THEORETICAL MICROMECHANICS OF DNA AND DNA-PROTEIN COMPLEXES

BY

ABHIJIT SARKAR
B. Sc., Jamia Millia Islamia, New Delhi, India, 1993
M. Sc., Jamia Millia Islamia, New Delhi, India, 1995

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Chicago, 2002

Chicago, Illinois

I dedicate this dissertation to my two best friends and only benefactors, my father Mr. Chinmoy Sarkar and my mother Mrs. Mili Sarkar. Their sacrifice and encouragement brought me opportunities they did not have. They are responsible for my truly privileged life.

ACKNOWLEDGMENT

I thank my dissertation director, Prof. John F. Marko, for giving me the opportunity to work in his lab and showing me the way through the thicket of graduate work. Whatever may be judged its contribution to the theory of single-molecule experiments, this dissertation reflects, in the greatest measure, the care, commitment and indefatigable patience of Prof. Marko with me. Having so taxed Prof. Marko's considerable intellectual, moral and managerial resources, I will gladly attest to how great an understatement my words here are to the true extent of his hand in the execution of my graduate research.

I thank Prof. Didier Chatenay of Université Louis Pasteur for visiting our laboratory in 1998 and sharing crucial insights into experiments on torsionally constrained DNA that were performed in his laboratory. His encouragement and guidance contributed significantly to the completion of the theoretical work we undertook to explain his results.

Other members of the UIC physics faculty, particularly Profs. Anjum Ansari and Mark Schlossman, have been a constant source of good advice and encouragement.

I thank Profs. Jie Liang and Serdar Ogut, who were on my thesis defense committee, for providing many helpful suggestions.

I have had the privilege of working with colleagues like Mr. Jie Yan, Dr. Michael G. Poirier, Mr. Diego Dominici, Ms. Dunja Skoko, and Ms. Chee Xiong, who have been both good friends and patient interlocutors on scientific matters, besides being excellent students and scientists.

ACKNOWLEDGMENT (Continued)

I thank Prof. Q. N. Usmani, Professor of Physics at Jamia Millia Islamia, New Delhi, India, who was instrumental in my gaining admission to the University of Illinois at Chicago. During my studies at the Jamia Millia Islamia, Profs. Pankaj Sharan, Mohammed Sami, and Lekha Nair, all of the Department of Physics there, were always generous with their time, encouragement and wisdom. For this, I thank them.

This research was supported by NSF Grants DMR-9734178 and DMR-0203963, by a grant from the Petroleum Research Foundation, by a Research Innovation Award from the Research Corporation, by a Biomedical Engineering Research Grant from the Whitaker Foundation, by a Focused Giving Grant from Johnson & Johnson Corporate Research, and by the University of Illinois Foundation.

AS

TABLE OF CONTENTS

CHAPTER **PAGE**

CHAPTER **PAGE**

LIST OF FIGURES

FIGURE **PAGE**

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
ssDNA	Single-Stranded Deoxyribonucleic Acid
dsDNA	Double-Stranded Deoxyribonucleic Acid
RNA	Ribonucleic Acid
bp	Basepair
kb	kilobase
RNApol	RNA Polymerase
FJC	Freely-Jointed Chain
WLC	Worm-Like Chain
AFM	Atomic Force Microscopy
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
ATP	Adenine Triphosphate
NTP	Any of Adenine, Guanine, Thymine, and Uracil Triphosphates

LIST OF ABBREVIATIONS (Continued)

N	Newtons, Unit of Force
pN	PicoNewton, 10^{-12} Newtons, Unit of Force
K	Degrees Kelvin, Unit of Temperature
rad	Radians, Measure of Angle
ns	Nanoseconds, 10^{-9} seconds
nm	Nanometer, 10^{-9} meters
k_BT	Unit of thermal energy, 4.1×10^{-21} Joules at room temparature

SUMMARY

This dissertation is a description of my research in theoretical problems in biomolecular micromechanics. Techniques of theoretical condensed matter physics are used to understand single-molecule micromanipulation experiments. Modeling of DNA-protein interactions involved in important genetic processes is used to explore different mechanisms for these processes and to guide further experimentation.

In Chapter 1, I summarize the history of DNA structural biology and describe the molecular structure of DNA. The essentials of the polymer description of DNA are introduced and a brief account of the force, energy, and time scales characteristic of the cellular world given. In subsequent sections of this chapter, I present background and summaries of my research in DNA elasticity (section 1.5), in long range twist propagation along DNA (section 1.6) and unzipping DNA with 'obstructions.' Each of these topics is discussed in greater detail in Chapters 2, 3, and 4, respectively.

In Chapter 2, I describe in greater detail the theoretical description of fixed-twist DNA elasticity experiments. It is shown that DNA has four novel stress-induced structural states besides its native B-form structure. Some structural parameters of these states, like mean twist rate, are determined by fitting the model to the data. The results are consistent with previous work on novel structures in DNA and extends our understanding of the full range of DNA's elastic response. The results of this work are summarized in two phase diagrams showing which structural states are dominant for given forces and torques (or twists).

Chapter 3 explores whether twist pulses can propagate far enough in DNA to affect the binding of other proteins. Given the ubiquity of proteins that apply twist pulses to DNA, this problem has significant biological relevance. The main tool is a simple equation of motion that describes twist transport in DNA. Taking up the cases of RNA polymerase (RNApol) and DNA gyrase in detail, arguments are presented for the particular form of the initial condition for twist pulses generated by these two protein motors. Solving the equation of motion leads to an estimate of how far along the DNA torques can be remotely delivered. In the conclusion, it is suggested that this work is relevant to the study of transcription and chromosome remodeling.

Chapter 4 reports work on DNA unzipping in the presence of sequence inhomogeneities and proteins which act like obstructions preventing uniform unzipping of the DNA. A simple statistical mechanical model is developed to characterize the force signals that would be seen experimentally when the unzipping fork reaches the obstruction. It is found that for both sequence inhomogeneities and bound proteins, there are distinctive force signals from the unzipping of DNA inside the inhomogeneity or in contact with the protein. These signals are large enough in force that they can be distinguished from the background. Detailed analytic calculations in the mean field approximation are presented and they show how, at least in case of sequence inhomogeneities, the shape of the force signal can be modulated by altering the length of the the inhomogeneity or its cohesive strength. These results are also established using exact calculations.

In Chapter 5, I summarize the results of this thesis and discuss its biological relevance. I present a list of experiments suggested by this research and indicate directions for future growth for theoretically-inclined work in biomolecular micromechanics.

CHAPTER 1

SYNOPSIS OF BIOLOGICAL, CHEMICAL AND POLYMER FEATURES OF DNA

1.1 Introduction

The discovery of the structure of DNA and its central role in life and, subsequently, the reduction of life to a set of well-defined and essentially universal chemical phenomena is surely one of the greatest advances in human knowledge. We shall, in Section 1.2, present a very brief summary of the key historical events connected to the discovery of the structure of DNA. We will also describe the chemical or primary structure and the spatial or secondary structure of DNA. In Section 1.3, we summarize the polymer properties of DNA since DNA in cells is usually found as very long polymers, much longer than any artificially synthesized polymer. Then, Section 1.4 presents the relevant force, energy, and time scales needed to understand single-DNA molecule experiments. Section 1.5 describes how DNA responds to an imposed stretching force. This section presents some of the background on DNA elasticity experiments and qualitatively discusses recent work studying how DNA with torsional constraints behaves under an applied force. Section 1.6 briefly describes theoretical work on the possibility of driving bound proteins of their DNA template through propagating twist pulses. Finally, Section 1.7 introduces our work on unzipping DNA with sequence inhomogeneities or bound proteins.

1.2 DNA's Biological Significance

1.2.1 DNA: Summary of Key Historical Events

DNA in living cells is the carrier of inheritable genetic information that allows cells to function normally and produce copies of themselves later on. Proteins do most of the work in cells and are key structural elements in the cell, and DNA contains the plans for building all the proteins the cell will need. Moreover, DNA controls when a given protein will be produced, through regulatory information elements, which do not serve as blueprints for proteins but rather as conductors orchestrating the creation of proteins in cells.

It was known since the work of Johann Friedrich Miescher and Ernst Felix Hoppe-Seyler in 1869 that DNA was present in cell nuclei. However, it was the work of Avery, Mcleod and McCarty in 1944 (1) that provided the first solid evidence that DNA was in fact the genetic material, the carrier of hereditary traits from parents to progeny, and the memory element of cells that held all the vital information needed to build and run them. Essentially, in their experiment, they transplanted DNA from one cell-type to another and saw the host cell acquire phenotypic characteristics of the donor cell.

Erwin Chargaff, in 1950, found that the four species of nucleotides, the basic repeat units of the DNA polymer, were found in certain defined ratios. Adenine (A) and thymine (T) were to be found in one-to-one proportion in DNA extracted from a number of unrelated organisms (2). Similarly, nucleotides guanine (G) and cytosine (C)(2) were also found in 1:1 ratio. This was a crucial piece of data because it suggested that these ratios were a universal feature of DNA, quite independent of where the DNA came from.

The first attempt to put forward a coherent account of the structure of DNA was by Linus Pauling (3). It was not successful, although, as we shall see in Chapter 2, recent single-molecule DNA micromanipulation experiments, where an individual DNA polymer is subjected to forces large enough to overwhelm the inter-nucleotide bonds, have found evidence that under extraordinary mechanical stress DNA can acquire a structure that has some features in common with Pauling's 1953 suggestion (28).

The correct elucidation of DNA's structure came in 1953, in the work of Francis H. C. Crick and James D. Watson (4). They proposed a model which synthesized the researches of Chargaff, the X-ray diffraction studies of Rosalind Frankin (5) and Maurice Wilkins (5), and quantum-chemical considerations about bonding schemes amongst the nucleotides. Watson and Crick proposed that B-DNA, i.e. DNA found in most living cells under normal conditions, consisted of two intertwining polynucleotide chains, held together by hydrogen bonding between certain chemical subunits of the nucleotides, called bases. The bases are responsible for the four flavors of nucleotides for they come in four chemical types; the rest of the nucleotide is common to A, G, C, and T. Quantum mechanical calculations about the strength of hydrogen bonding between bases established that base A preferentially bonds with T, and G with C. Thus, Watson and Crick proposed that base A in one chain will always face base T on the other and so too for G and C in a scheme of complementary basepairing. In summary, DNA was found to be a double-helical structure built out of two mutually-twisting polymer ropes with the base sequence of each strand a strict complement of the other.

With the role of DNA as the genetic memory element of cells and its three-dimensional structure established, the search began for the code which allowed the sequence of nucleotides along DNA to store genetic information. The importance of knowing the 3D structure of DNA was now fully appreciated since it immediately suggested how the coding scheme works i.e. how the sequence of bases specifies the blueprints for proteins and the regulates protein production and destruction in cells. This problem was essentially solved by Crick *et al* in 1961 (6). They found that in parts of the DNA which code for proteins, every consecutive triplet of bases stands for one of the 20 amino acids that make up proteins (amino acids are the building blocks of proteins, which are their polymers, like nucleotides are the building blocks of DNA). We shall not discuss the details of the code but simply want to emphasize its central role in cells and therefore all of biology.

1.2.2 DNA's Chemistry and Structure

As described above, deoxyribonucleic acid or DNA is a linear polymer composed of two separate strands, each a covalently linked chain of nucleotides, which, in the absence of appreciable stress, firmly hydrogen-bond together to form a right-handed double helix (Fig. 1.1). This classical 'B-DNA' structure is the basic DNA conformation found inside living cells. B-DNA is stable in physiological aqueous solution, i.e. water with buffered pH near 7.5, and with univalent salt of roughly 0.1 M concentration.

Nucleotides, which are the monomers that make up each arm of the DNA, come in four varieties and are composed of three distinct parts: a sugar called deoxyribose, a phosphate group, and one of four units, adenine (A), guanine (G), thymine (T), and cytosine (C) called

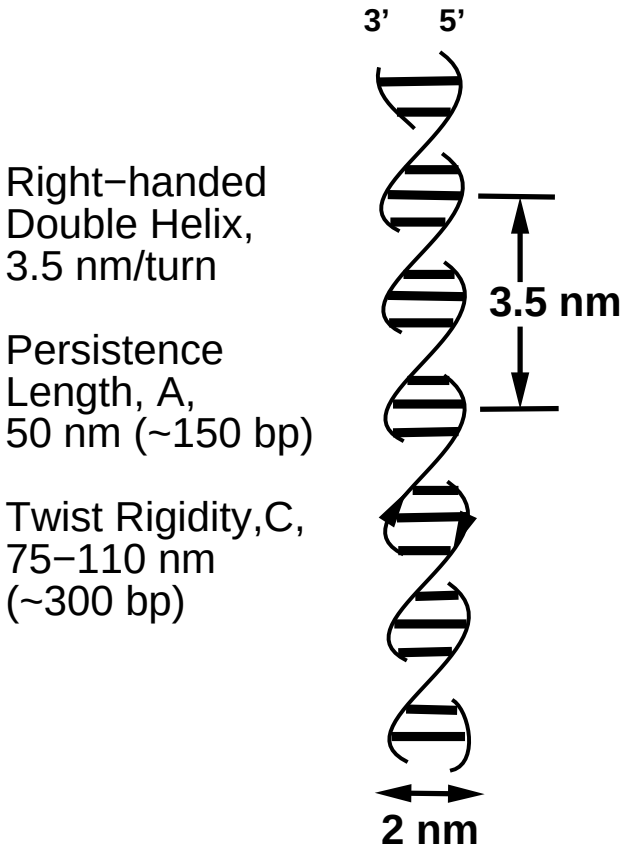


Figure 1.1. The Watson-Crick double helix. It takes 3.5 nm or 10.5 bp to complete one turn of the double helix. The labels 3' and 5' indicate the orientations of the two strands: the strand on the left runs 5' to 3', the other 3' to 5'. We have only labeled one terminus of each strand; arrows on the strands also indicate their direction. The directedness of the strands helps guide proteins that bind to and then sequentially 'read' DNA. Structurally, oppositely directed strands maximize hydrogen bonding between complementary bases. The persistence length of DNA, which is the distance over which DNA can be thought of as a rigid rod, is 50 nm; bends are significant for molecules many persistence lengths long. The twist rigidity is the distance over which fluctuations in the DNA twist rate are insignificant, i.e less than 1 radian. DNA radius is $\approx$ 1 nm.

bases. The sugar is attached on one side to a base, and on the other to a phosphate group by covalent bonds (Fig. 1.2). Each of the four bases attached to nucleotides is hydrophobic and capable of hydrogen-bonding with another complementary base according to the complementary basepairing rule: A always bonds with T and G always bonds with C.

