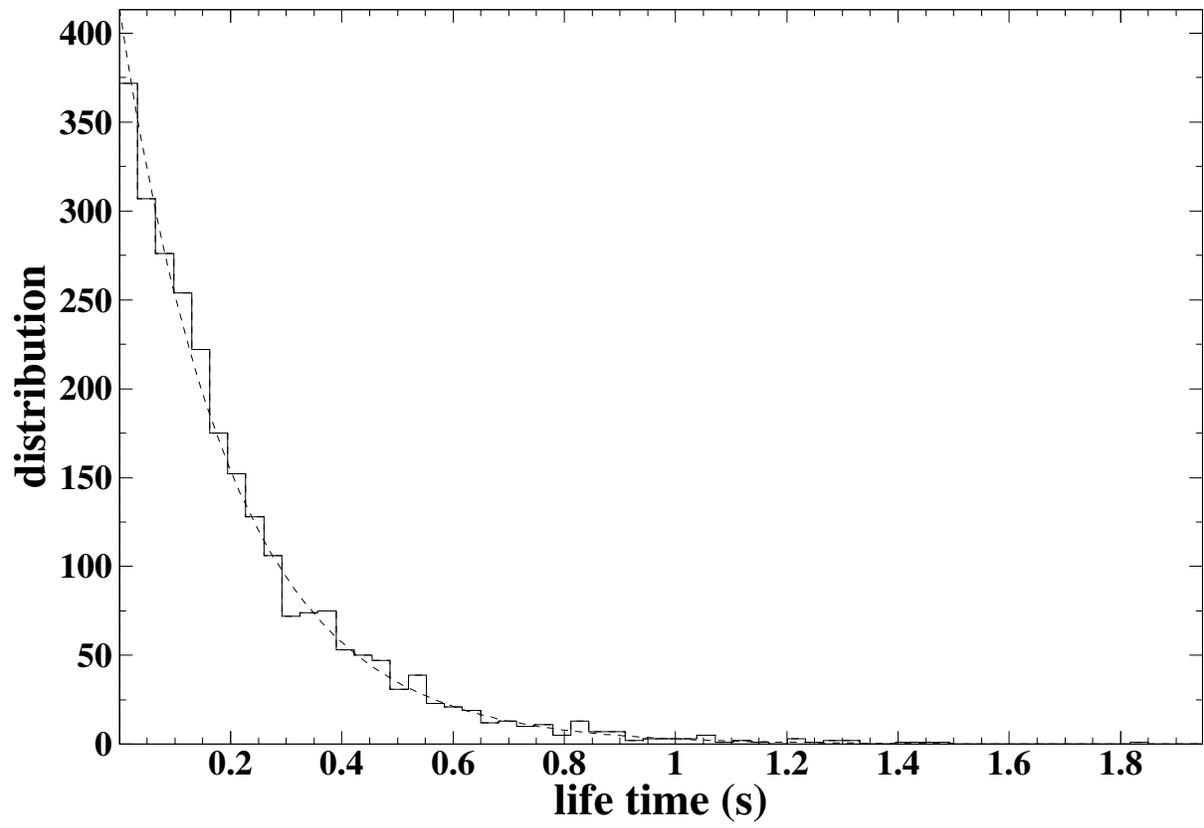


Figure 13

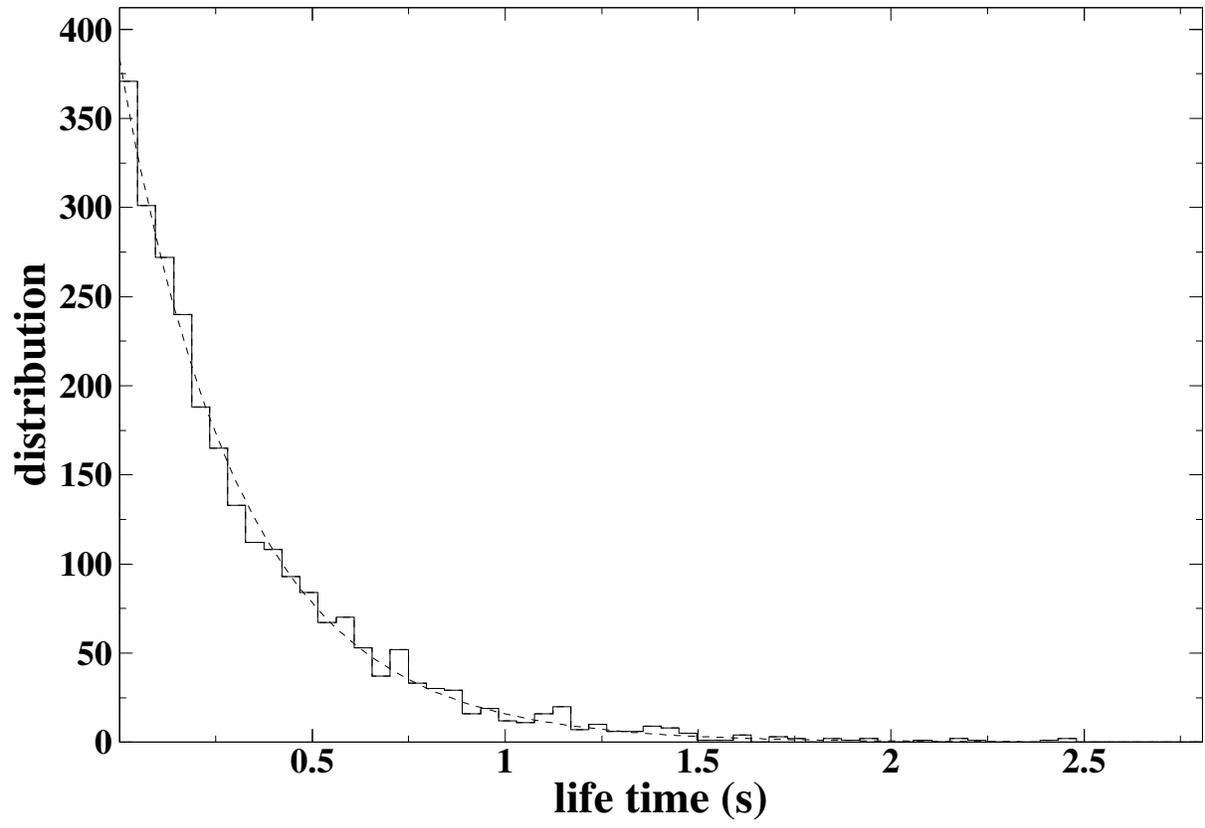


Figure 14

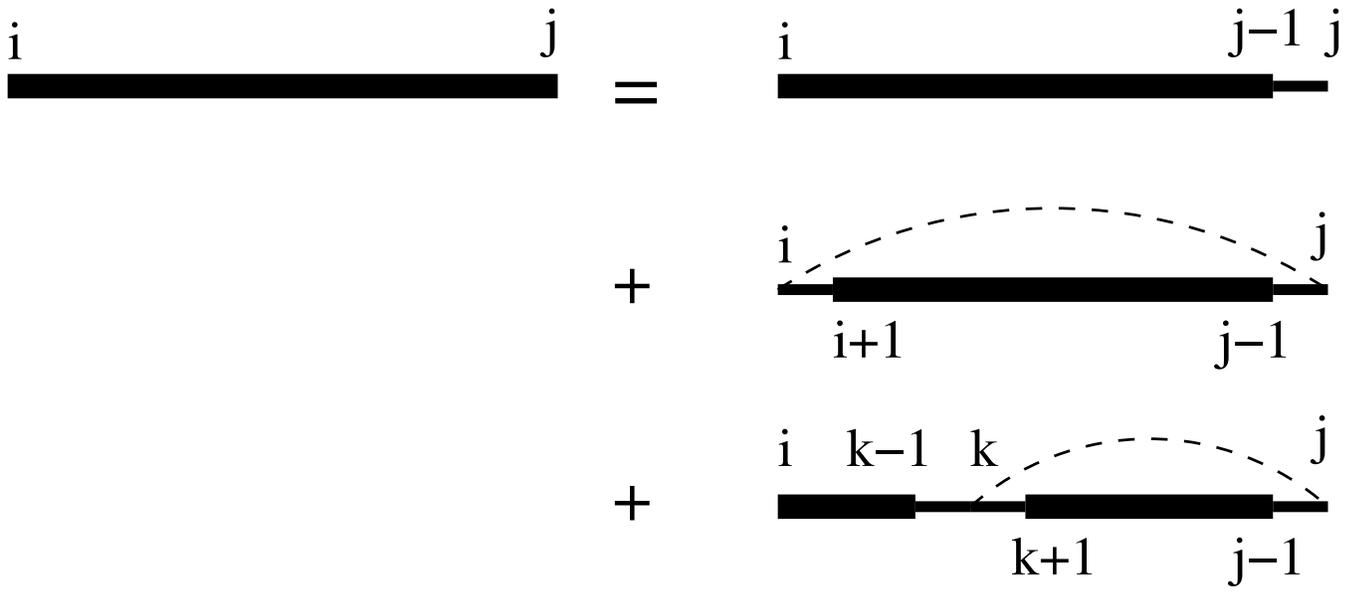


Figure 15

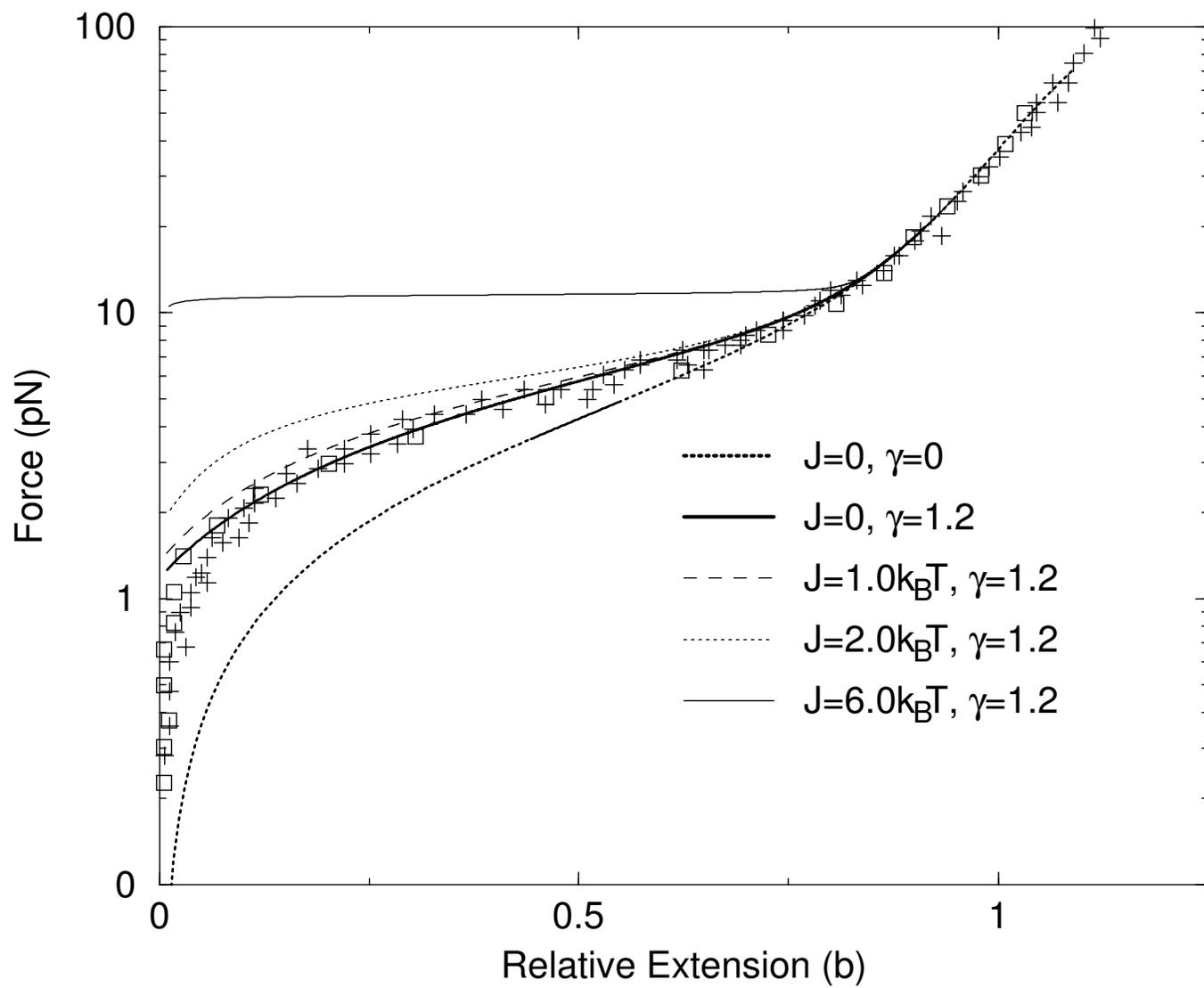


Figure 16

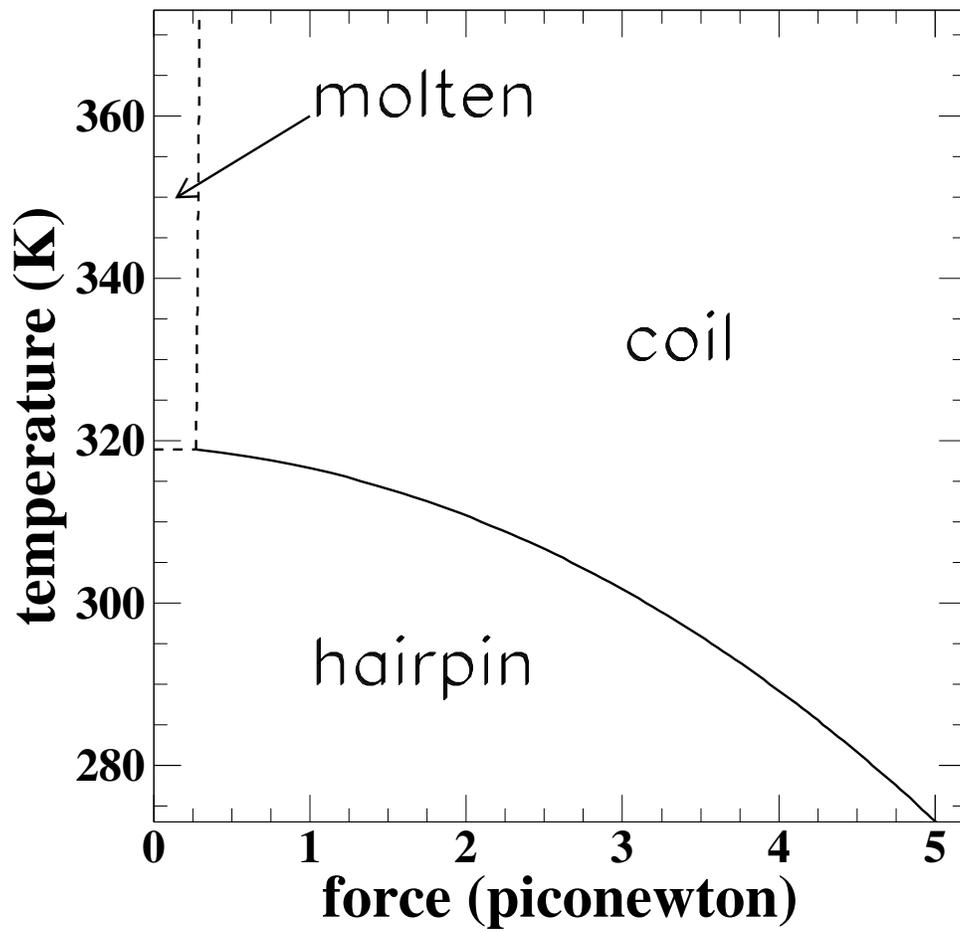
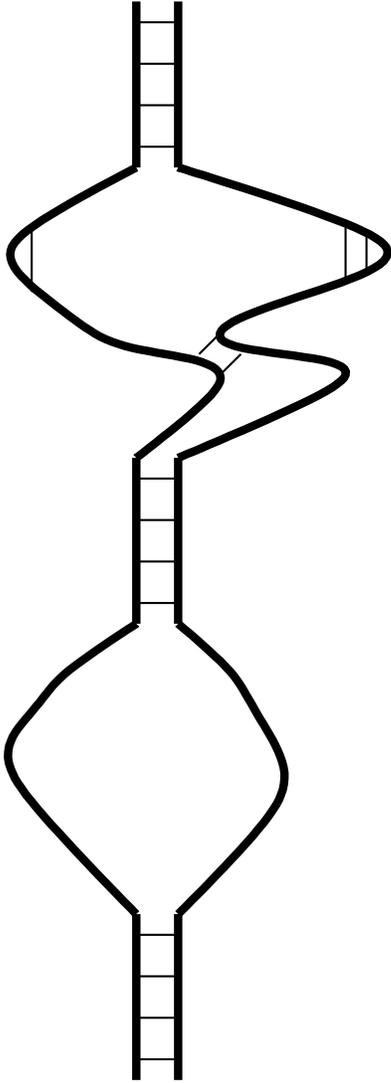