Each arm of the DNA is formed, in turn, by a covalent bond between the phosphate group of one nucleotide and the hydroxyl group of its neighbor nucleotide's sugar (Fig. 1.2). This bond is often referred to as the phosphodiester bond and the two arms collectively the DNA backbone. Since hydrogen bonding between complimentary bases is what holds the two strands of DNA together, knowing the sequence of bases along one strand of the double helix is sufficient to deduce the sequence of the other.

In the assembled DNA molecule, the phosphate groups on the nucleotide polymers are the outliers, protruding into the medium (basically water) in which DNA is suspended, while the bases, which are hydrophobic, point inward facing each other. Water is thus expelled from the interior of the DNA molecule. The two polymer strands of DNA thus embrace each other through (1) hydrogen-bonding or basepairing between complementary bases on opposite arms and (2) entropically-driven hydrophobic or stacking interactions, which result in a net B-DNA stabilization energy of few k_BT per base-pair. The resulting minimum free energy configuration of the two mutually enveloping strands of ordinary DNA is the Watson-Crick double helix sketched in Fig. 1.1. The distance between any two neighboring basepairs in DNA is 0.34 nm and it takes 10.5 basepairs (bp) to complete on turn of the helix, so that the helix repeat is $\approx$ 3.5 nm. The van der Walls diameter of the DNA molecule is $\approx$ 2 nm.

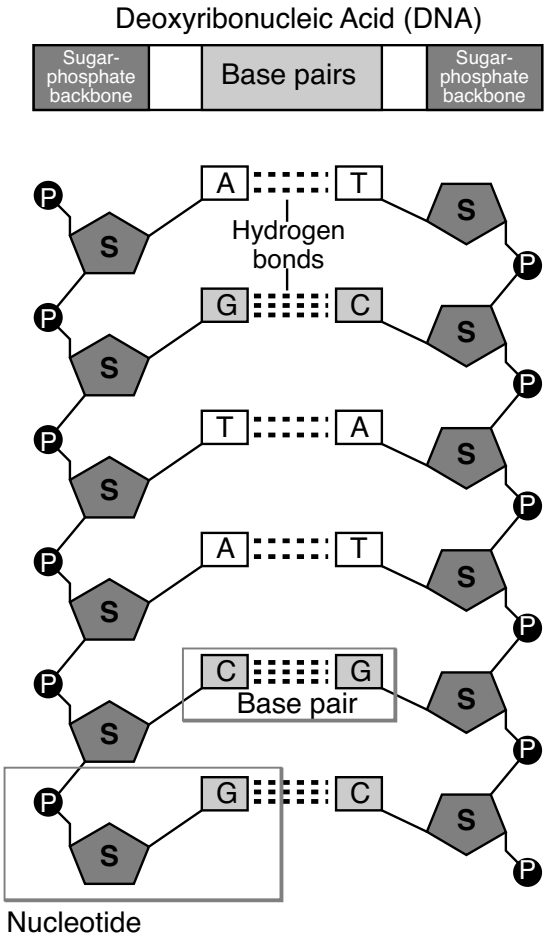


Figure 1.2. The primary structure of DNA. A nucleotide consists of a sugar linked on one side to a phosphate group and on the other side to one of four possible bases. Each nucleotide links up with another to form the strands of DNA. The two strands are held together by complementary hydrogen bonding between bases which, due to their hydrophobicity, point 'inwards' and thus 'face' each other. This figure courtesy of National Human Genome Research Institute; see www.nhgri.nih.gov/DIR/VIP/Glossary/ Illustration/Pdf/nucleotide.pdf.

DNA has a charge density of two electrons per basepair, with the phosphate groups in a complementary basepair supplying one electron each. This high charge density is screened strongly in physiological solutions, so that repulsion between the two sister strands is reduced sufficiently that the double helix is stable at room temperature.

1.3 Double-Stranded DNA as a Polymer

Single contiguous double-stranded DNA molecules (dsDNA) in cells can have total contour lengths (i.e. the total length of the central axis of the molecule) ranging from 10 μm ($\approx 10^4$ base pairs) for viruses to greater than 1 cm ($\approx 10^8$ base pairs) for higher organisms. Because of constant kicks imparted by the molecules of the surrounding aqueous medium, dsDNA spontaneously and randomly undergoes thermal bending fluctuations throughout its length. However, steric constraints, arising from the local geometry of the double helix, couple the orientations of neighboring monomers. This coupling becomes weak on large scales; the orientation of base pairs $>$ 50 nm (150 bp) (34) apart are statistically independent. This length of 50 nm is the persistence length of dsDNA. Long unstressed dsDNA molecules look like random walks on length scales much bigger than the persistence length, but like rigid rods within a persistence length.

The surrounding heat bath also produces transient fluctuations around the mean twist angle of $\approx 34^o$ between adjacent basepairs. Once again, each backbone constrains the spontaneous torsional motion of the other, correlating twist orientations over a length C $\approx$ 100 nm (43; 69). Walking along the contour of a DNA molecule, one would find random twist variations bigger

than 1 radian over distances greater than C. In other words, fluctuations in relative angle of two adjacent base pair stacks are statistically independent over base pairs 100 nm apart.

1.4 Force, Energy and Time Scales Relevant to Biological Micromechanics

Since cells are immersed in a heat bath at room temperature, the basic unit of energy for them is k_BT, which for T=300 K is 4.1×10^{-21} J = 4.1 pN nm (1 pN = 10^{-12} N and 1 nm = 10^{-9} m). The persistence length of DNA, 50 nm, sets the relevant length scale for understanding DNA's low force elasticity and attendant phenomena. Another important length scale, already mentioned, is 1 bp = 0.34 nm. Also, as we mentioned previously, it takes about 3.5 nm, or 10.4 bp, to complete one full turn of the double helix. Since water, at room temperature, surrounds and damps fluctuations of biological molecules, its viscosity often sets the dominant time scale, $t_{biol} = \eta l^3/k_BT$. With $l \approx 1$ nm, $\eta = 0.001$ Pa·sec, and $k_BT = 4.1$ pN nm, $t_{biol} = 1$ ns.

1.5 Static Response of DNA to External Forces and Torques

1.5.1 Force Transduction in PicoNewtons

Thermal fluctuations in the medium surrounding a dsDNA will introduce kinks in the polymer so that a typical polymer configuration will look like a random walk in space. Straightening out these thermal bends will require a force sufficient to overcome 1 k_BT of energy per persistence length of DNA. Since $b = 50$ nm for DNA and 1 $k_BT = 4.1$ pN nm, we expect a force of ≈ 0.1 pN to straighten these bends.

On the other hand, from our discussion of the binding energy of the two complementary strands of DNA (free energies of only a few k_BT per base-pair (8)), we estimate that forces on the order of $\approx 10\ k_BT$/nm, when applied to B-DNA, will induce conformational change to some

other DNA structure. Again, since k_BT/nm is 4.1 pN, we should now expect DNA structural transitions to be associated with forces on the order of 40 pN.

The advent of picoNewton force transducers has allowed single-molecule micromechanics to move beyond theoretical speculations and into the realm of experimental practice. We sketch, in Fig. 1.3a, a single-molecule stress-strain experiment. DNA with specially adapted ends is made to stick to coated beads or micropipettes. Beads can be manipulated by being trapped in laser spots, or by micropipettes using suction. With both DNA ends held by beads in laser traps and/or micropipettes, one end is moved by a fixed amount with respect to the other, which is held fixed, and the elastic force on the force transducer generated by the molecule is measured. Forces on laser-trapped beads are measured from their thermal fluctuations in the laser spot. For micropipettes, direct observation of their bending allows calculation of force through calibration of pipette bending moment.

Another method for mechanically manipulating single DNA molecules is to attach a paramagnetic bead to one end of the DNA and, on the other end, a chemical tag which sticks to a glass cover slip. With a single DNA molecule interpolating between the two beads, a magnetic field can be used to apply a force to the magnetic bead at the 'free' end. This arrangement is often referred to as a magnetic tweezer (31; 21). The unique feature of magnetic tweezers are that the apparatus itself has zero force constant, i.e. the force acting on the molecule is for practical purposes spatially constant. Magnetic tweezers are not good for positional control but are very well-suited for low-force experiments.

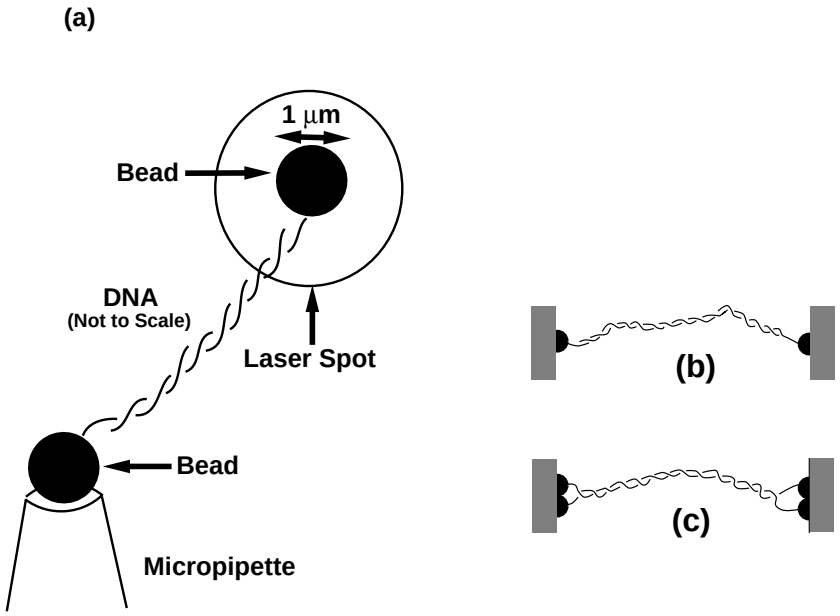


Figure 1.3. Schematic of a DNA micromanipulation experiment (a) and the different attachment schemes for a single molecule of double-stranded DNA (coils) to the force transducer (bulges) (b) and (c). Fig. 1.3a shows a DNA molecule, imagined to be many kilobases long, attached at its two ends to beads, which themselves are trapped by a laser spot, suction from a micropipette, or held in a magnetic field. Beads are typically ≈ 1 μm is radius. Once the DNA+beads complex has been trapped, one end is held fixed while the other is moved in fixed steps. Force can be measured through a variety of calibration schemes which depends on the specifics of the force transduction technique. Fig. 1.3b shows single strand attachment which allows Lk to fluctuate freely because the DNA can rotate about itself if pulled. Fig. 1.3c illustrates topology-fixing double strand attachment.

1.5.2 Summary of the History of Biomolecular Micromechanics

Even though Wilkins *et al* (9) in 1951 were able to reversibly stretch entwined DNA fibers upto 1.5 times their native length, the explosion of activity in the field of single-DNA molecule micromanipulation can be traced back to the pioneering work of Steven B. Smith and A. J. Bendich in 1990 (10). They used an electrophoretic force field to stretch a single DNA molecule by forcing it to snake its way through agorose gel. The next important step was the work of Smith *et al*, in 1992. They used a magnetic tweezer apparatus to stretch a single, isolated DNA molecule from λ-phage (31), a virus that infects E. coli bacteria. In 1994, Bustamante *et al*, in collaboration with the theorists (34), established conclusively that DNA's elastic response, as seen in single molecule stretching experiments, was described by a worm-like chain (WLC) model rather than the freely jointed chain (FJC) used by Smith *et al* in their 1992 paper. In the same year, Steve Chu's group reported experiments on stretching λ-DNA using hydrodynamic flows (32; 33), which culminated in the 1995 paper by Perkins *et al*, where they observed DNA's low force (less than 10 pN) elastic response that reproduced and strengthened the results of Smith *et al* in (34).

A major challenge in the field after 1995 was how to explore the high force, i.e. greater than 10 pN, elastic regime of DNA. Magnetic tweezers, hydrodynamic flows, and optical traps used by various groups were inadequate for the purpose of applying larger forces on DNA. The search was motivated, at least in part, by predictions (11) that for sufficiently large forces, $\sim$40 pN, DNA should abruptly transform from B-DNA to another novel structural state, S-DNA. Understanding high-force stretching of DNA was also biologically interesting since a number

of important DNA-acting proteins apply forces $\sim$ 10 pN when they bind to DNA. Indeed, as Leger *et al* found in (20), an important class of DNA binding proteins, RecA, generates forces of $\sim$ 100 pN when it binds to DNA, thereby significantly stretching its DNA template.

Pioneering experiments by Cluzel et al (12) and Smith et al (13) showed that indeed a sharp structural transition occurs in dsDNA when under roughly 65 pN of tension. This dramatic transition is reversible and is possible because the covalently bonded sugar-phosphate backbones along each strand are helically coiled inside B-DNA. This new stretched DNA state is widely called 'S-DNA.' This transition has been observed in many laboratories, and is used as a force calibration in many experiments.

A significant shortcoming of all the experimental designs described so far was that, while they stretched DNA, they were unable to control the DNA's twisting, i.e. the rate at which one strand winds around the other. The total number of turns of one strand of DNA around the other is the molecule's twist, Tw. When the DNA is held taut, Tw is also its linking number, Lk (the general definition of Lk is presented in Chapter 2). As we stretch DNA, its two strands will tend to unwind if we allow the molecule to freely rotate about its inter-helical axis. In Fig. 1.3b, we show a cartoon of DNA attached to 'beads' through only one of its two arms; this arrangement allows the molecule to freely swivel around its body axis when it is being stretched. In Fig. 1.3c, we show double-stranded attachment of the DNA to 'beads'; this prevents the DNA from spinning about itself so that the molecule does not unwind as we stretch it.

In order to study how constraining DNA's twist affects its elastic response, two groups developed experimental techniques to torsionally constrain DNA, that of Strick *et al* (21) and

Cluzel *et al* (22; 23). These experiments allow one to measure the force-extension curve for a dsDNA subject to the constraint that its double-helix linking number is fixed. Early experiments by Strick *et al* (21) showed that for low forces < 5 pN, effects on the entropic elasticity of DNA could be observed and studied, as a function of twisting. Studies of this low-force regime verified theoretical predictions of coexistence of supercoiled and extended domains (24), and denaturation of undertwisted DNA at forces of a few pN. Subsequent and detailed theoretical studies (25; 26; 27) have painted a rather complete picture of the low-force behavior of twisted DNA.

Slightly later experiments by the same group explored the behavior of twisted DNA at higher forces, and showed that dsDNA could be denatured by untwisting (28; 29), and very heavily overtwisted (30). Finally, work of Leger *et al* (23) made a global study of the force-extension behavior over a wide range of twisting, and for large forces in the range 1-100 pN. Remarkably, those experiments found four force-plateau-type transitions, suggesting that four DNA states in addition to B-DNA can be accessed using different forces and twisting. However, a full interpretation of the experimental results requires a theoretical analysis (see Chapter 2). Finally, other DNA-stretching experiments have indicated that dsDNA can be stretched to as much as double its B-form length (14; 15), close to full extension of the sugar-phosphate backbones.

1.5.3 DNA Elasticity in Absence of Twist Constraint

In Fig. 1.4 (dashed), we plot DNA's force-extension curve when it is free to adjust its twist under external stress. At low forces, i.e forces in the range 0.1 pN to 10.0 pN, the applied force

does work against the random thermal bends introduced by the immersion of the polymer in an ambient thermal bath. Initially, DNA behaves like a linear spring, with extension proportional to force, but as more and more kinks are taken out, linear elasticity gives way to highly nonlinear behavior. In fact, by the time the molecule has been extended to 30% contour length, the force response is already non-Hookean. When $f = 10$ pN, the molecule is taut and thermal bending fluctuations no longer play a role.

A theory for the non-linear part of DNA's elasticity up to 10 pN has been developed (11). The crucial finding was that the force scales as $1/(1 - x/L)^2$ with relative extension in the nonlinear regime. For lower forces, the usual flexible-joint chain (FJC) model works well. A uniform extrapolation formula for force vs. distance valid across the range 0.1 pN to 10.0 pN has been worked out:

$$f = \frac{k_B T}{A}\left[\frac{x}{L} + \frac{1}{4(1 - x/L)^2} - \frac{1}{4}\right]. \tag{1.1}$$

For forces bigger than 10 pN, the molecular extension initially varies linearly with force. But, at 65 pN, the molecule extends abruptly over a narrow range of force ($\approx$ 5 pN), from about 1.1 times to about 1.7 times its unstressed B-form contour length. The linear relationship between force and extension is recovered after the transition but with a different slope. The B-S transition is reversible and robust to the extent that it is often used to calibrate other single-molecule experiments.

We should point out in passing that almost all single-molecule experiments, involving DNA, including the experiment whose data we are reporting, use DNA from the phage λ, which has 48,502 basepairs and a total contour length of ≈ 16.3 μm.

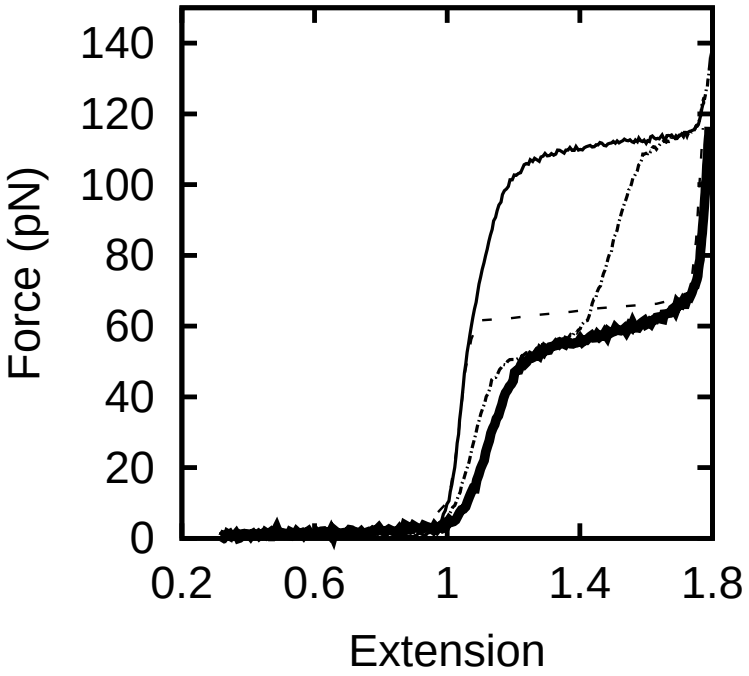


Figure 1.4. DNA's elastic response when its twist is free to fluctuate (dashed) and when not (thin solid, dots-dashes, thick solid). When twist is allowed to freely adjust (dashed), the low force entropic elasticity, i.e. stretching response when applied forces are less than 10 pN, gives way to an initially linear behavior beyond an extension of 1.1. Around 65 pN, a highly cooperative transition ensues. For even higher forces, linear behavior is restored, although with a different slope. When $\sigma = 0$ (thin solid), i.e. when λ-DNA has its twist fixed to its native value, the 65 pN transition is absent, replaced by another at 100 pN. Since S-DNA is undertwisted, the 65 pN is forbidden since this would violate the twist constraint. When 35% of turns are initially taken out of the DNA, i.e. when $\sigma = -0.35$ (dots-dashes), another plateau appears at 45 pN, while the 100 pN plateau is greatly diminished. At $\sigma = -0.71$ (thick solid), the 100 pN plateau has completely disappeared.

1.5.4 DNA Elasticity in Presence of Twist Constraint

We shall only summarize here the salient features of DNA elasticity under conditions where the molecule's twist is fixed throughout an experimental run, since Chapter 2 is dedicated to the theoretical description of these experiments. In Fig. 1.4, σ is the fraction of twists taken out or put into the DNA over and above the native Tw of λ-DNA. We plot force traces for $\sigma = 0$ (thin solid), -0.35 (unevenly dashed), -0.71 (thick solid). It is apparent in the data that when $\sigma = 0$, there is no transition at 65 pN; instead it is shifted to 100 pN. For negative σ's, a second plateau, also distinct from the 65 pN plateau, arises. As we make σ more negative, the upper 100 pN plateau diminishes while the lower 45 pN plateau grows. Eventually, at $\sigma = -0.7$, the 100 pN plateau has completely disappeared. For positive σ, the 100 pN plateau is supplanted by one at 25 pN which, like its negative σ cousin, progressively reduces the 100 pN plateau for larger σ's.

The explanation of these phenomena begins with the observation that forcing DNA to stretch also compels it to untwist. When stretched DNA is not allowed to untwist, say, through the use of a mechanical constraint as is the case here, we forbid it to transform into other states the violate the twist constraint, e.g. S-DNA, since S-DNA is undertwisted w.r.t B-DNA. This eliminates the 65 pN transition; only at the much higher force of 100 pN does the DNA find a way to solve its problem (find the lowest free-energy state), when it becomes energetically favorable to create a *mixture* of overtwisted P-DNA and undertwisted S-DNA. With P- and S-DNA is the right proportion, the twist constraint, say $\sigma = 0$ can be satisfied and the molecule can relax to a lower energy state.

Why the lower plateau for σ-negative? When we take turns out of the DNA, we create a mixture, even at low forces, of undertwisted Z-DNA, which stores the untwists, and B-DNA. The first transition corresponds to the transformation of Z-DNA into a mixture of P- and S-DNA, again in the ratio needed to satisfy the twist constraint. The rest of the molecule, which is still in B-form and therefore has its σ fixed to 0, transforms to a mixture of P- and S-DNA only at 110 pN. More negative the initial σ's, more of the molecule transforms to Z-DNA initially, wider the 45 pN transition and shorter the 100 pN transition. Essentially the same argument can be made for positive σ's, except that instead of Z-DNA, we need to introduce a state capable of absorbing overtwists (sc-P-DNA).

In Chapter 2, we construct a simple statistical mechanical model of mechanically stressed double stranded DNA with its twist fixed externally and exactly evaluate its partition function by numerically diagonalizing the transfer matrix. We determine the parameters of our model by a fit of the theoretical force-extension curves to those of experiment (23).

1.6 Stability of Proteins Bound to DNA Under Transient Twist Strains

DNA can be stretched and twisted. Stretching DNA with a protein bound to it can remove the protein (55). To see why, suppose that protein binding uses up a length D of DNA out of the total length L. This leaves a length $L - D$ to absorb an applied stretching force. Since the length D has been captured by the protein and is not available support the stress, the DNA molecule has a larger free energy. But this is compensated by the extra thermodynamic stability that the bound protein brings to the system as measured in a positive protein binding free energy μ. So, as long as $\mu > fD$, where fD is extra lowering of the free energy that can

be achieved by releasing the captured length D, the protein remains bound. This implies that for typical proteins stretching forces less than a few pNs are enough to dislodge them.

Experiments studying force-induced destabilization of bound proteins have recently been performed (64; 65; 66). Cui *et al* (64) and Bennink *et al* (65) have studied the force-response of single chromatin fibers. Ali *et al* (66) have studied force-generation in IHF, a protein used in bacteria to fold DNA up. The picture provided by theory has generally been accepted as a useful baseline in discussions of the experimental results.

Twisting DNA can also destabilize a DNA-bound protein. When proteins bind, they to some degree constrain the twist of the segment of DNA to which they are bound; the physical contact between the protein and the DNA may prevent twist distortions from traveling into the DNA region "underneath" the protein. In this case, when a static twist strain is applied to some length L of DNA, which has a segment of length D 'absorbed' by a protein, say in the form of a loop as shown in Fig. 1.5b, the remaining length $L - D$ must absorb the extra twist. The work that needs to be done to spread a total twist Θ over a length $L - D$ is $C\Theta^2/[2(L - D)]$ in k_BT units (we consider DNA not more than a few persistence length long so that it can be idealized as a rigid rod), where C is the twist persistence length, ≈ 100 nm. Clearly, if the length D of DNA was also available, the elastic twist energy would be lower. But, when the protein binds it lowers the free energy of the system by μ, the protein binding free energy. So we expect the protein to remain bound, even in the presence of an applied torque, as long as $C\Theta^2/2D < \mu$. In terms of a torque τ, when $\tau^2 D/2C = \mu k_B T$, the protein should have probability $1/2$ to come unbound. Since $\mu \approx 5k_BT$ for typical proteins and $D \sim 10$ nm, $\tau \approx 10k_BT$. This magnitude of

torque is not atypical of that produced by molecular motors like gyrase and RNA polymerase (RNApol).

In Fig. 1.5, we illustrate different ways in which DNA-protein structures constrain DNA twisting. Torques are imagined to be applied to DNA from the left end while the right end of molecule is firmly anchored so that DNA+protein cannot rigidly rotate as a whole. In cells, the anchoring may be achieved by attachment of the DNA to the nuclear skeleton or other cellular structures, or excess hydrodynamic drag that prevents the whole complex from turning about its horizontal axis.

The simplest case (Fig. 1.5a) is the typical situation of a protein which binds to some specific short (8-20 bp) DNA sequence. Because of the defined shape of the protein which must interact with the DNA bases, the twist of the bound region is constrained. A rather extreme example of this situation is the case of DNA bound to the octamer of histone proteins in the nucleosome, where 146 bp of DNA are twist-constrained (18). In this case, one can imagine that an applied torque has to be quite large to change the equilibrium in favor of the protein being dissociated, since the free energy of binding of such proteins is typically 10 to 20 k_BT. On the other hand, one might imagine the DNA to respond by partially unbinding from the proteins, possibly leading to gradual transfer of twist *through* a protein-DNA complex (52).