Figure 17

A



B

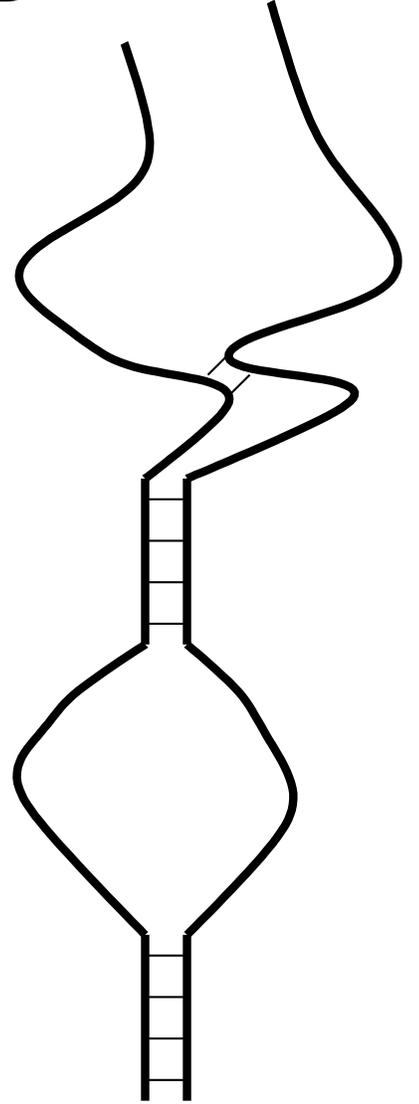


Figure 18

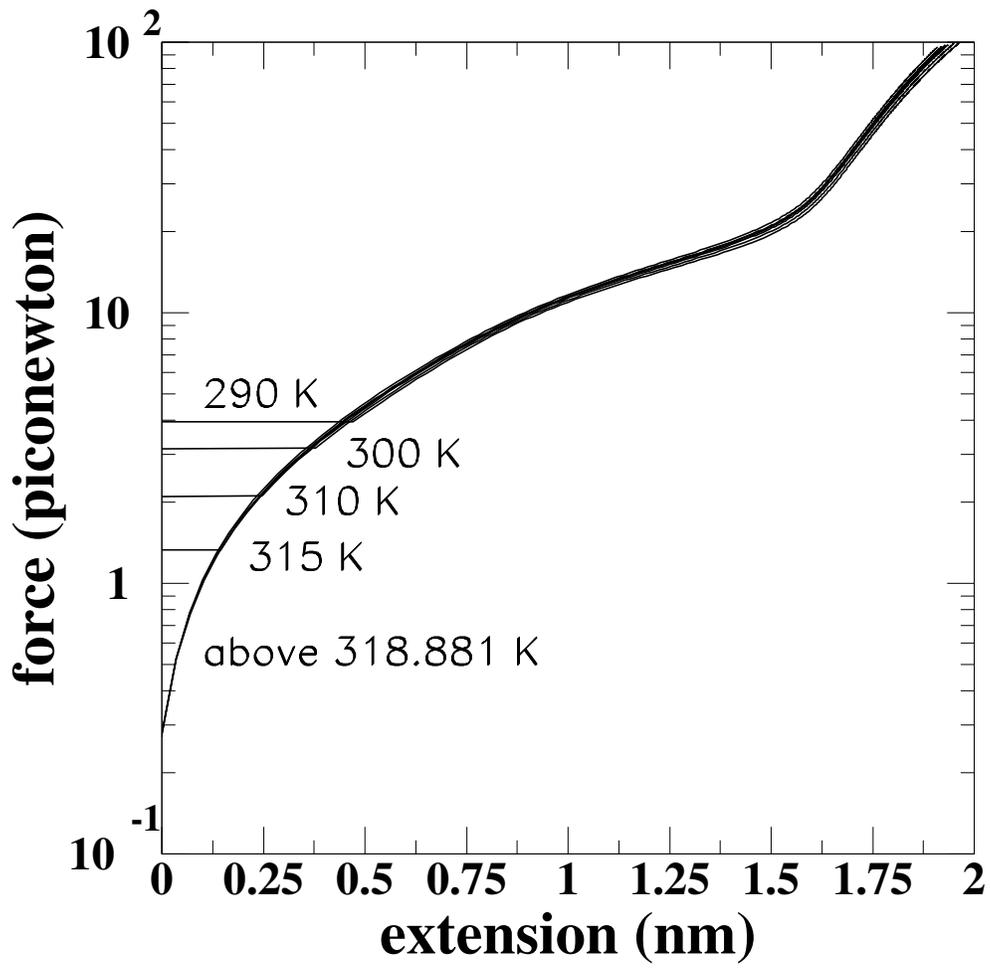


Figure 19

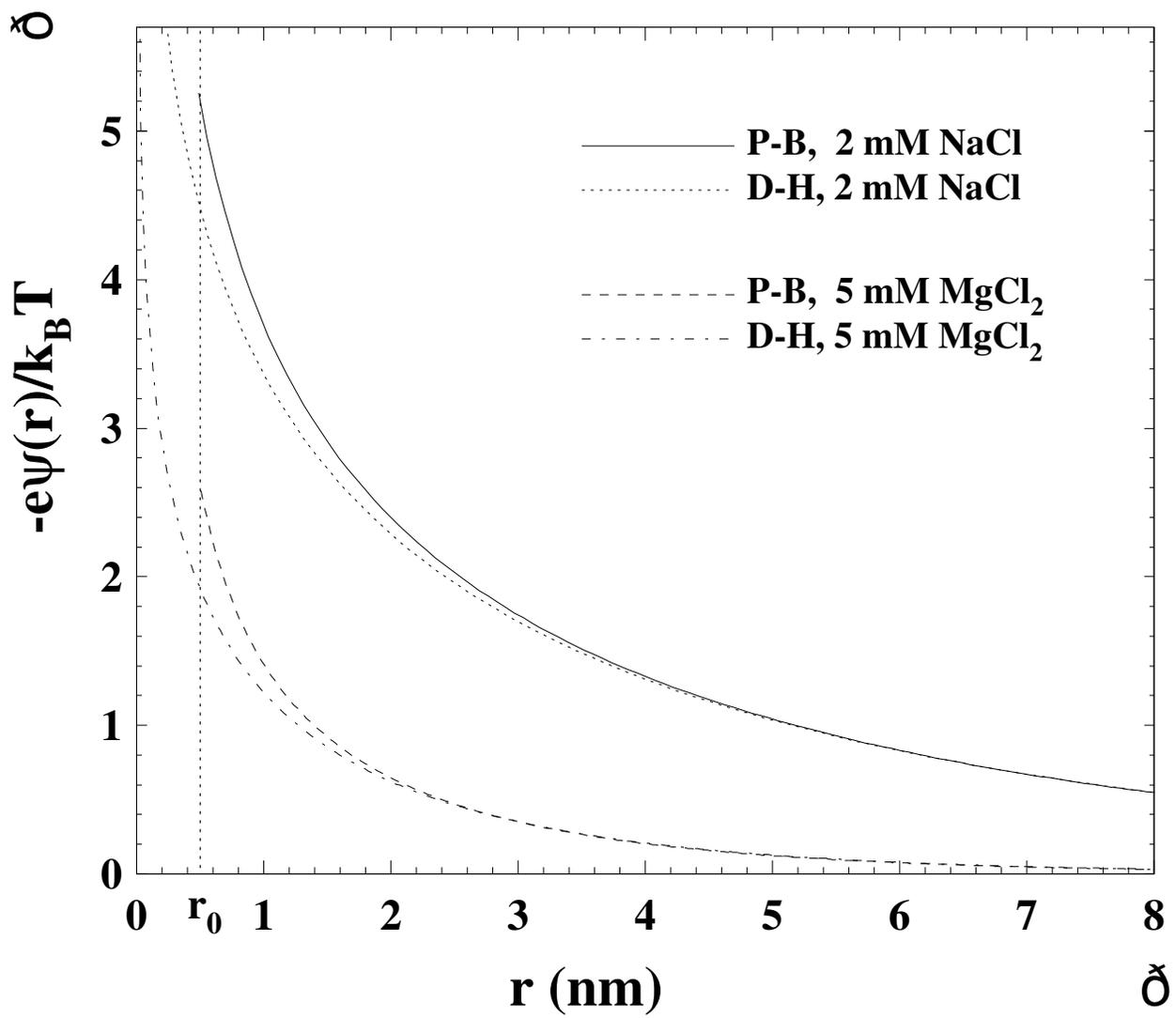


Figure 20