In Fig. 1.5b, we show a variation on the first case, where a DNA loop is formed by interaction of one protein (or protein complex) with two (or more) DNA sites. There are many examples known where loops of thousands of bases of DNA are stabilized in this manner, for example the bacterial lac-repressor protein complex, and DNA loop-protein complexes formed by other

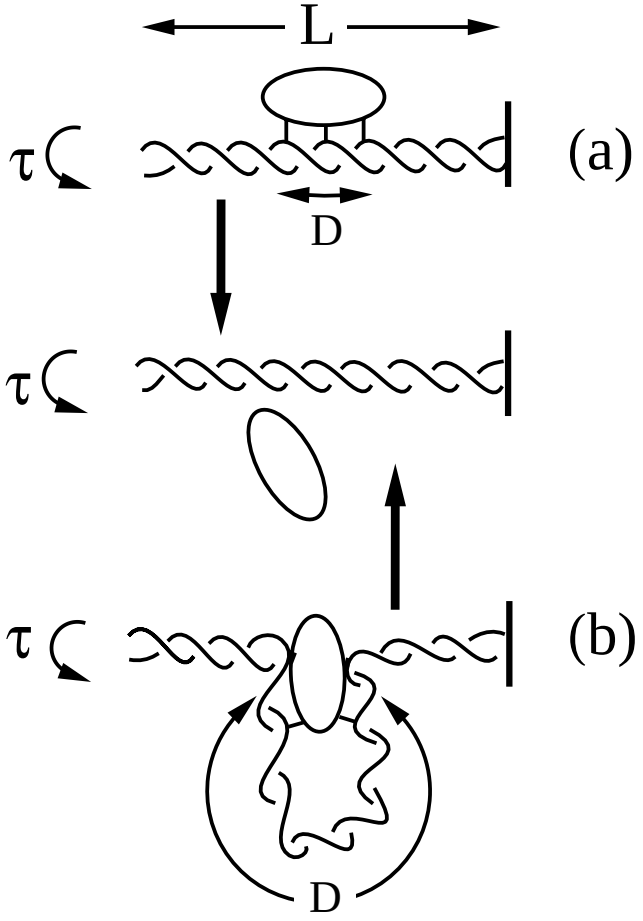


Figure 1.5. Ejection of DNA-bound protein (oval) by applied torque τ for different DNA-protein binding geometries. (a) Length D of DNA is used up when protein binds; the remaining length L-D must absorb the imposed twist strain since D has its twist fixed to 10.5 bp/turn. For $\tau < \tau^*$, inability of D bp to absorb some of the excess twist is compensated by the protein binding free energy μ. But for $\tau \geq \tau^*$ the protein-DNA complex dissociates since ΔF can be lowered by spreading out the extra twist strain over the remaining length D. In nucleosomes, D is the length of DNA wrapped around the histone core. (b) DNA-bound protein captures length D of DNA in a loop. Since the length of the loop is much bigger (typically 1 kb) than the protein-DNA interaction region, we expect a lower τ^* than above. The lac repressor bound to DNA is an example. The heavy arrows in (a) and (b) signify that the dissociated state of the DNA-protein complex is more favored when $\tau \geq \tau^*$.

transcription factors (53). DNA loops are also hypothesized to be stabilized by chromosome-folding protein complexes such as SMC elements (54). In all these cases, one can imagine that the binding of the DNA loop could be controlled by either static or transient torsional stress. In the case of twisting generated by RNA polymerase, this would provide an example of transcription-generated chromosome 'remodeling', and could be part of regulation of genes near to the transcription site (52).

Another situation close to that shown in Fig. 1.5b is the binding of a dsDNA to a surface; twisting DNA on either side of this binding site could then drive unbinding. If the binding surface is fixed in space (e.g. the nuclear envelope or some other large cell structure) the far end of the DNA need not be constrained in order for the twisting to drive unbinding of the DNA.

Large torques can be produced by a number of sources in living cells. In our work we have considered the case of DNA gyrase and RNA polymerase (RNApol) which are both proteins which act on DNA and introduce large twist strains in the process of catalyzing different reactions. Since these torque sources only produce transient torsional strain pulses, we need to understand how these strains relax along the DNA molecule. We expect that as the pulse propagates out from the initially affected region, it will generate torque at the different places it visits. We are interested in how far a typical twist pulse propagates and still produces enough torque to disturb the protein-binding equilibrium.

How does the excess twist decay away? Let $\theta(s,t)$ be the *excess* angle by which one base-pair stack is rotated with respect to its neighbor. Again considering DNA as a rigid rod, the

net torque on a cross-section Δs thick, in a medium of viscosity η, is the difference between the elastic torque, $k_BTC\partial^2\theta/\partial^2 s\Delta s$, restoring $\theta(s,t)$ to 0, and the frictional torque, $\eta\partial\theta/\partial t\Delta s$, opposing the rotational motion of the segment in the medium. Ignoring inertial effects, the equation of motion for $\theta(s,t)$ is $D\partial^2\theta/\partial^2 s = \partial\theta/\partial t$, where $D = Ck_Bt/\eta$. Using this equation, we find the typical distance, R_{max}, to which transient torques can propagate and still have enough strength left to produce torques that can affect protein binding. We find $R_{max} \approx 25$ nm for RNA polymerase and $R_{max} \approx 150$ nm for DNA gyrase, producing a torque $\approx 2k_BT$.

In Chapter 3, we examine the interplay between twisting of DNA generated by active DNA-twisting enzymes, and the stability of proteins which in their binding to DNA, constrain DNA twist. Our aim will be to estimate conditions under which twist-constraining proteins bound to DNA can be removed by torque. Our general approach will be to simplify the treatment of the DNA conformations, and to largely ignore the partition of linking number into twisting and writhing (43). Instead we will consider relatively short DNA segments where DNA twist and linking number will be treated essentially interchangeably. Besides torque-induced protein ejection from DNA, the non-zero twist modulus of DNA leads to many new phenomena including the supercoiling of DNA under torsional stress (43; 46), twist dynamics (47; 48), and twist transport (49; 50).

1.7 Unzipping DNA With Sequence Inhomogeneities and Bound Proteins

In cells, the separation of DNA strands is a ubiquitous phenomenon. For example, the two strands must be completely separated during DNA replication, and partially separated during DNA transcription. In cells, the separation of DNA strands occurs via forces applied

by DNA-processing machinery. Force-driven dsDNA 'unzipping' is therefore of direct biological relevance.

The two strands of the DNA double helix can be 'unzipped' by application of $\approx$ 15 pN force. Fig. 1.6 is a cartoon of this process showing the production of single-stranded DNA (ssDNA) with the application of equal and opposite forces at one end of the DNA. Since inhomogeneities in sequence generate inhomogeneities in strand binding energies, force variations occur, correlated with DNA sequence. A few groups have carried out single-molecule studies of DNA unzipping by force. Essevaz-Roulet *et al* (70; 71; 72; 73; 74) and Rief *et al* (16) have studied the $\approx$ 10 pN forces encountered during unzipping of 50 kb λ DNA, and that group has been able to resolve the variation in force with sequence. GC-rich regions were observed to require higher force for their unzipping than AT-rich regions, in accord with the generally higher base-pairing free energy associated with GC-rich sequences.

In Chapter 4, we theoretically analyze the force signal expected during unzipping through a single DNA region which is more strongly bound than the surrounding DNA. Our calculations are done for the situation where the extension between the unzipped molecule ends are fixed, and where force is measured. Two different types of force traces are obtained depending on the inhomogeneity length and strength. For short inhomogeneities, a 'sawtooth' force trace is obtained, with a force jump corresponding to release of the entire inhomogeneity at once. For longer inhomogeneities, traces with force plateaus are obtained, corresponding to gradual unpeeling of the strongly bound region. At force jump points, there are energy barriers that must be crossed via thermally activated kinetics. We also analyze the related situation where a

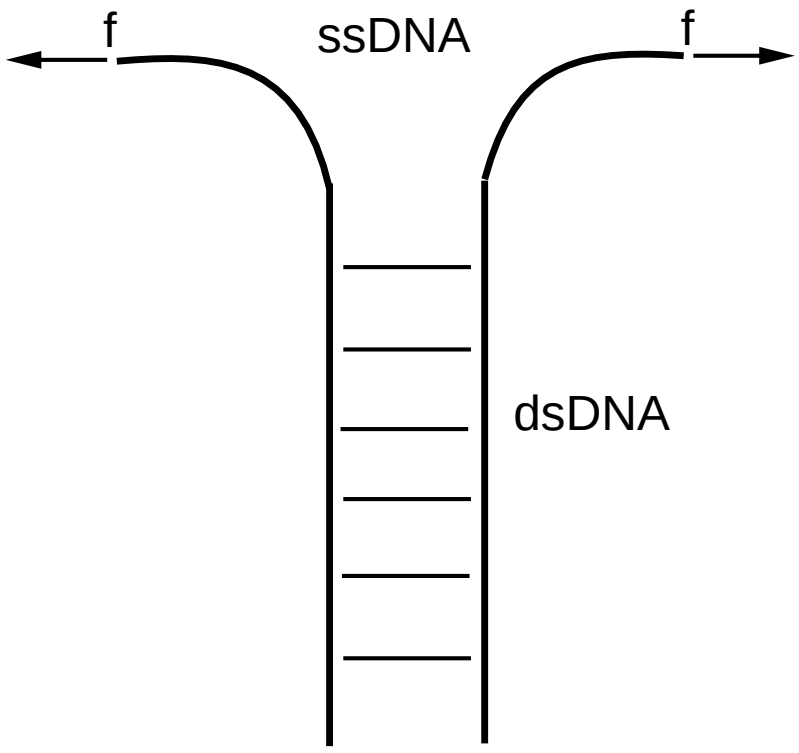


Figure 1.6. Forcing apart the two strands of the DNA double helix by applying equal and opposite forces to one end of the molecule. By applying $\approx$ 15 pN of force, the DNA can be unzipped. In the simple case sketched in this figure, we ignore sequence inhomogeneities and other obstructions to uniformly unzipping DNA. The two single-stranded DNAs (ssDNA) that result from unzipping are very floppy, having very small persistence length.

protein is bound at a specific point on the DNA, and where its removal is required for unzipping to proceed past it. Similar force signals are observed in this case, with the force jumps now corresponding to dissociation of the protein.

In focusing on the how sequence-dependence and bound proteins may affect the strand-separation process, we do not develop a detailed microscopic theory of the inter-nucleotide interactions that hold the two polynucleotide strands together (for one such attempt, see (29)). Instead, we attempt to model, using a simple theory, two scenarios that in the immediate future may be accessible to direct laboratory experimentation.

CHAPTER 2

PULLING AND TWISTING DNA

2.1 Introduction

In this chapter, we construct a simple statistical mechanical model of mechanically stressed double stranded DNA with its twist fixed externally. We exactly evaluate its partition function by numerically diagonalizing the model's transfer matrix. We determine the parameters of our model by a fit of the theoretical force-extension curves to those of experiment (23). The plan of the chapter is as follows. In Section 2.2 we discuss the experimental results and the DNA structures that have been proposed to explain them. In Section 2.3 we present our model. In Section 2.4, we show the results of the theory in comparison with experiment, and we discuss its implications. We conclude in Section 2.5 with a summary of our principal findings and discussion of other proposals for explaining the DNA force-distance curves that we have analyzed.

The material in this chapter has been published in Phys. Rev. Lett. **83**, 1066 (1999) (23) and Phys. Rev. E **63**, 51903 (2001) (61).

2.2 Experimental Data on DNA Overstretching

2.2.1 Freely Untwisting Molecule

In micromanipulation experiments, DNA linkage will be free to change if single-strand attachments are made to the molecule (Fig. 1.3b), or if there is a break (a 'nick') in either of

the two sugar-phosphate backbones. The elastic response of nicked DNA has four regimes (Fig. 2.1a): first, thermal bending fluctuations are removed by application of small forces (< 10 pN). By 10 pN the molecule is completely extended to its full B-form contour length, ≈ 16.3 μm for λ-DNA (31; 32; 33; 34; 35). The force needed to stretch the molecule further rises linearly with extension up to ≈ 65 pN, and a linear stretching elastic constant can be roughly estimated from the slope ($220 k_B T/\text{nm} = 900$ pN). Then, at 65 pN, the molecule extends abruptly over a narrow range of force (≈ 5 pN), from about 1.1 times to about 1.7 times its unstressed B-form contour length (12; 13). The linear relationship between force and extension resumes after the transition but with a larger elastic constant (12; 13).

The behavior at ≈ 65 pN has been interpreted as a cooperative transition from one well-defined structural state of DNA (B-form DNA at low forces) to another (S-form DNA at high forces) (12). The precise structure of S-DNA is unknown. Although a ladder-like form has been proposed (13; 36), S-DNA is less than the 2 times B-form length expected for complete unwinding of the double helix. In comparison with experiments on isolated single-stranded DNA, S-DNA is seen to have a different force-extension behavior (13), suggesting that the two strands of an S-DNA are strongly interacting.

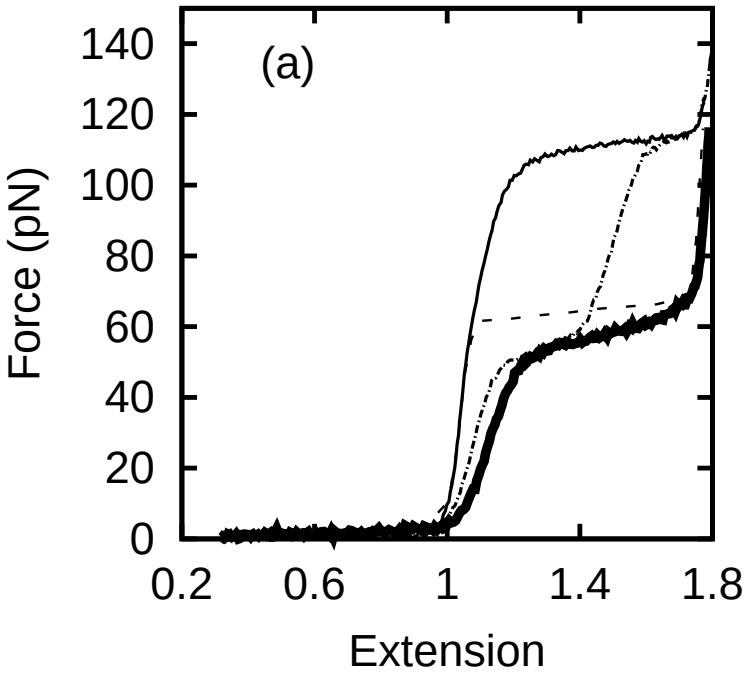
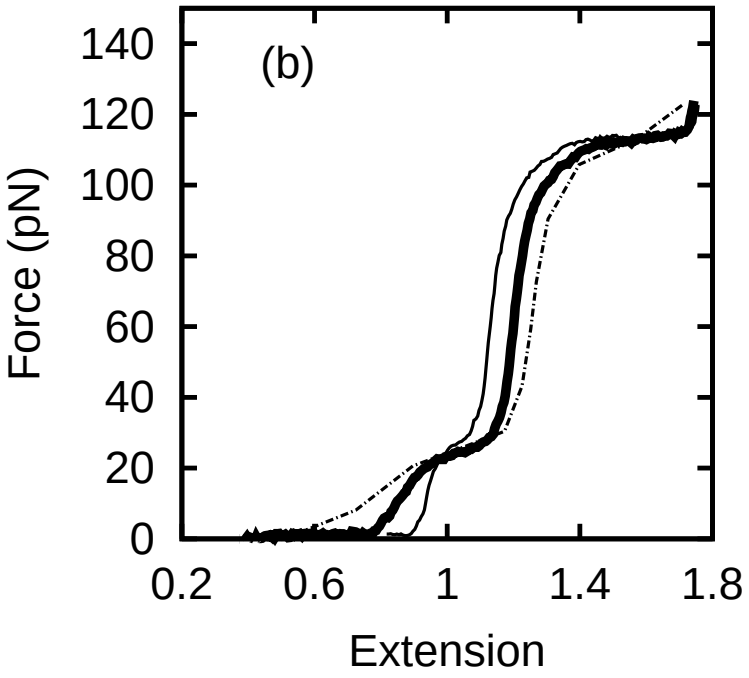


Figure 2.1. Force versus extension for DNA with different fixed negative and positive σ. Fig. 2.1a shows data for Nicked DNA (dashed), $\sigma = 0.00$ (solid thin), -0.357 (dot-dashed), and -0.714 (solid thick). Note that $\sigma = -0.714$ curve coincides with the nicked DNA curve at the end of the 65 pN plateau. This is used to extract the helicity of S-DNA. Fig. 2.1b (see next page) shows the elastic response of overwound DNA with $\sigma = 0.357$ (solid thin), 0.714 (solid thick) and 1.071 (dot-dashed).


(b)
Force (pN)
Extension
140
120
100
80
60
40
20
0
0.2
0.6
1
1.4
1.8

2.2.2 Fixed Linking Number

For dsDNA with both strands affixed to surfaces (Fig. 1.3c), the linking number Lk of the two strands is a topological invariant provided the surfaces are fixed in space. Lk is the sum of the twist (Tw, the number of times one sugar-phosphate backbone wraps around the other) and the writhe (Wr, the average over all projections of the number of non-local self-crossings of the double helix) (39; 40):

$$Lk = Tw + Wr \tag{2.1}$$

Experiments which fix DNA topology are usually done with initially relaxed molecules, with Wr near 0. For a long straight and unstressed DNA, $Lk = Lk_0$, the number of turns of the double helix. Since there is one turn every 10.5 bp for unstressed DNA, $Lk_0 \approx \mathrm{N_{bp}}/10.5$, where $\mathrm{N_{bp}}$ is the total number of base pairs. This implies for λ-DNA a Lk_0 of ≈ 4200. It is convenient to describe the linkage by the fractions of turns added or removed:

$$\sigma = \frac{Lk - Lk_0}{Lk_0}. \tag{2.2}$$

For either a ladder-like DNA structure, or for separated single strands, $\sigma = -1$. By comparison, DNA in vivo is usually constrained to have $\sigma \approx -0.05$.

Fig. 2.1a shows the force-extension response of torsionally constrained DNA with $\sigma = 0$, i.e. $Lk = Lk_0$. A force plateau is observed at 110 pN, nearly double the force of the B-S transition on a nicked molecule. The transition with $\sigma = 0$ is also somewhat more gradual and 'rounded' than the nicked molecule transition. These data already make clear that S-DNA

is not simply denatured DNA, and that the fixed linkage constraint introduces another DNA state characterized by a higher free energy.

For $\sigma < 0$, a new force plateau appears in Fig 2.1a at ≈ 50 pN. As the molecule is progressively underwound (as σ is made more negative) the width of the lower 50 pN plateau increases at the expense of the upper 110 pN transition. The 110 pN transition completely disappears at $\sigma \approx -0.7$. The $\sigma = -0.714$ curve coincides with the nicked DNA curve at the end of the 65 pN plateau. This indicates that 'pure' S-DNA has $\sigma \approx -0.7$. A progressive softening of the DNA with underwinding is also apparent in the data (23).

Our explanation of the $\sigma = 0$ transition is that the linking number constraint obstructs the production of 'pure' S-DNA at 65 pN. Instead, a mixture of two phases, S-form and some other form of dsDNA, must arise to satisfy the twist constraint. Work of the group of Allemand *et al* (30) has shown that dsDNA can organize into a highly overtwisted 'P'-form with a 2.4 bp helix repeat, or $\sigma = +3.375$. It is possible that P-DNA is created along with S in suitable proportion to produce the necessary total linking number, and that the 110 pN force is characteristic of the transition to the P-state. Given the twist of P measured by Allemand *et al* and the S-DNA twist of $\sigma \approx -0.7$, for $\sigma = 0$ this mixture should be about 4/5 S to 1/5 P. This is in accord with the observed overextension of 1.66 times the native λ-DNA length at 110 pN for $\sigma = 0$, the appropriately weighed average of of S- and P-form extensions (23).

The appearance of the second transition plateau at ≈ 50 pN and the progressive narrowing of the 110 pN plateau for $\sigma < 0$ can be fit into this picture as follows. Initially, unwinding is stored in an untwisted double helix. Once a force of ≈ 50 pN is applied, the untwisted part of

the molecule transforms to S-form and the remainder of the molecule is still B-DNA with $\sigma = 0$ and will display a transition at ≈ 110 pN. Further unwinding narrows the 110 pN plateau while widening the one at 50 pN since a progressively larger fraction of the molecule can transform to S-DNA at 50 pN. This picture of the $\sigma < 0$ transitions means that the σ for which the 110 pN completely disappears for the first time is a direct measurement of the helicity of S-DNA. This happens at $\sigma = -0.714$, implying a S-DNA helix pitch of 22 nm and 37.5 base pairs per turn.

Fig 2.1b shows data for force-extension response of overwound DNA for $\sigma = 0.357, 0.714, 1.071$ (23). A new force plateau near 25 pN occurs, which was first reported by Allemand *et al* (30). With increasing σ, the 25 pN transition widens while the 110 pN transition narrows. For higher σ's, the molecule stretching modulus shows considerable softening for forces < 20 pN.

The 25 pN transition for overwound DNA is thought to be a straightening of plectonemically supercoiled P-DNA(23; 30). Overwinding the molecule from $\sigma = 0$ to $\sigma = 1.071$, the molecule fraction which must transform to unwrithed P-DNA increases, widening the 25 pN plateau (from initially zero width) while narrowing the 110 pN plateau.

The σ for which the 110 pN transition completely disappears corresponds to the point where the entire molecule is able to transform to P-DNA at the 25 pN transition. In this way, Leger *et al* measured P-DNA to have a helix repeat of ≈ 2.62 bp/turn and an extension of 1.6 times the B-form length (23). This is close to the 2.4 bp/turn and the 1.75 extension estimates of Allemand *et al* (30).

The interpretation of the fixed-σ experiments raises a number of questions. For $\sigma < 0$ and forces less than 50 pN, is the molecule initially in a pure B-DNA structure, or is it coexisting with some other structure so that there can be *two* reasonably sharp transitions as force is increased? If B-DNA does coexist with some other structure below 50 pN, what does this structure look like? Since the 50 pN plateau is seen only for $\sigma < 0$ and not for $\sigma = 0$, we conclude that this phase must have non-zero helicity. From inspection of the experimental data we also expect the phase to have a contour length similar to B-DNA.

2.3 Model

We have constructed a simple 1d statistical mechanical lattice model of DNA to test the proposed transition mechanisms and to explore in detail questions such as those raised in the previous paragraph. Our model is a generalization of a previous model proposed in Cluzel *et al* (12) to explain the B-form to S-form transition. We modify it to take into account the fixed twist constraint and the local harmonic fluctuations of twist density and extension about their equilibrium values. This model is also a discretized version of a continuum model presented in (23) to explain the fixed twist experiments. Our discrete model leads to much faster numerical calculations and lets us explore the phase diagram of DNA more thoroughly.

As might be expected, there are quite large differences in the force-driven structural transitions displayed by synthetic pure AT and pure GC molecules(15; 16). Sequence effects can in principle be added to our calculations, but the available data are far from adequate to fix the many parameters required. Given this state of affairs, plus the observation of little change

of the B-S transition when different natural-sequence molecules are used (20), we consider a homogeneous model in this chapter.

2.3.1 Definition of the Model

To understand the complicated plateau curves of Fig. 2.1 and their evolution with σ, we treat them as being due to transitions between five microscopically distinct structural states. Which states participate in a transition depends on σ. The micromanipulation experiments control only extension and linkage number, and therefore our model specifies only the microscopic extensions and linkage numbers of the five states. Further details of the secondary structures of the various DNA states cannot be directly inferred from the available experimental data, and therefore we do not attempt to predict them.

The five states that we consider include B-DNA which is stable at zero force and zero torque, and the overstretched S- and P-DNA states, which are underwound and overwound, respectively. This leaves two more states, which we have found to be essential in producing the two-plateau force-distance curves found in the intermediate-σ parts of Fig. 2.1. To fit the data, one of these states must be highly overwound and highly contracted, implying the supercoiled P state (sc-P-DNA) proposed by Allemand *et al* (30). The final state must be more underwound than the S-state, but extended to a length comparable to B-DNA. Since our fitting suggests a highly untwisted double helix, we call this state 'Z'-DNA in analogy with an actual left-handed double helix structure observed for certain sequences and chemical conditions(37).

DNA is modeled as a 1d lattice with each lattice site degree of freedom, n_i, (with the subscript i labelling the lattice sites) having one of five possible integer values from 0 to 4

representing 5 different microscopic structures (0: B-DNA, 1: S-DNA, 2: P-DNA, 3: 'Z'-DNA and 4: sc-P-DNA).

Each structure has a characteristic dimensionless linkage number density, $\overline{\theta}_n$, which measures the excess linkage density in radians per base pair of one of the DNA structural states relative to B-form DNA, and extension $\overline{s}_n$ (fractional change in contour length relative to B-DNA) about which we allow harmonic fluctuations. By comparision, for completely untwisted, i.e. ladder-like, DNA, $\overline{\theta} = -0.62$. The free energy cost relative to B-form DNA associated with creation of the stressed DNA states is taken into account with five more parameters ϵ_n. By definition, $\overline{\theta}_0 = 0$, $\overline{s}_0 = 0$ and $\epsilon_0 = 0$.

The Hamiltonian for our model is:

$$\frac{H([n_i])}{k_B T} = \sum_i \left[\frac{C'}{2}(\theta_{n_i} - \overline{\theta}_{n_i})^2 + \frac{K'}{2}(s_{n_i} - \overline{s}_{n_i})^2 + J(1 - \delta_{n_i,n_{i+1}}) + \epsilon_{n_i} - f s_{n_i} - \tau \theta_{n_i} \right]. \quad (2.3)$$

C' in our Hamiltonian is the usual twist persistence length C (between 75 nm and 110 nm; see below) divided by 0.34 nm to give a dimensionless harmonic twist fluctuation energy. Similarly, K' is the extensional elastic constant K (= 300 nm^{-1}; see below) times 0.34 nm. The force f and torque τ are similarly reduced by factors of $k_B T$ and 0.34 nm so as to be dimensionless.

The quadratic terms in θ_n and s_n simply account for small deviations of linkage density and extension from their usual values, $\overline{\theta}_n$ and $\overline{s}_n$. The "domain wall" term, $J(1-\delta_{n_i,n_{i+1}})$, states the molecule's preference for structural uniformity, or equivalently the cooperativity of the various

structural transitions. The force and torque couplings transmit the externally imposed stresses to the molecule.

At zero force and torque, the parameters of the model should be chosen so that state 0, B-DNA, is the lowest-energy state. The basic idea of the model is that for sufficiently large forces or torques, the other states can become lower in energy, causing first-order-like structural transitions. Of course a one-dimensional model will not produce truly discontinuous transitions, but instead will produce smoothed transitions similar to those observed experimentally (Fig. 2.1).

Our model does not include low force (< 4 pN) entropic elasticity of DNA, which is not important at the high forces we are considering. The "softening" of the DNA extensional modulus in the linear regime < 50 pN for $\sigma < 0$ is also not captured by this model. These effects will be further discussed in Section 2.5.

2.3.2 Transfer Matrix Calculation

We begin by integrating out the local harmonic fluctuations in θ_n and s_n. This leads to additive contributions to the free energy and $\langle\theta\rangle$ and $\langle s\rangle$ are

$$\langle\theta_{n_i}\rangle = \overline{\theta}_{n_i} + \frac{\tau}{C'} \tag{2.4}$$

$$\langle s_{n_i}\rangle = \overline{n}_i + \frac{f}{K'} \tag{2.5}$$

Calculation of the partition function is now reduced to a transfer matrix diagonalization problem by rewriting it as

$$= \sum_{\{n_1\},\ldots,\{n_N\}} \prod_i T(n_i, n_{i+1}). \tag{2.6}$$

where

$$T(n_i, n_{i+1}) = \exp\left[-\beta H_{symm}\left(n_i, n_{i+1}\right)\right]. \tag{2.7}$$

H_{symm} is the symmetric nearest-neighbor Hamiltonian.