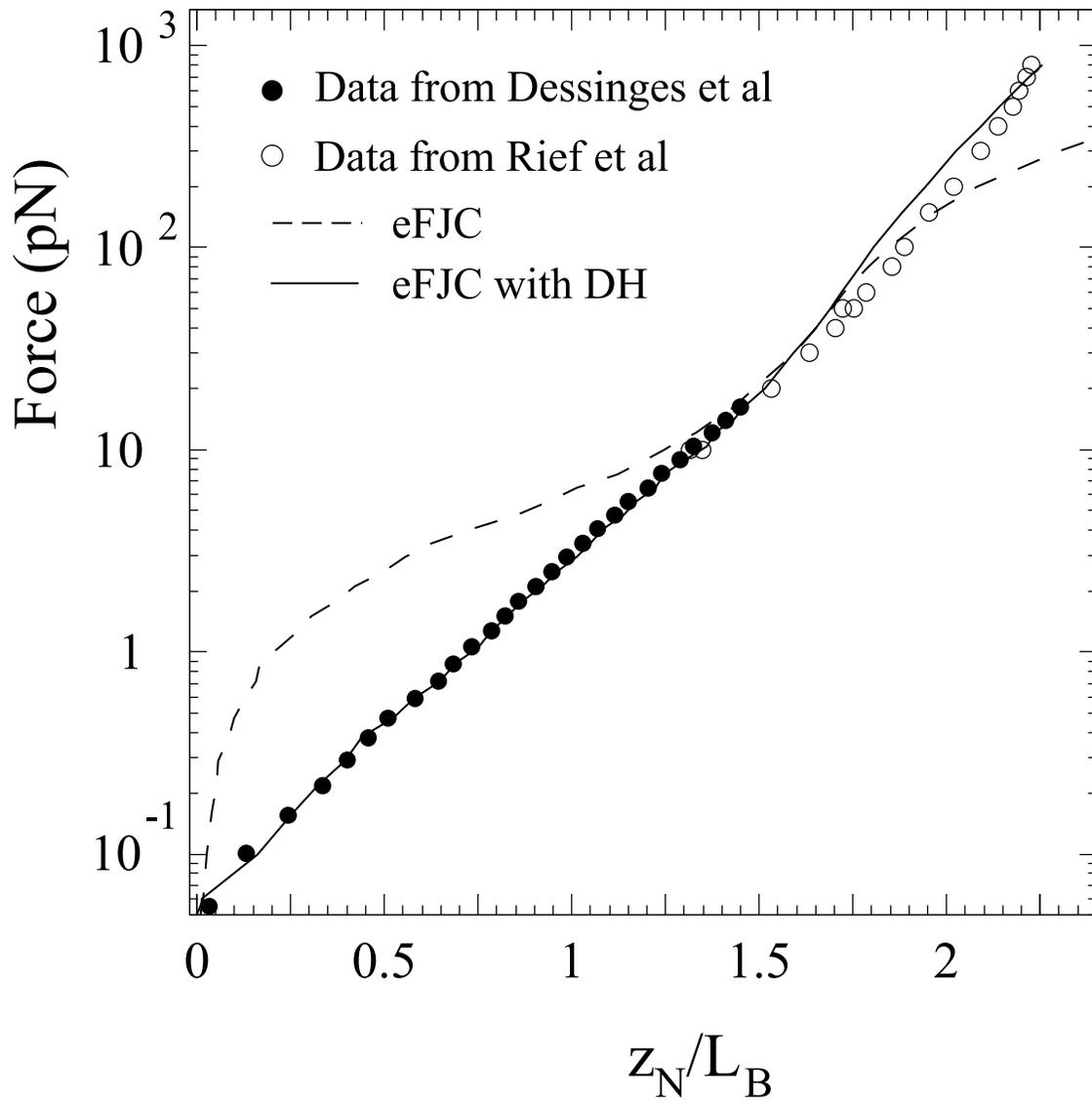


Figure 21

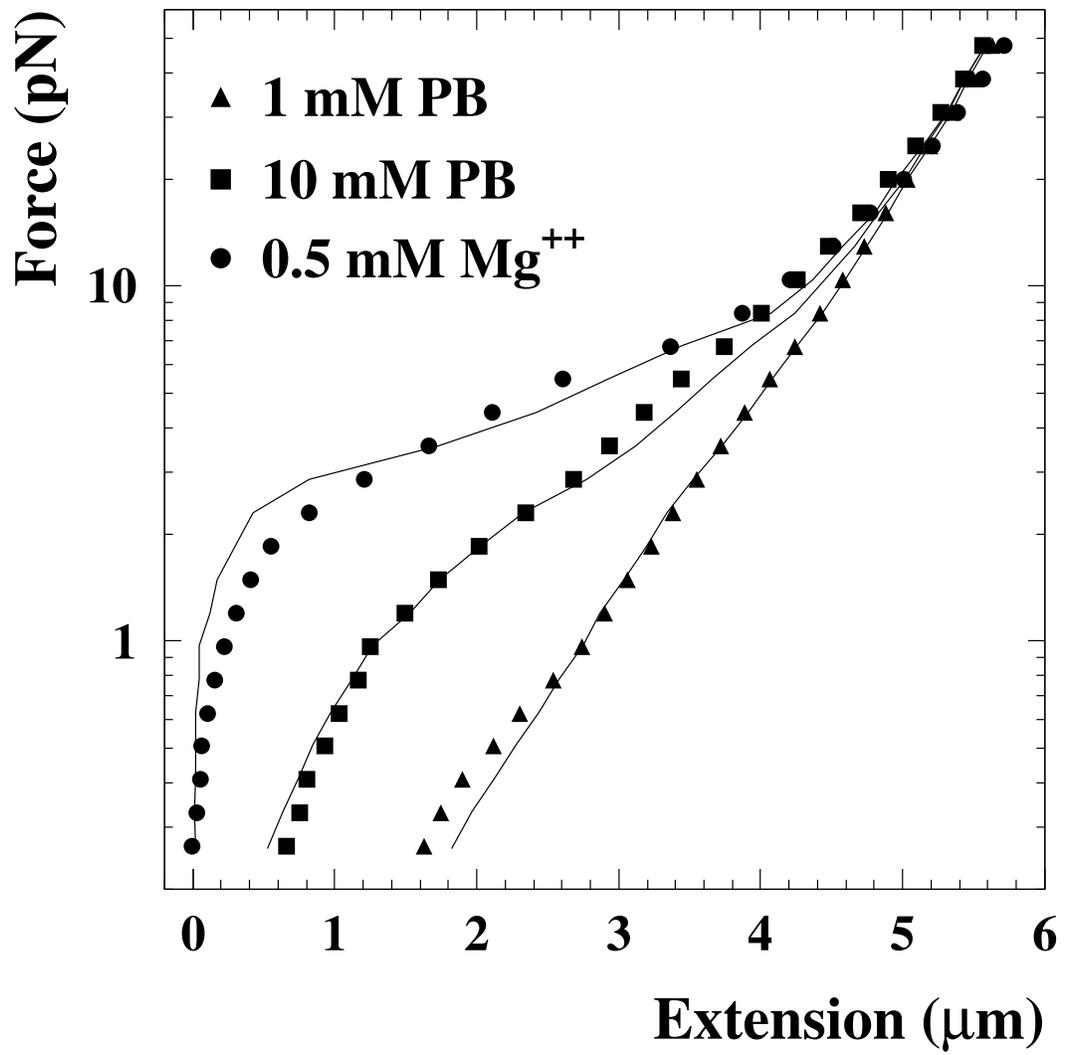


Figure 22

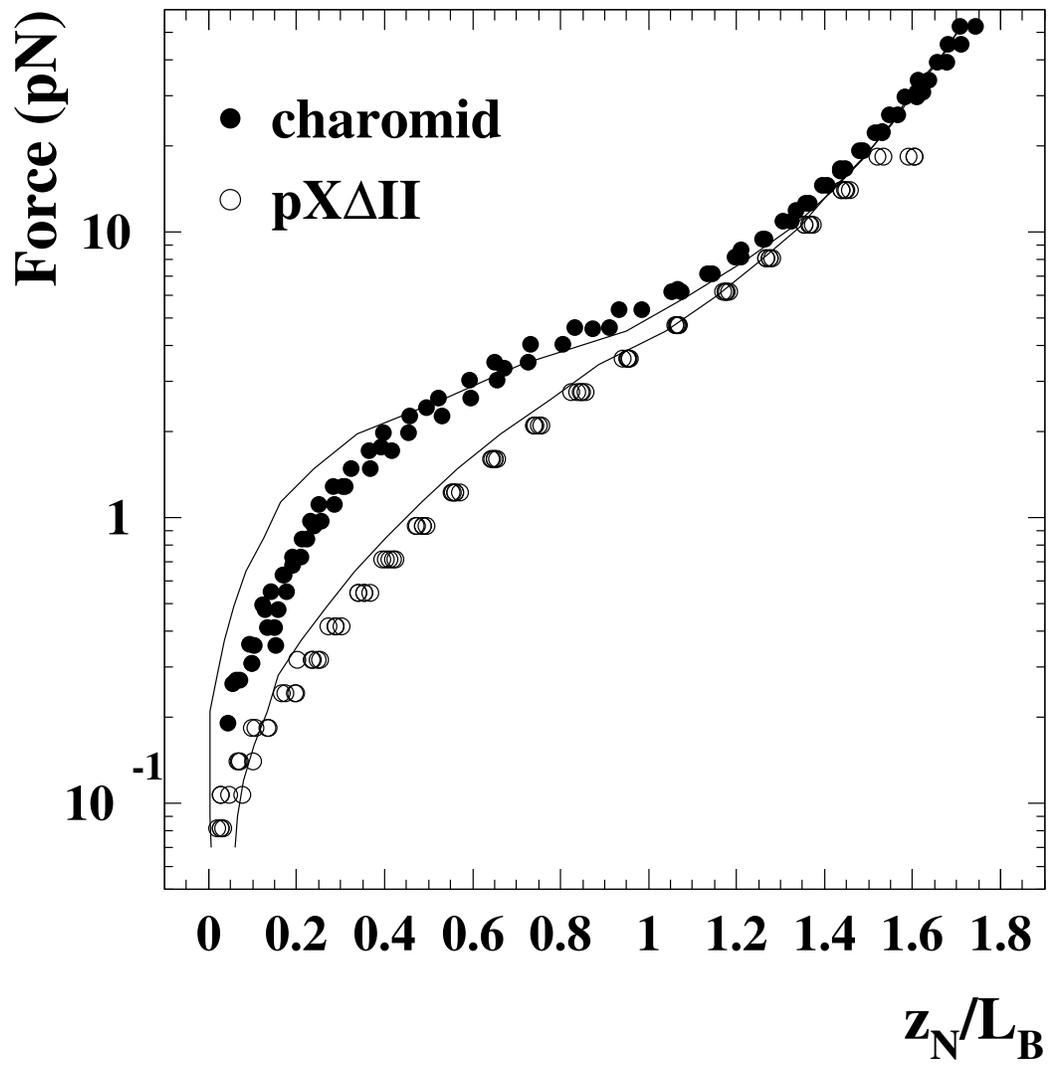


Figure 23

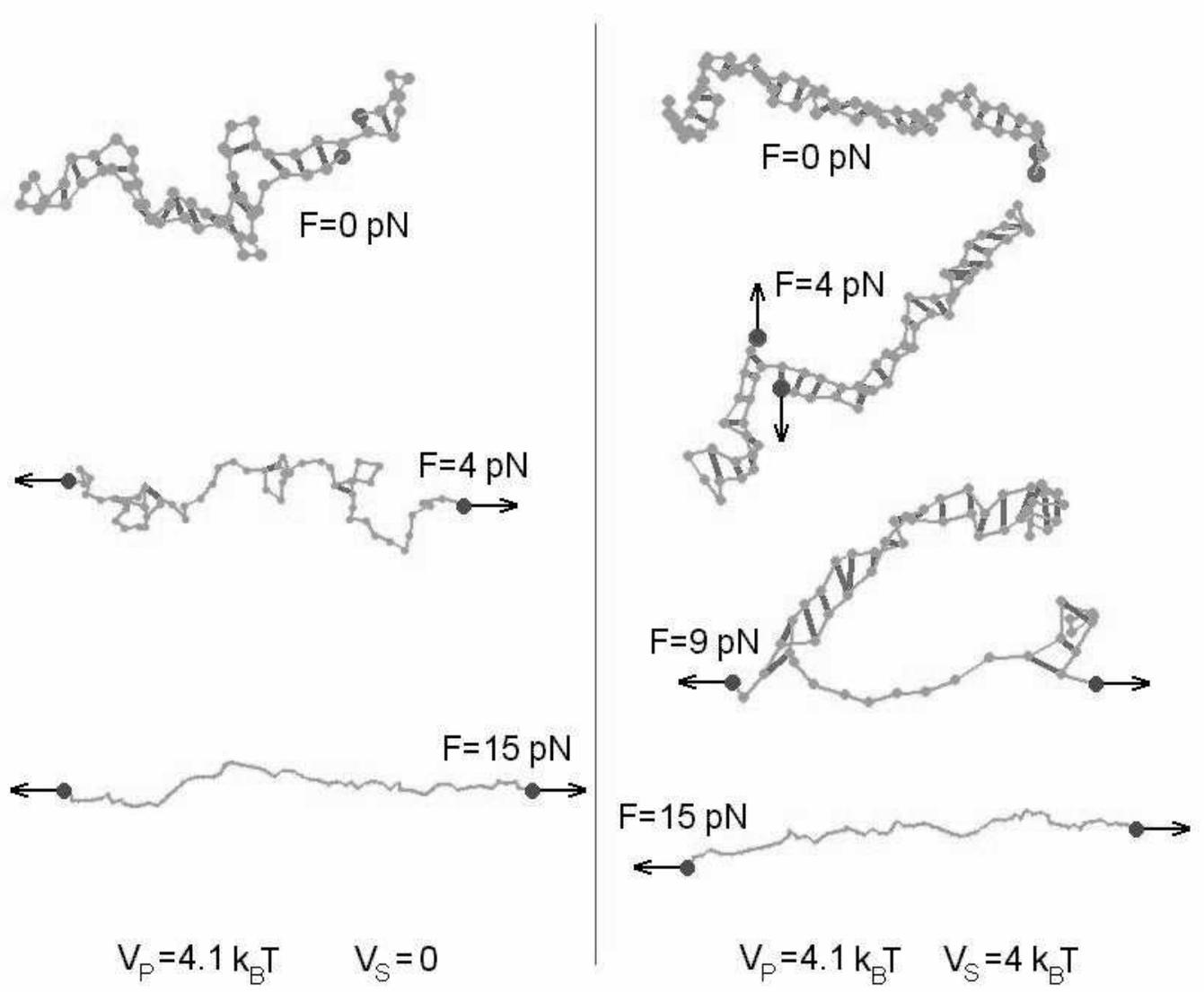


Figure 24

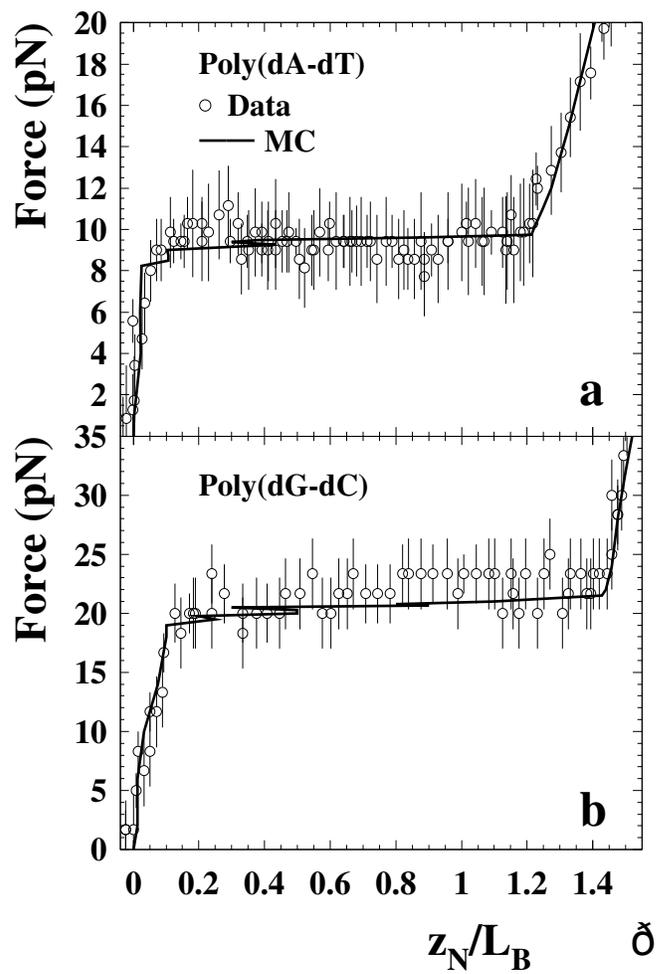


Figure 25

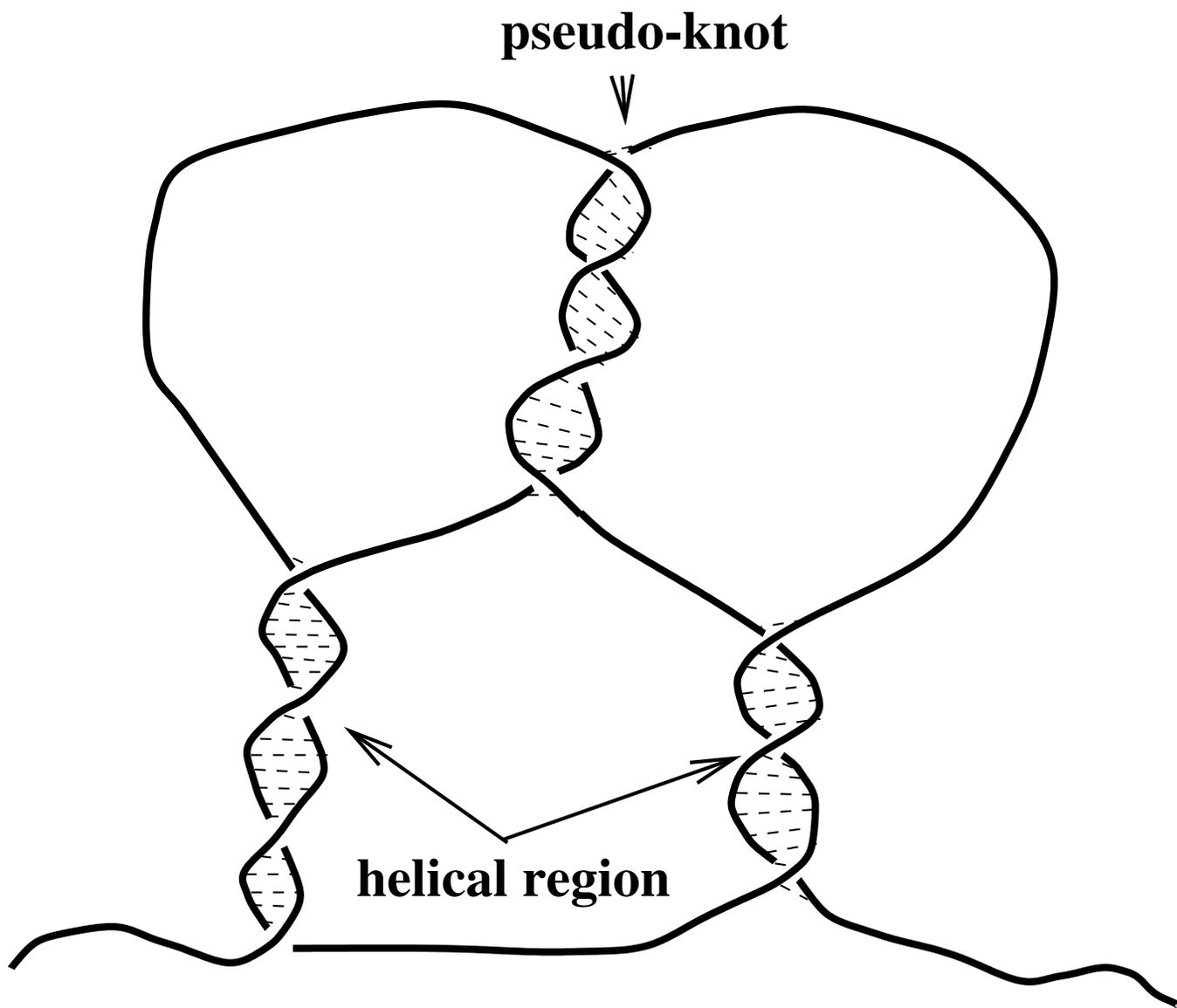


Figure 26

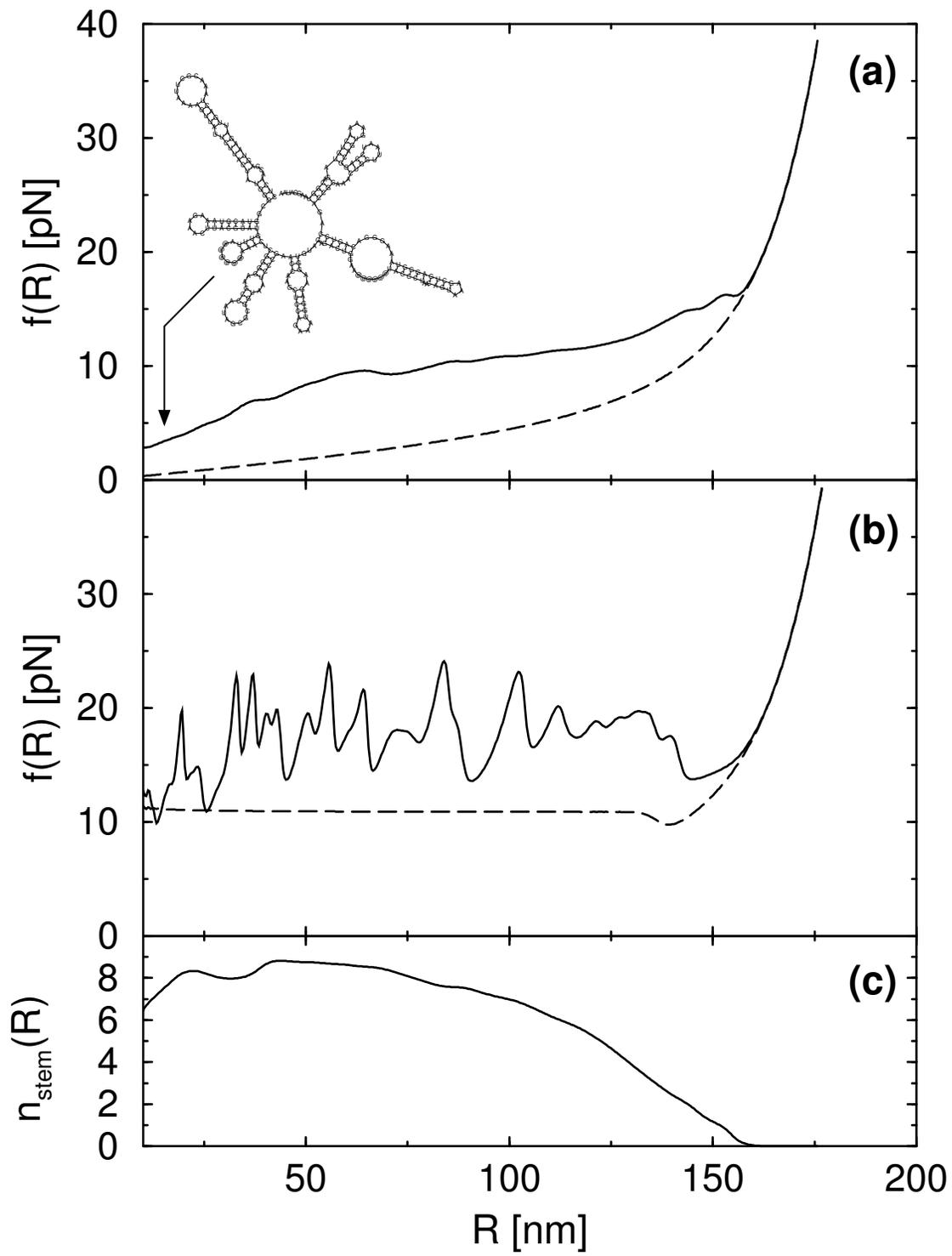
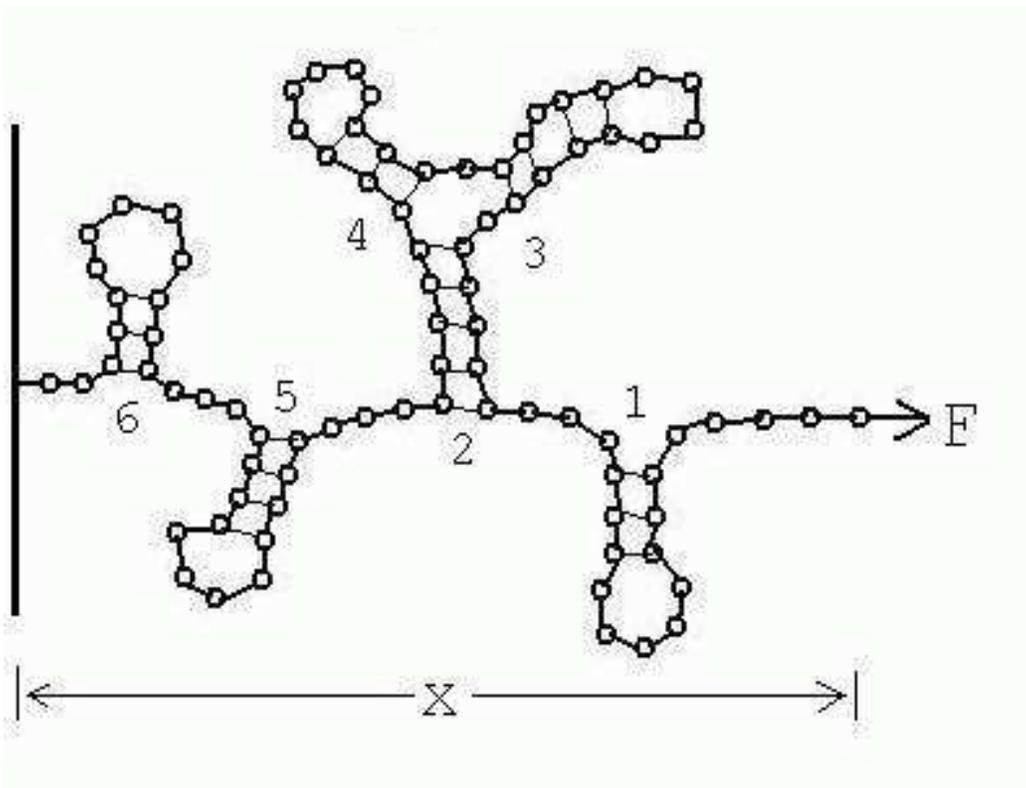