For a one dimensional system with site-independent couplings and periodic boundary condition on the transfer matrix ($n_{N+1} \equiv n_1$), Eq. (2.6) becomes

$$Z = \mathrm{Tr}(T^N) = \sum_n \lambda_n^N \tag{2.8}$$

where the λ_n's are the eigenvalues of the transfer matrix. The free energy $F = -\beta^{-1} \ln(\sum_n (\lambda_n)^N)$. In the thermodynamic limit, only the largest eigenvalue λ_{max} contributes to the sum so we have

$$F = -N\beta^{-1} \ln(\lambda_{max}). \tag{2.9}$$

The choice of boundary condition does not affect this result.

For our model, the transfer matrix is independent of which pairs of neighboring lattice sites we look at since we have homogeneous couplings. Also, the elastic fluctuations in s_{n_i} and θ_{n_i}

have been integrated out so they do not enter the model's transfer matrix. The transfer matrix is given by

$$T(n_i, n_{i+1}) = \exp\left[J - \delta_{n_i,n_{i+1}} - \frac{1}{2}(\epsilon_{n_i} + \epsilon_{n_{i+1}} - f(\overline{s}_{n_i} + \overline{s}_{n_{i+1}}) - \tau(\overline{\theta}_{n_i} + \overline{\theta}_{n_{i+1}}))\right] \tag{2.10}$$

The average value of an operator dependent on adjacent sites, $< O >$, can also be calculated by this method:

$$\langle O \rangle = \frac{\mathrm{Tr}(SOS^{-1}T_d^N)}{Z} \tag{2.11}$$

where T_d is the diagonalized transfer matrix and S is the diagonalizing matrix. Two-point correlation functions can be similarly calculated using

$$\langle O_i O_{i+L} \rangle = \frac{\mathrm{Tr}(SO_iS^{-1}T_d^L SO_{i+L}S^{-1}T_d^{N-L})}{Z} - \langle O_i \rangle \langle O_{i+L} \rangle \tag{2.12}$$

The model is solved exactly by numerically diagonalizing the transfer matrix. This allows us to compute directly statistical averages for given values of force and torque. However, to contact the experimental results of Fig. 2.1 we also need to compute the torque necessary to make $\langle \sigma \rangle = \sigma_{\mathrm{ext}}$. This must be done for each force, and for each σ_{ext} of interest.

2.3.3 Choice of Values of Fitting Parameters

In all eighteen parameters need to be specified in our Hamiltonian to obtain the theoretical force-extension curves. These are: C' (or C), K' (or K) , J, ϵ_n, $\overline{\theta}_n$, and $\overline{s}_n$, with n going from 0 to 4. However, many of these parameters are determined by previous experiments.

The stretch modulus K is known to be roughly 300 nm^{-1} (41; 42), giving $K' = 102$. C is thought to be between 75 and 100 nm. Moroz and Nelson (25) have fit experimental data to obtain $C = 110$ nm, while Bouchiat and Mezard find $C \approx 70$ nm (26) and a value $C = 75$ nm has been determined from studies of linkage number fluctuations in circular DNAs (43). We use a C' of 220 which corresponds to $C = 75$ nm. Finally, the parameters for the B-DNA state are defined as $\overline{s}_0 = 0$, $\overline{\theta}_0 = 0$ and $\epsilon_0 = 0$.

S-DNA is known to be 1.7 times longer than B-DNA (12) and from Leger *et al* (23), it is known to have a helicity (relative to B-DNA) of -1.33 nm^{-1} (37.5 base pairs per turn). This fixes $\overline{s}_1 = 0.7$ and $\overline{\theta}_1 = -0.45$. Next, we use the free-twist experimental data to fix the S-DNA free energy at $\epsilon_1 = 3.7$ and the transition cooperativity at $J = 2.0$, in precisely the manner used by Cluzel *et al* in their model of the B-S transition(12).

The P-DNA state was previously studied by two groups (30; 23), and we take values for the stretch and linkage number which are between the values obtained in those experiments, $\overline{s}_2 = 0.6$ and $\overline{\theta}_2 = 1.87$. Since the two experiments agree to about 10% on these two parameters, we are highly constrained as to their choice.

The parameters for the related sc-P-DNA state are also essentially determined by experiment (30; 23). Since supercoiled P-DNA will form plectonemic coils, its length will be essentially zero, i.e. $\overline{s}_4 = -1$; on the other hand its linkage number will be close to that of P; we take $\overline{\theta}_4 = 1.87$.

Finally, we have the 'Z'-DNA state, which from the experimental data is qualitatively understood to not be highly extended, and to be underwound. The 'Z' stretch and linkage parameters must be determined by fitting of our model to experiment, and we determine $\overline{s}_3 = 0.13$ and

$\overline{\theta}_3 = -1.30$. This means that 'Z'-DNA has a left-handed twist. It is the only left-handed state in our model. Overall, the structural parameters for the five states are compatible to those used previously (23).

The parameters which are truly free for fitting are the free energies of the P, sc-P and 'Z' states. These are determined essentially by the plateau forces for the fixed-σ transitions. We find $\epsilon_3 = 2.3$ ('Z') , $\epsilon_2 = 17.0$ (P), and $\epsilon_4 = 13.5$ (sc-P). For S-, P-, and 'Z'-DNA the free energies are similar (to within 30% in the worst case) to the corresponding parameters in our previous work (23). In the present chapter the free energy of sc-P-DNA differs substantially from what was used in Leger *et al*, being quite close to the free energy of P-DNA. The different free energy parameters found here are a consequence of differences between the two models, our earlier one being a continuum model vs. the present one being discrete. In our earlier work there were additional free parameters for twist and stretch stiffnesses of each state, which make varying contributions to the free energies of the five states. In our present work, all states have the same stretch and twist stiffness, which make the relative free energies more closely related to the free energies of the five states.

2.4 Results and Discussion

2.4.1 DNA With Unconstrained Linking Number

When we allow DNA linking number to fluctuate freely with zero applied torque, our model reproduces the usual B-form to S-form transition seen in experiments on nicked or single-strand-attached molecules. Fig. 2.2a superposes the experimental data on our calculated force-extension curve. The length fraction, $\langle \delta_{n_i,m} \rangle$, of each of the states (m = 0,1,2,3,4) is shown

adjacent to the force extension curve as a function of force. Near 65 pN, we see B-DNA give way to S-DNA, resulting in a pure phase of S-DNA after $\approx$ 70 pN. The length fractions of the other phase are extremely close to zero due to their much higher free energies.

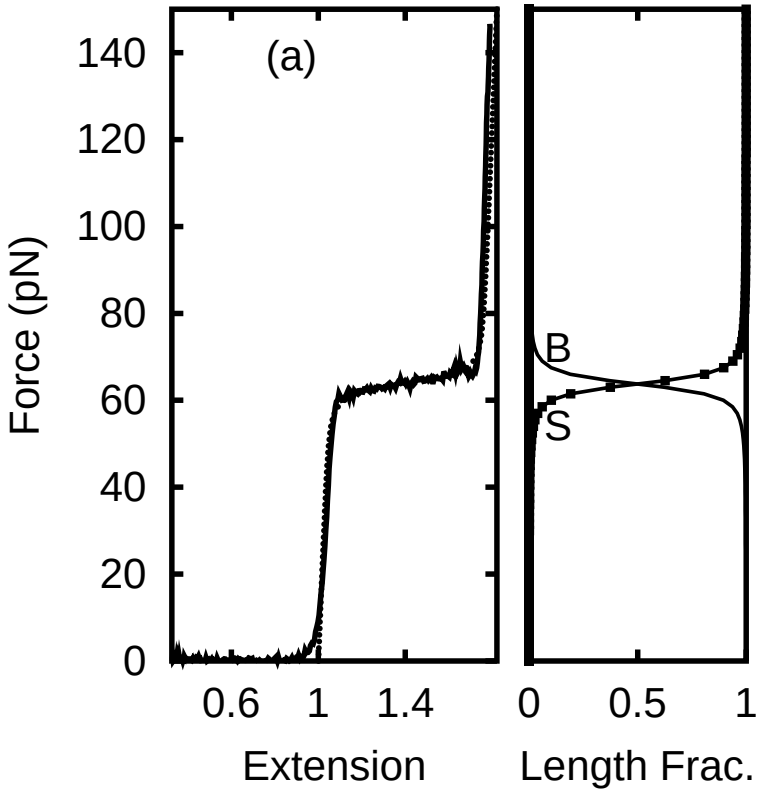
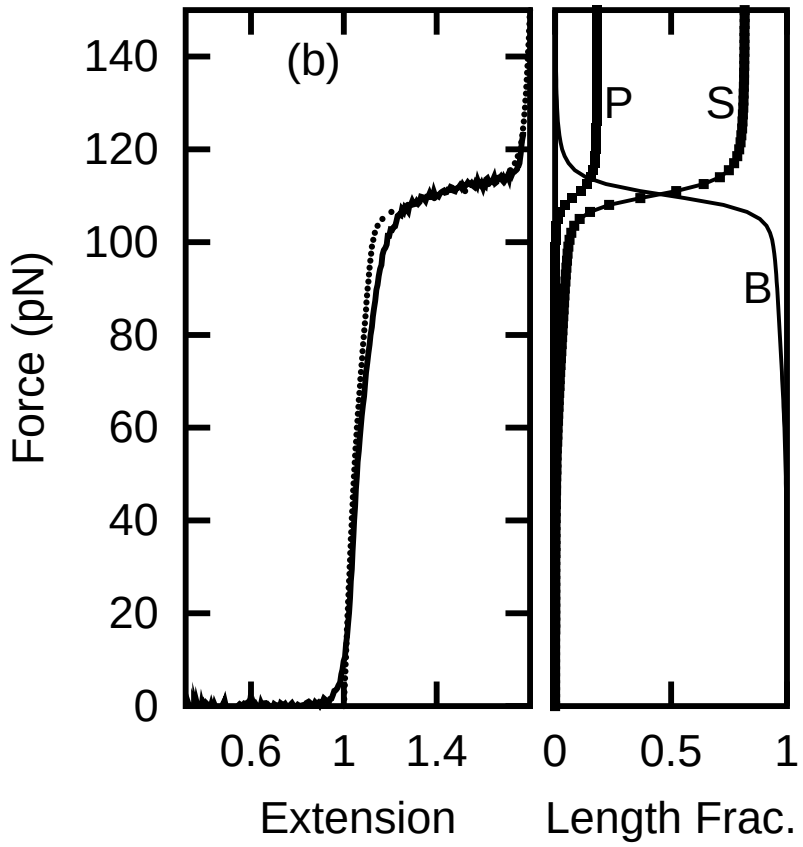
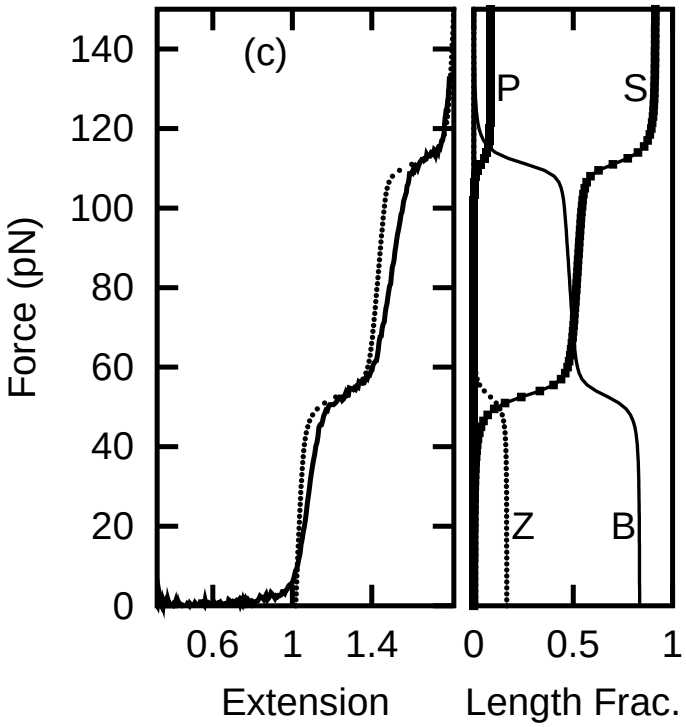
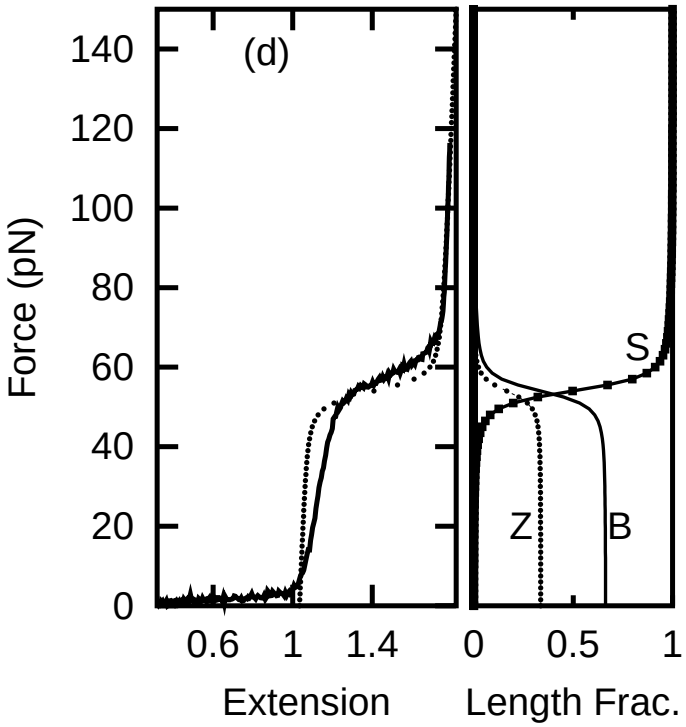
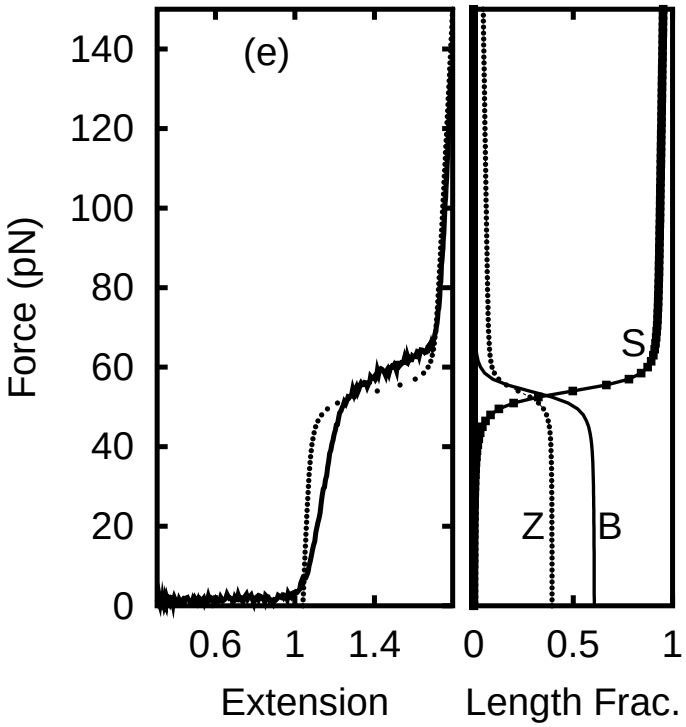