Figure 27

A



B

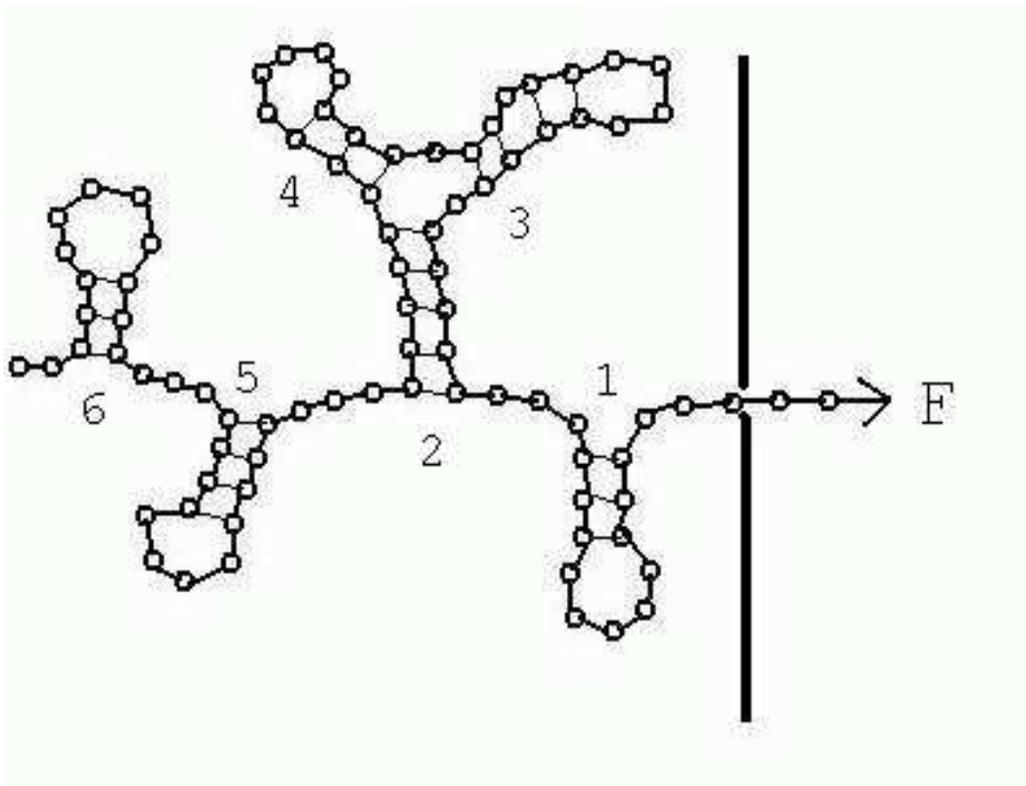


Figure 28

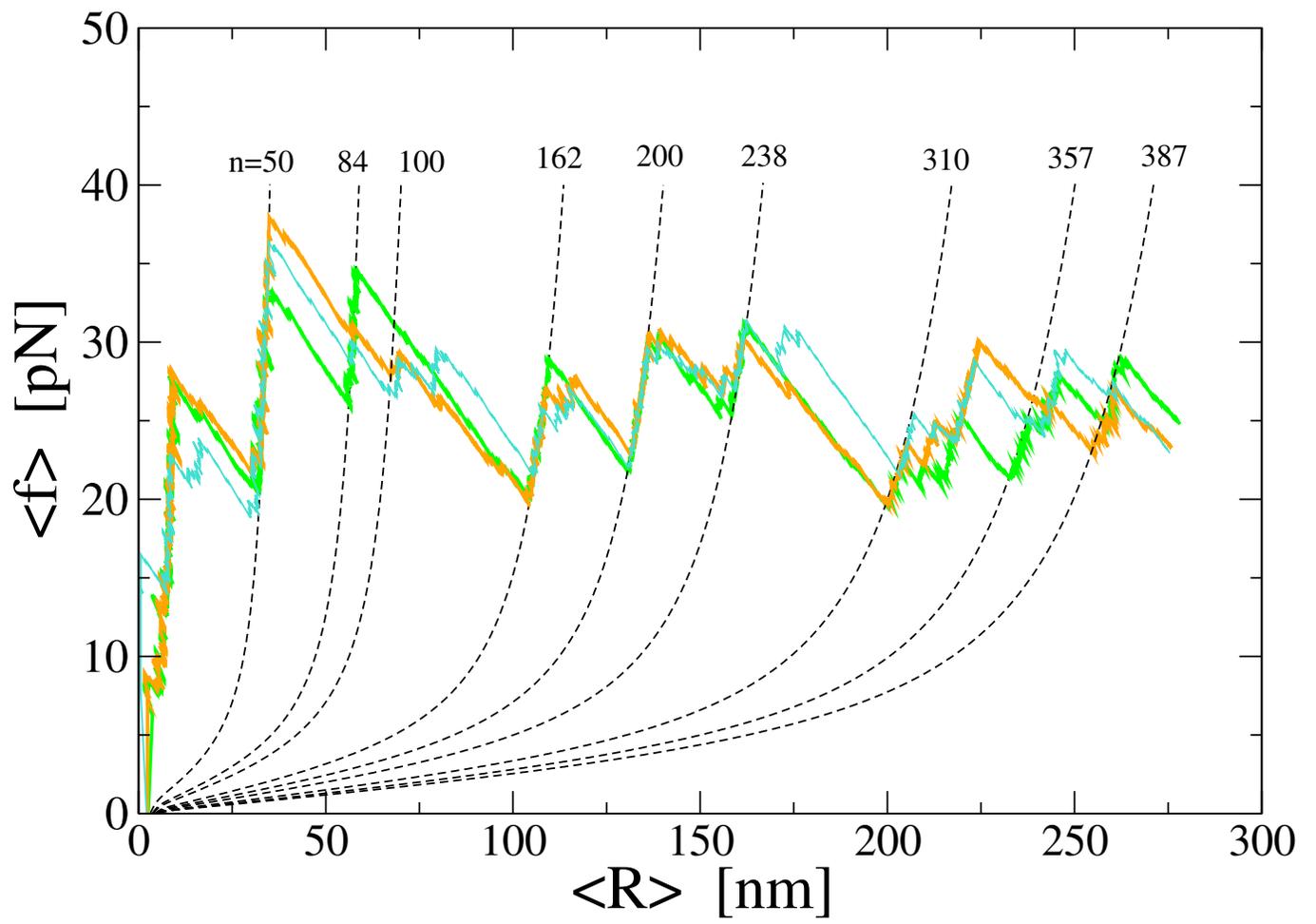


Figure 29

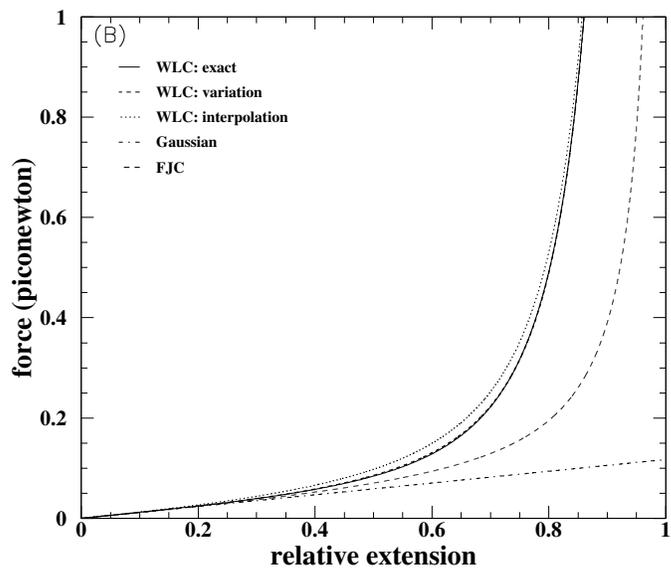
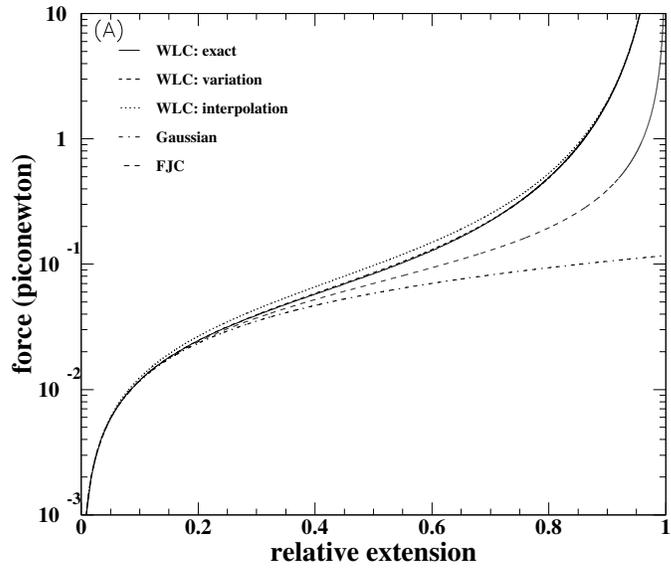


Figure 30

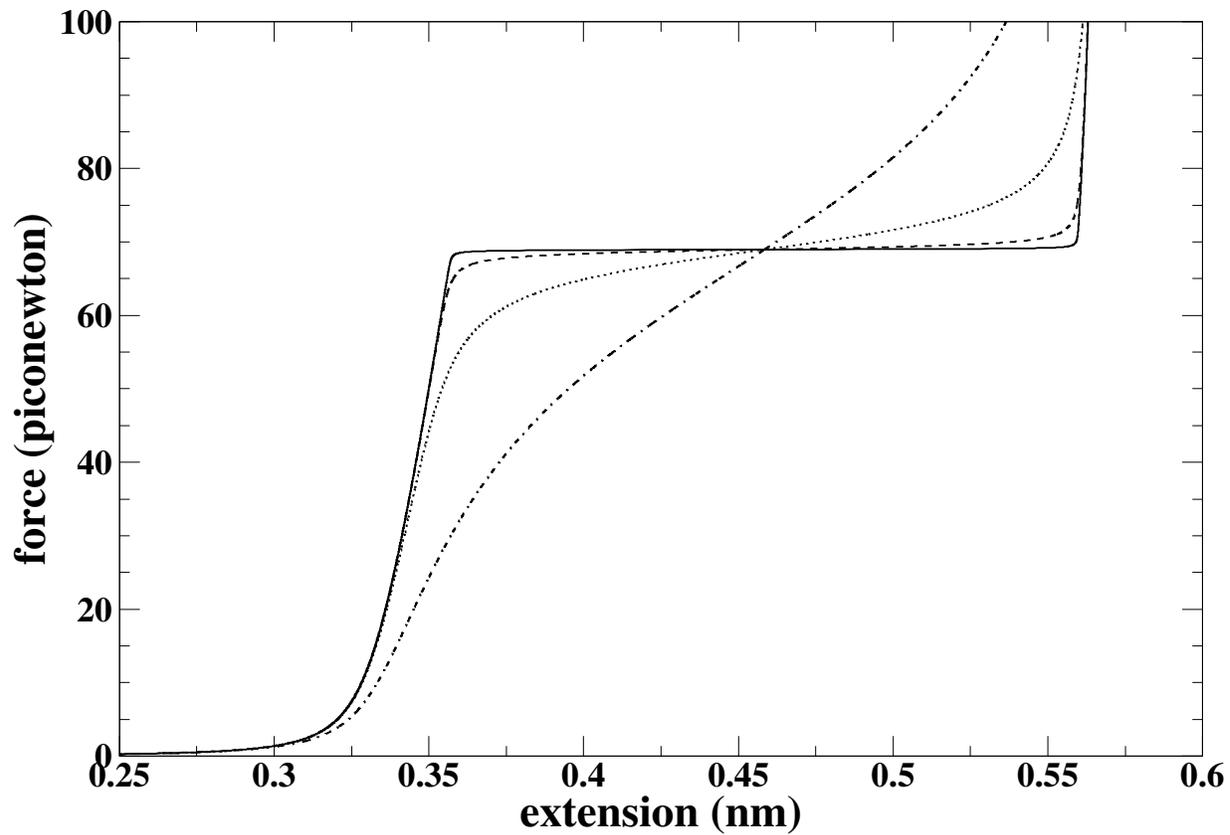
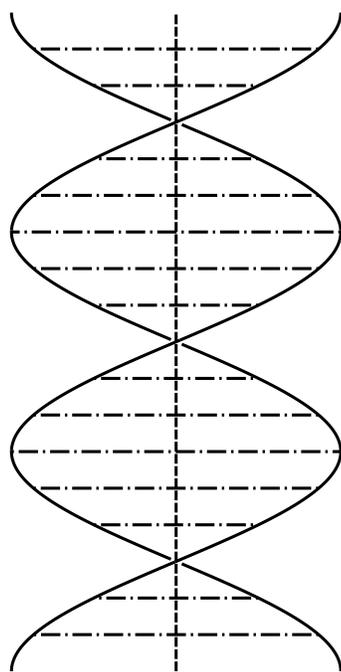


Figure 31



— **Backbones**
- - - **Central axis**
- - - **Basepairs**

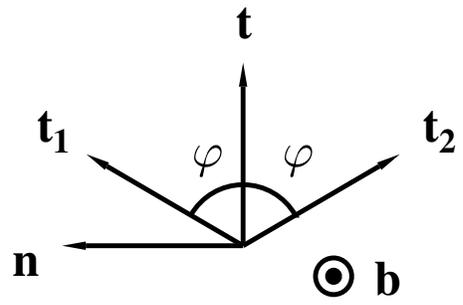


Figure 32

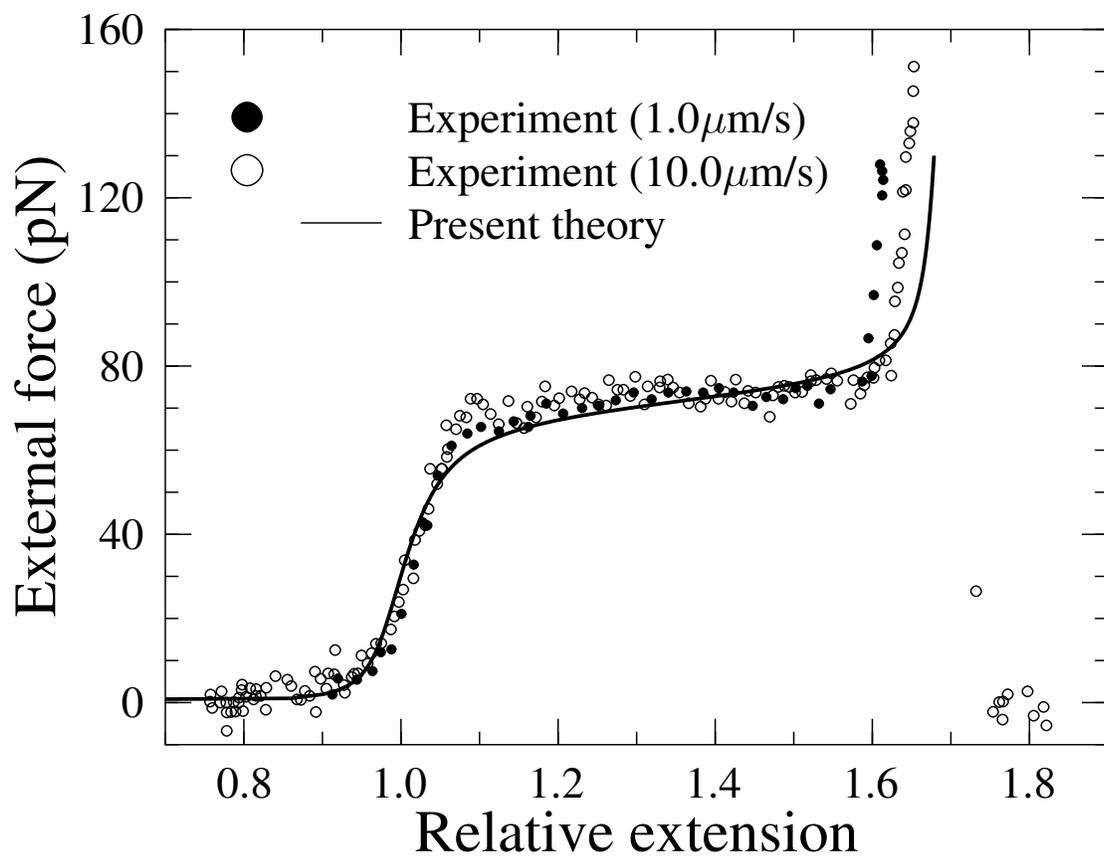


Figure 33

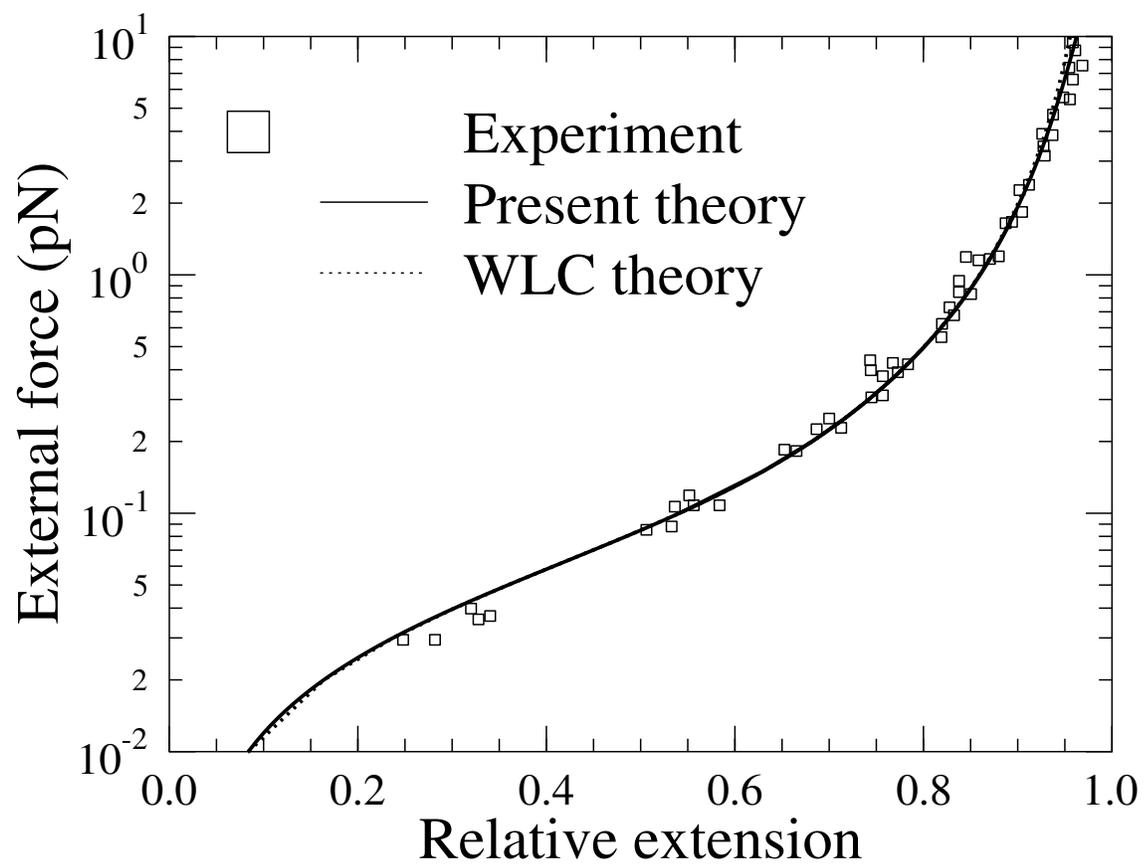


Figure 34

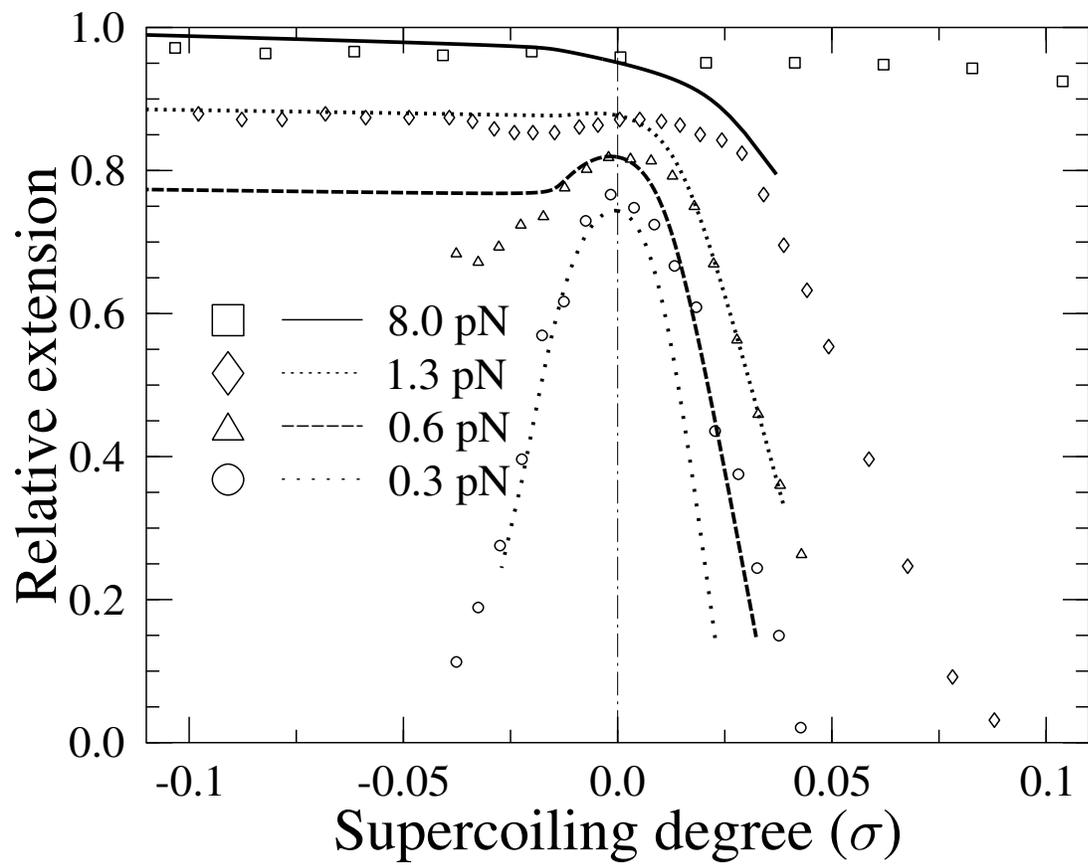