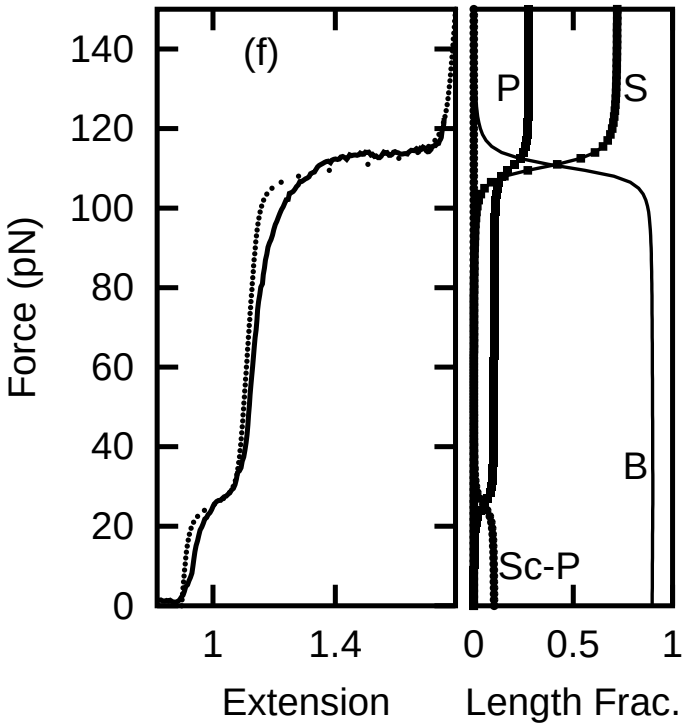
Figure 2.2. Experimental force versus extension data for positive and negative σ and nicked DNA superposed on the corresponding theory curves together with the plots of force versus length fraction ($< \delta_{n_i,n_{i+1}} >$) for fixed σ. In all force-extension curves, the solid black line is the experimental data and the solid dot curve the theoretical fits. Fig 2.2a shows computed (solid dots) and experimentally determined (solid black) force-extension data for nicked DNA. The adjacent B and S length fractions intersect at the transition force. Other phases are absent. Fig. 2.2b-e correspond to $\sigma = 0.00, -0.357, -0.714$, and -0.833 respectively. Fig. 2.2f-h correspond to $\sigma = 0.357, 0.714, 1.071$. In all these plots, $< \delta_{n_i,n_{i+1}} >$ is adjacent to the force-extension data. The fixed-topology constraint determines which structural states participate in the transition(s) for given σ. In all the plots, length fraction for B-DNA is shown by the solid line, for P- by successive squares, for S- by alternating solid squares and dashes, for 'Z'-DNA by the solid dot curve, and for sc-P by alternating solid dots separated by lines.

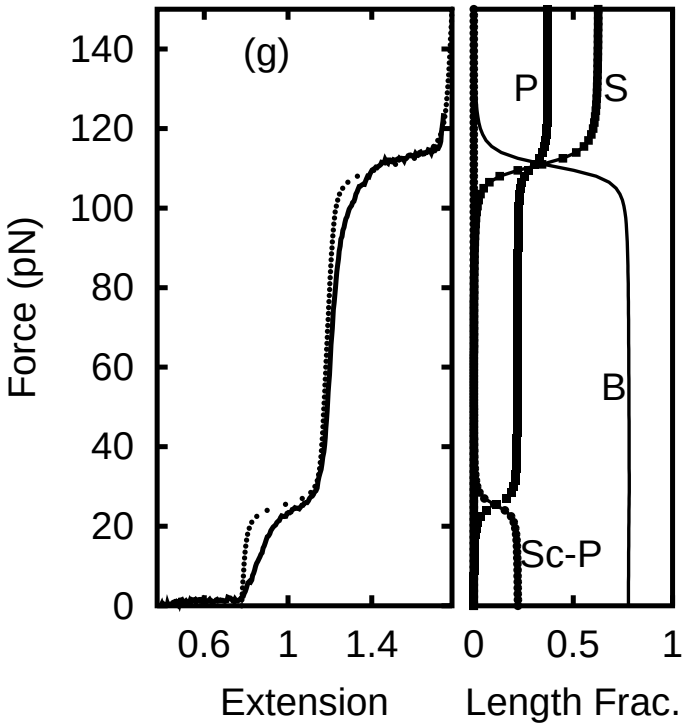

(b)
P
S
B
Force (pN)
140
120
100
80
60
40
20
0
0.6
1
1.4
0
0.5
1
Extension
Length Frac.

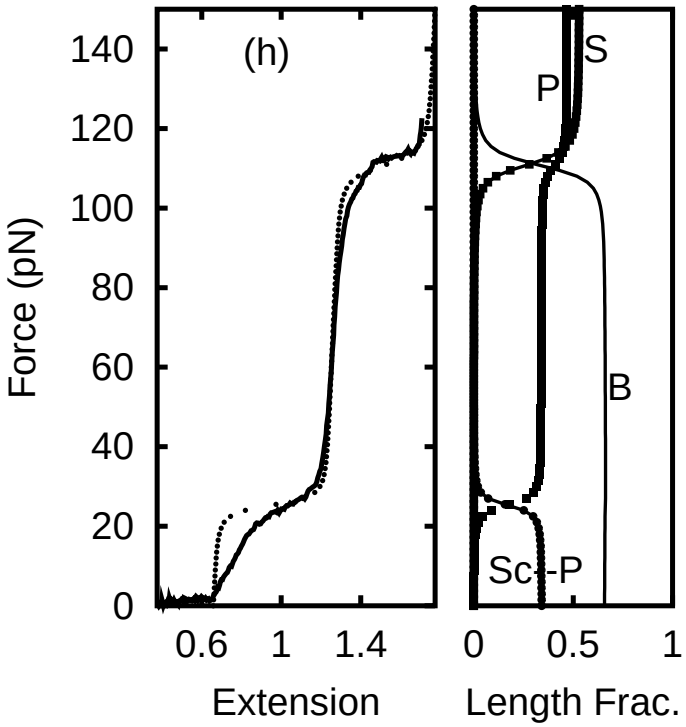

(c)
P
S
Z
B
Force (pN)
140
120
100
80
60
40
20
0
0.6
1
1.4
0
0.5
1
Extension
Length Frac.


(d)
140
120
100
80
60
40
20
0
Force (pN)
0.6
1
1.4
Extension
0
0.5
1
Length Frac.
S
Z
B


(e)
140
120
100
80
60
40
20
0
Force (pN)
0.6
1
1.4
Extension
0
0.5
1
Length Frac.
S
Z
B


(f)
P
S
B
Sc-P
Force (pN)
140
120
100
80
60
40
20
0
1
1.4
0
0.5
1
Extension
Length Frac.


(g)
Force (pN)
140
120
100
80
60
40
20
0
0.6
1
1.4
Extension
P
S
B
Sc-P
0
0.5
1
Length Frac.


(h)
Force (pN)
140
120
100
80
60
40
20
0
0.6
1
1.4
Extension
S
P
B
Sc-P
0
0.5
1
Length Frac.

2.4.2 DNA With Fixed $\sigma \leq 0$

Fig. 2.2b shows the theoretical force response curve superposed on the experimental data, for linking number $\sigma = 0$. Initially, the molecule starts off in pure B-form. Near 65 pN, an appreciable amount of S-DNA cannot form as in the unconstrained-σ case since this would either violate the twist constraint (S-DNA is undertwisted), or cost a large amount of twist energy. Thus only a small density of S-DNA 'bubbles' form as force is raised from 65 to 110 pN. At 110 pN, the force is large enough for the molecule to transform from B into a mixture of S- form and P-form DNA. Since P-form DNA is overtwisted, it can compensate for the creation of the undertwisted S-DNA. This is evident in the roughly 4:1 proportionality of the length fractions for S- and P-DNA for forces > 100 pN, which matches the proportionality of the S- and P- twisting.

The computed force-extension curves for underwound DNA for $\sigma = -0.357$, -0.714 and -0.833 are shown in Fig. 2.2c, 2.2d and 2.2e respectively, along with experimental data. For $\sigma = -0.357$ two transitions occur, one at ≈ 50 pN, followed by a second at ≈ 110 pN. The existence of two transitions requires the undertwisted 'Z'-DNA state to absorb the underwindings at low force. Without the 'Z'-state, we have found that S-DNA always appears immediately at low forces (without 'Z', S would be the only underwound state in the model). Then, there could be one transition, corresponding to replacement of the B fraction with P. An additional problem in this case is that the zero-force state has an extension which is too long to fit the experimental data; note that this is a consequence of the zero-torque results which fix the S-DNA structural parameters.

Instead, we have a mixture of 'Z-' and B-DNA at zero force in appropriate fractions to fix the net linking number at $\sigma = -0.357$. At ≈ 50 pN, the 'Z'-DNA portion and some of the B-DNA portion of the molecule transform to S-form, giving a B-S mixture for forces between 50 pN and 100 pN. The width of this first transition is determined by the relative fractions of B and S necessitated by the underwinding. The exchange of S-DNA for 'Z'-DNA is seen clearly in the computed length fractions for these phases.

At 110 pN, the B-DNA portion is converted to a mixture of P- and S-form DNA, in the same way as occurs in the $\sigma = 0$ case. At larger forces, the molecule is thus predominantly in S-form, with a small fraction of P-DNA as required to satisfy the linkage number constraint. At $\sigma = -0.357$, we start at low forces with 'Z'+B, and then in two steps we go to B+S and then to S+P. Without the 'Z'-state to start with, there would be no possibility for this sequence of two transitions.

As σ is reduced further, the fraction of the molecule in the 'Z'-DNA phase at low force progressively increases, and the fraction of the molecule which ends up as S-DNA after the 50 pN transition correspondingly increases. Thus, less and less B-DNA remains to participate in the 110 pN transition, and it becomes progressively narrower. Finally, for $\sigma = -0.714$, the 110 pN transition completely disappears indicating conversion of all the B+'Z' DNA below 50 pN to 'pure' S-DNA at 50 pN (Fig. 2.2d).

For even more underwinding (Fig. 2.2e), a large amount of 'Z'-DNA must initially be created. Then, in order to satisfy the twist constraint an appreciable fraction of 'Z' survives at high forces. This occurs because the 'Z' state is more underwound than is S.

A few discrepancies between theoretical and experimental curves are evident. The experimental data show a marked reduction in the low-force extensional modulus with underwinding. This effect is not in our model since we have only one stretch constant parameter for all five states; additional parameters might be able to take this effect into account. This difference in modulus causes the theoretical extension widths of the 50 pN B+'Z' $\rightarrow$ B+S transitions to be wider than those observed experimentally. Most likely, these differences are due to our oversimplification of the 'Z' state, which is clearly appreciably softer in both twist and stretch than the other DNA states. Finally, we have ignored low force (< 10 pN) entropic elasticity of B-DNA; our model of course does not reproduce the low force elastic regime where partially extended and plectonemically supercoiled B-DNA are found.

2.4.3 DNA With $\sigma > 0$

Fig. 2.2f-h show the computed elastic response of DNA for the overwound cases $\sigma = +0.357$, $+0.714$, and $+1.071$, overlaid on the experimental data. At zero force, the extra twists are absorbed in a collapsed supercoiled state, sc-P-DNA (30). This leads to coexistence of B- with sc-P-DNA at low forces. The experimental signature of this supercoiled state is the length reduction well below the B-form length. Near 25 pN, the supercoiled P-DNA extends to become P-DNA, giving the first transition. This is exactly analogous to similar transitions between plectonemically supercoiled B-DNA and extended B-DNA observed at forces ≈ 1 pN, although without the added complexity of appreciable conformational fluctuations(21; 24).

Through this first 25 pN transition the B-DNA fraction is almost unchanged. Then, the B region undergoes the 110 pN transition to a mixture of S- and P-DNA with the proportion

of DNA participating in these two states as is required by σ. The larger the number of extra twists inserted into the molecule, the greater the proportion of DNA that is in sc-P-form at zero force, and the greater the proportion of P-DNA at high force. Similarly, the extension width of the 25 pN transition broadens at the expense of the 110 pN transition as σ increases.

Again not all of the features of the transitions are captured by our simple model. As for $\sigma < 0$, the stretching modulus of the low-force (B+sc-P) drops as the molecule is overwound progressively more. This is not so surprising since the P-form is known to have its base-pairing disrupted (see Section 2.5). Overall, the two-stage transitions are well-described by our simple five-state model over a wide range of linkage number and force.