Figure 35

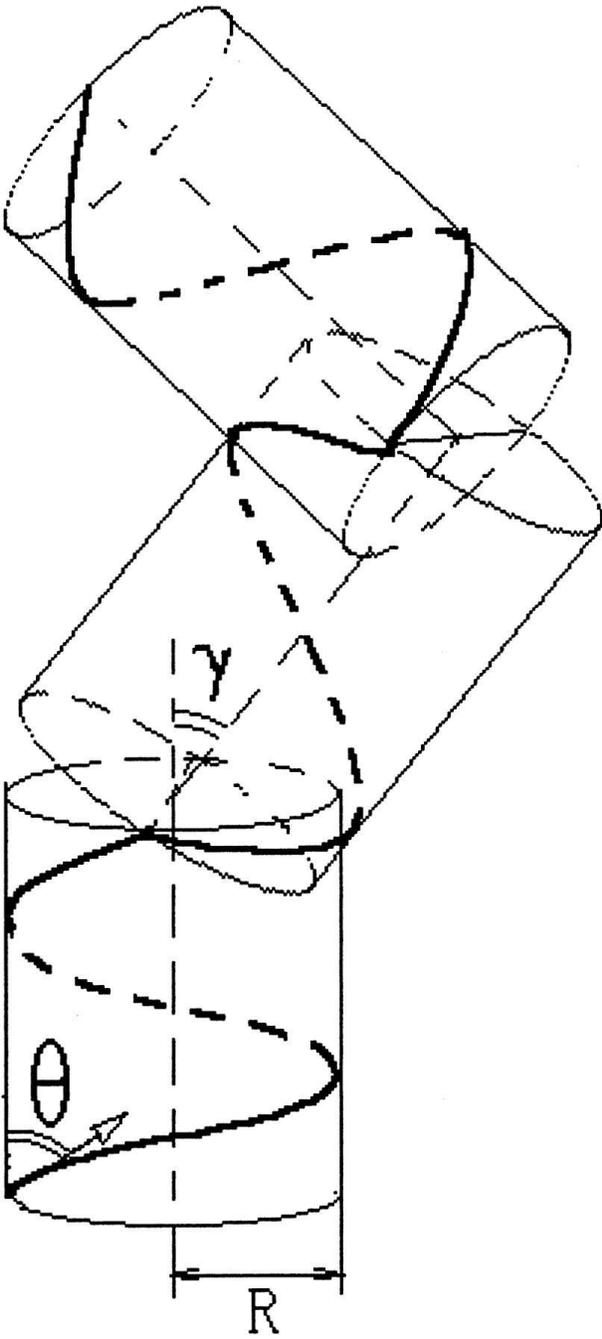


Figure 36

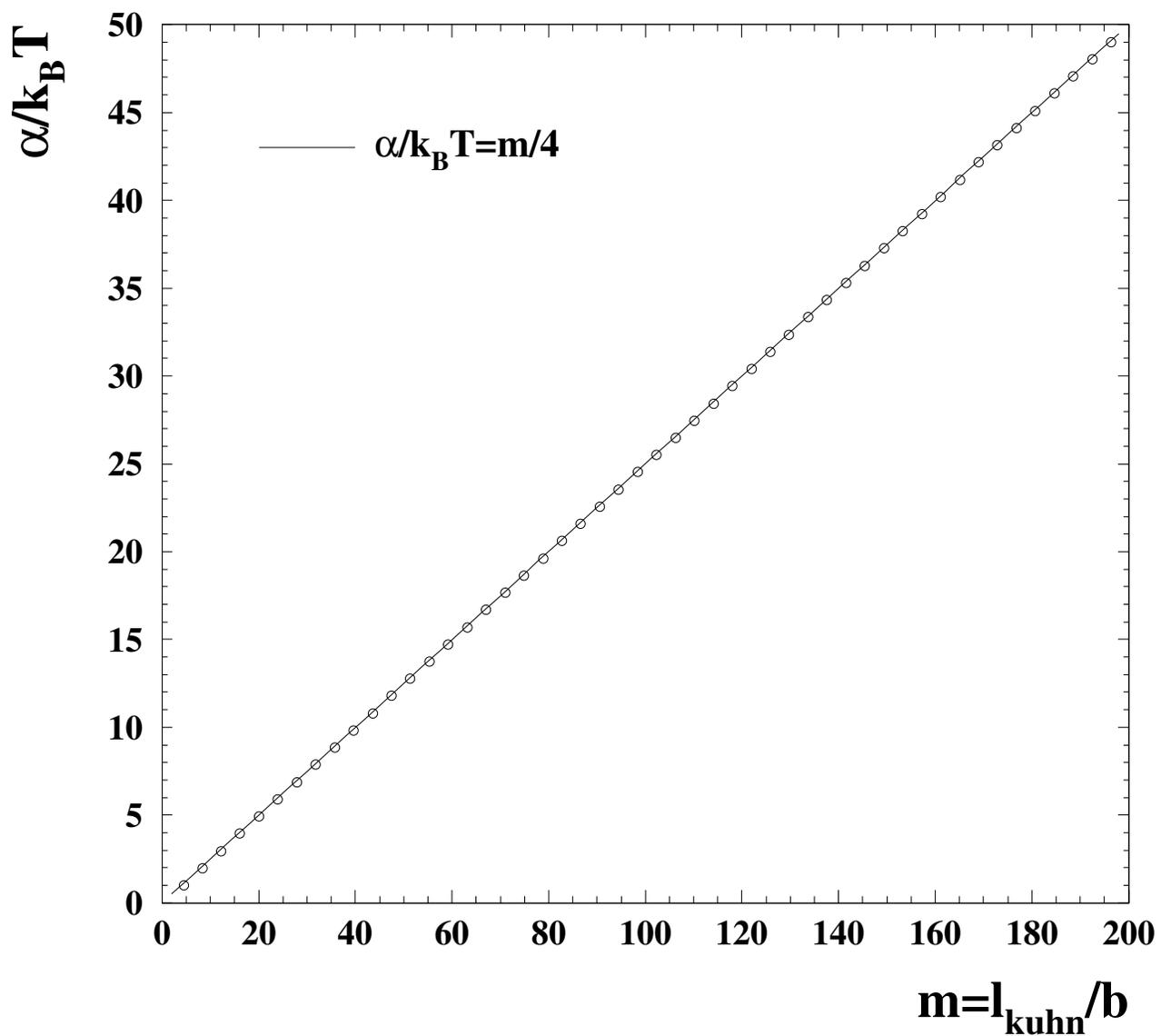


Figure 37

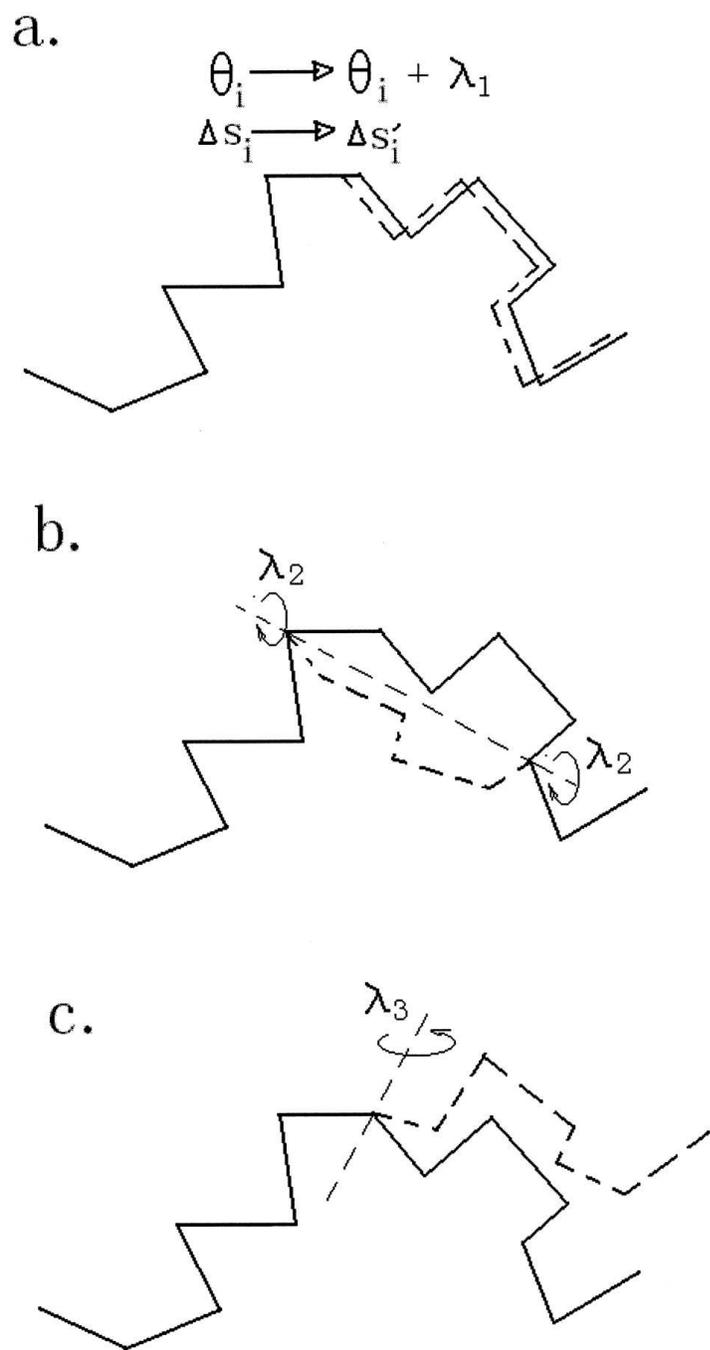


Figure 38

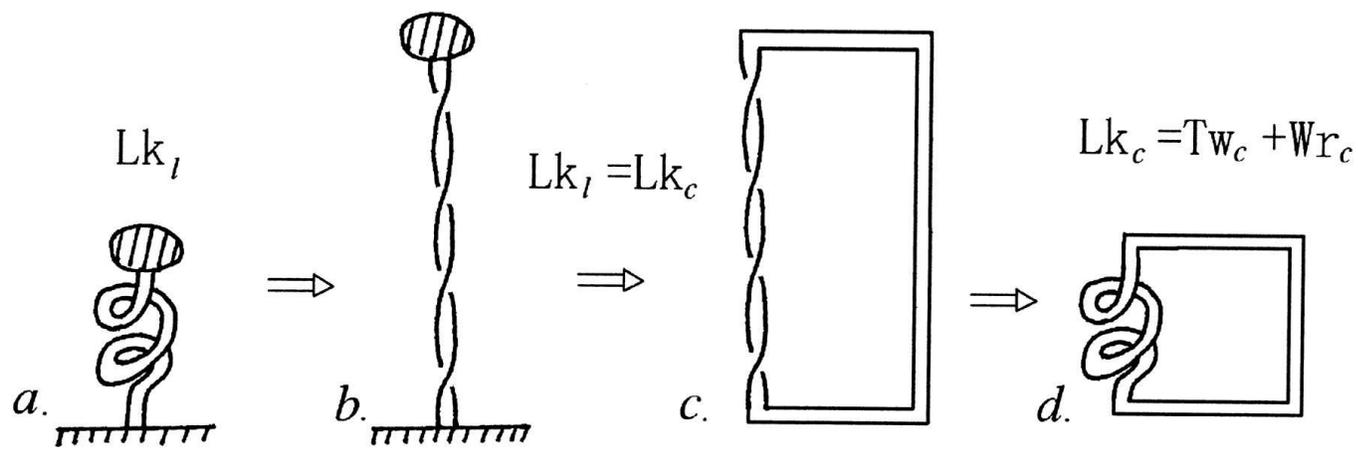


Figure 39

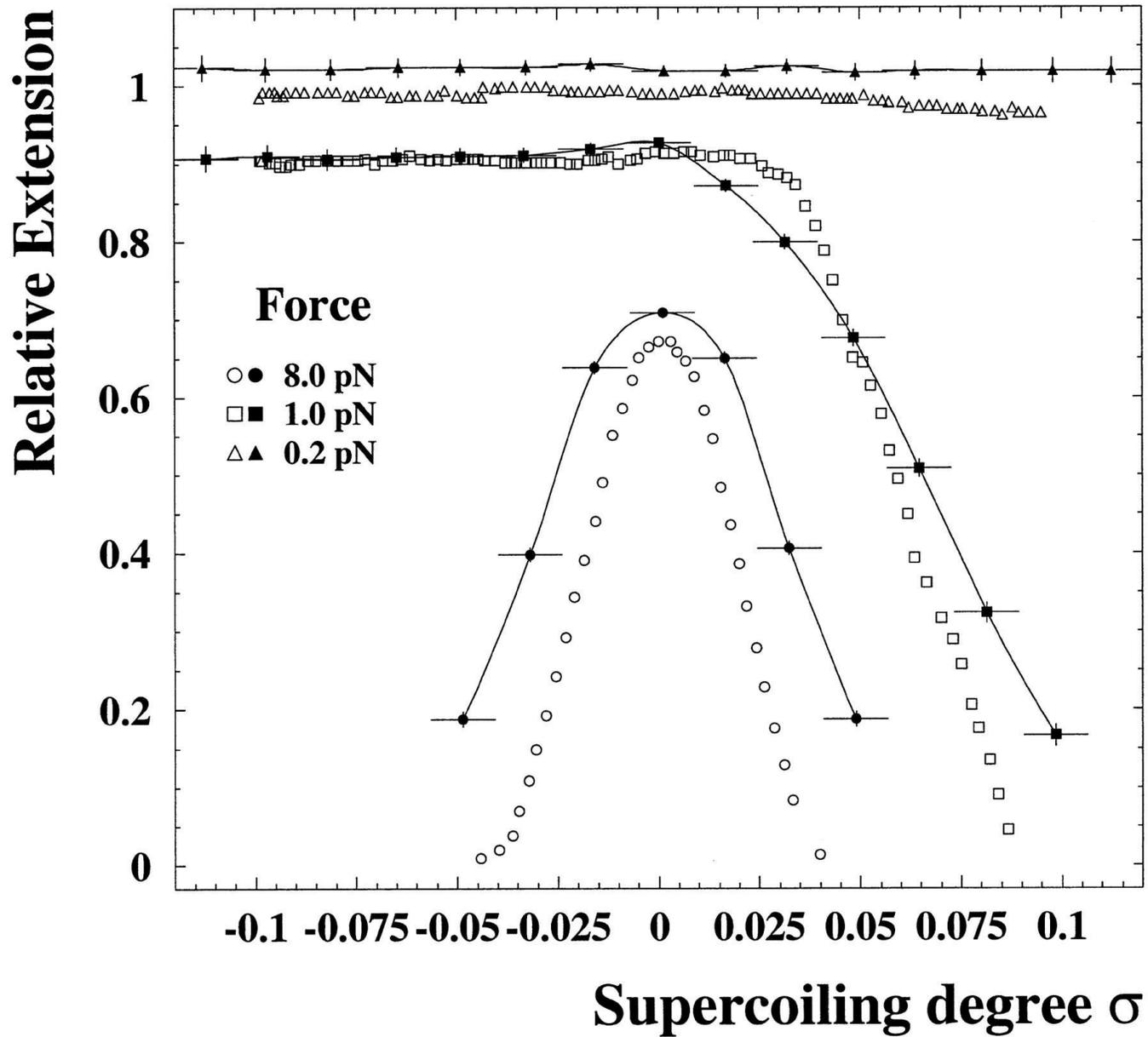


Figure 40