2.4.4 Global Phase Diagram of DNA Under Stress

The preceding sections propose a simple model which semi-quantitatively describes the structural transitions of a DNA driven by forces and torques. In fact, the transitions are experimentally sharp enough that it makes sense to talk about a 'phase diagram'. This is most simply defined in terms of pure 'phases' which are stable in different regions of the force-torque plane. We do not consider the modification of this phase diagram by changes in other parameters, e.g. temperature, pH, salt concentration, etc.

The thick lines of Fig. 2.3a show the boundaries between the five pure states as a function of force and torque. The lines themselves are the locus of points in the force-torque plane where there is no state which occurs with greater than 90% probability, indicating that the transitions are all rather sharp and first-order-like, with a typical width of 1 k_BT in torque, and 4 pN in force. The five phases also involve three triple-point-like regions where three phases can coexist.

For zero force and zero torque, we have B-DNA; then for $f < 10$ pN, a transition to 'Z'-DNA occurs for unwinding torque $\approx -2\ k_BT$, and a transition to sc-P occurs for overwinding with $\approx +7\ k_BT$ of torque. These estimates are in very good agreement with the results of the independent prior analysis performed by Strick *et al* (44) on force-extension data for fixed-twist molecules. Moreover, recent work, based on microscopic description of inter-nucleotide interactions, by Cocco *et al* (29), has also led to an independent estimate of this torque that agree well with ours.

For forces > 20 pN, sc-P DNA is replaced by P-DNA; the transition force between these two phases is essentially independent of torque since the P and sc-P states are nearly the same, except for having very different extensions. At a force ≈ 50 pN, S-DNA appears for the first time, for unwinding torques $\approx -3\ k_BT$. S-DNA remains stable at zero torque up to the highest forces that can be studied by micromanipulation, in agreement with experiments at zero torque which show only one transition, corresponding to B-S in our model.

Since all the transitions are relatively sharp, the state which is stable for given force and torque corresponds to that with lowest energy, $\Delta E_{n_i} = \epsilon_{n_i} - f\overline{s}_{n_i} - \tau\overline{\theta}_{n_i}$. The results of the 'zero-temperature' free energy minimization shows the same phase diagram topology as in the thermal phase diagram with zero temperature phase boundaries overlapping with the exact ($T \neq 0$) phase boundaries, from which we conclude that the true force-torque phase diagram essentially corresponds to this 'zero-temperature' estimate. Since force and torque appear linearly in the energy, the phase boundaries appear as a network of nearly straight line segments.

Fig. 2.3b gives a phase diagram in the force-linking number ensemble. Since σ is a conserved quantity, it displays phase separation, so we have a phase diagram with coexistence regions (solid black) between the pure states (open regions). The three triple points of Fig. 2.3a have turned into 'regions' with a continuum of forces and σ's for which the DNA is structurally in three pure phase coexistence. Finally, in each phase-diagram, the boundaries between pure phases correspond to force-τ or force-σ pairs for which the length fraction contributed by each of the five pure phases is less than 90%.

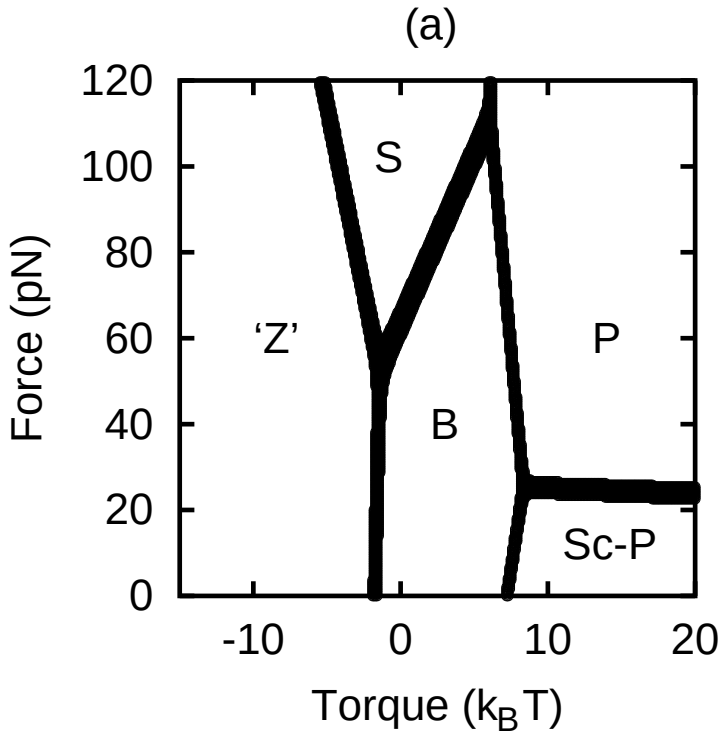
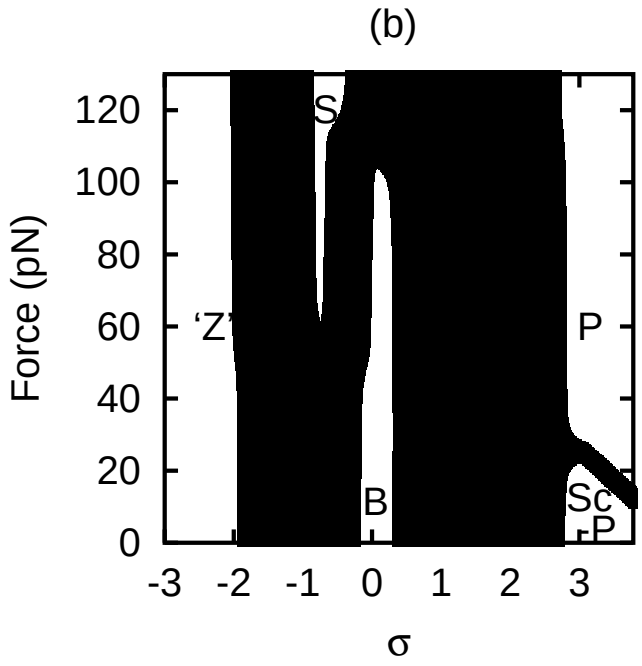


Figure 2.3. The force-torque (Fig. 2.3a) and force-σ (Fig. 2.3b) phase diagrams. In Fig. 4a, regions separated by the diffuse border "lines" correspond to force-torque pairs for which the equilibrium DNA state is a pure structural phase. Inside the boundaries, adjacent phases coexist. Intersection of three phase boundary lines corresponds to a region of three phase coexistence–a 'triple point.' In Fig. 2.3b, the phase boundary lines of Fig. 2.3a are now greatly thickened leaving only slivers of pure S-form and B-form phases. The force-σ pairs which fall inside the phase-boundaries correspond to to DNA structural states with coexistence of adjacent pure phases.


(b)
Force (pN)
120
100
80
60
40
20
0
-3
-2
-1
0
1
2
3
σ
S
'Z'
P
B
Sc
P

2.5 Conclusion

2.5.1 Five Structural States Occur in the DNA Force-Torque 'Phase Diagram'

Our main result follows from experimental data (23), which show four distinct force 'plateaus' which we take as the signatures of cooperative structural transitions to four non-B-DNA states. Although we have no direct microscopic information about these four new DNA states, the stretching-twisting experiments do allow determination of their mean extensions and linking numbers. For positive (right-handed) twisting, we find that one new state is extended and highly overtwisted (P-DNA), and another is its plectonemically supercoiled version (sc-P-DNA), which is similarly overtwisted but with zero extension. For negative (left-handed) twisting, a highly extended and undertwisted DNA state (S-DNA) occurs, plus an additional undertwisted state with extension only slightly greater than that of B-DNA ('Z'-DNA). All of our results are also consistent with other experiments of Allemand *et al* (30) which focused on transitions between B, P and sc-P states.

The physical picture of simple structural transitions between these five states, driven by forces and torques, is supported by the good global fit that we can make for a simple five-state model to the experimental force-extension curves. From our model we predict a force-torque 'phase diagram' in which the five structural states occur, separated by first-order-like transition lines. Given a micromanipulation experiment which can directly measure torque (e.g. by angular fluctuations of one end of the molecule), this phase diagram might be experimentally verified. All existing experimental data taken to date on dsDNA structural transitions have

been done in the ensemble of fixed linkage number, and an outstanding experimental problem is how to measure the torque applied in those experiments.

Furthermore, Clausen-Schaumann *et al* (15), while studying high force (up to 200 pN) elastic response of single λ-DNA molecules using AFM, have found that in 5% of cases, pulling DNA that is connected at one end to the substrate and at the other end to the AFM tip results in the elevated 100 pN transition, instead of the more usual 65 pN B-S transition. They interpret these force traces as resulting from stretching DNA whose 3' and 5' ends are both stuck to the substrate and which lack nicks, thus preventing rotation of the molecule about the helical axis. They observe that their data agree well with the results of Leger *et al* (23) and the theoretical description, using multiple-(structural)state model, contained therein and in (42).

2.5.2 'Z'-DNA State and DNA Strand Separation

An important result of our analysis is that we need two undertwisted states in order to produce the two-plateau structure of the force-distance curves obtained for underwound DNA, namely S-DNA and 'Z'-DNA. The latter state is only 10 % longer than the B-DNA double helix, but our best fit occurs when its helicity is roughly the reverse of that of B-DNA, i.e. a left-handed double helix with $\approx$ 12 bp per helix repeat. This state is distinct from S-DNA (S-DNA is about 70 % longer than B-DNA and has a right-handed helix with 38 bp repeat. We use the name 'Z'-DNA in imitation of the name given to the left-handed structure taken by some DNA sequences with similar length and helicity(37). We stress that our 'Z'-DNA occurs for undertwisted λ-DNA, and is not stable at zero applied torque. Our 'Z'-DNA is not obviously related to that of true Z-DNA.

It is possible that the torque-induced 'Z'-DNA is to some degree strand-separated DNA(38). However, this is difficult to determine from the current data since the micromanipulation experiments done to date do not directly probe details of DNA structure. We did try to fit the experiments with $\sigma = -1$, which corresponds to parallel separated strands. Our model indicates that this cannot provide a good fit to the experimental plateau, and we conclude that if the 'Z'-DNA state is in fact strand separated, those separated strands are then interwound to form a left-handed double helix with a 12 bp helix repeat and extension only slightly larger than that of B-DNA.

Finally, strand separation alone cannot explain the entire phase diagram since there are force-extension curves with two plateaus. There must be some differences in secondary structure between adjacent distinct phases, with accompanying energetic differences, to account for the observed transitions. We certainly do not mean to rule out the possibility of some degree of disruption of double helix secondary structure in the four non-B states. For example, the P and sc-P states have been proposed to have no hydrogen bonding on the basis of experiments showing them to be chemically exposed(30), and this is certainly not in any contradiction with our model. In fact, complete disruption of basepairing is required to explain the very short (2.4 bp) helix repeat inferred for P-DNA from experimental data (30; 23).

CHAPTER 3

TWISTING PROTEINS OFF DNA

3.1 Introduction

In this chapter, we take up the question of how DNA relieves excess twist that some outside agency, such as a mechanical twist-transducer or some enzyme, has imposed on it and what biological consequences arise from the twist relaxation process thus initiated. We start, in Section 3.2, by using a simple on-off model of protein binding to estimate the static torques necessary to drive-off a bound protein. Then, in Section 3.3, we first summarize relevant facts about two biologically important enzymes that are known to apply significant amounts of torque to DNA, then develop a model for twist relaxation and apply it to study the decay of typical twist pulses expected to arise from the action of these two enzymes. We estimate the maximum distance to which a decaying twist pulse can deliver sufficient amount of torque to drive off a bound protein. We conclude with a summary of our results and a discussion of possible applications of these ideas to chromosome remodeling and other cellular phenomena.

The material in this chapter has been published in Phys. Rev. E **63**, 61909 (2001) (45).

3.2 Driving Proteins Off DNA With Applied Torque

In this section, we analyze the stability of a protein which, when bound to DNA, constrains DNA twist (Fig. 1.5). We consider a segment of DNA of length L, the ends of which can be subject to a total twist Θ. In between the ends, we suppose a protein may be bound by free

energy μ (relative to the ideal-gas entropy of the proteins in solution i.e. $\mu \equiv k_B T \log \phi + \epsilon$, where ϕ is the protein concentration, typically $\approx$ nM to μM (55) in cells, and ϵ, the binding enthalpy, typically $\approx 20 k_B T$ for DNA-binding proteins). When bound, we suppose that the protein constrains the twist of a length D of DNA to be the equilibrium DNA twisting (constraint of twist to a different value is a straightforward generalization).

The length D of DNA with fixed linkage may be the linear sequence of base pairs covered up by the protein when it binds to DNA (Fig. 1.5a), or, in the case of a protein that can simultaneously bind two different parts of a single DNA molecule and form a loop, D is the contour length of the loop. We note that a loop forming protein need not necessarily bind both strands of the double helix at each of the two DNA binding sites to constrain DNA twist. Actually, two single strand protein-DNA attachments at each binding site are sufficient to fix the topology of the intervening DNA.

We take the twisting energy of bare DNA to be:

$$\frac{E}{k_B T} = \frac{C}{2} \int_0^L ds \left(\frac{d\theta}{ds} \right)^2 \tag{3.1}$$

where C is the twist persistence length of DNA, which has been determined to be between 75 and 100 nm from supercoiling and micromanipulation experiments. (43; 56) The excess

local twist angle $\theta(s)$ is measured relative to the equilibrium DNA twisting (i.e. the elastic equilibrium state is $d\theta/ds = 0$). The total twisting angle along the bare DNA is

$$\Theta = \int_0^L ds \frac{d\theta}{ds} \tag{3.2}$$

Assuming uniform twisting, the twist energy is simply

$$\frac{E}{k_B T} = \frac{C}{2L}\Theta^2 \tag{3.3}$$

If our protein is bound, the twisting imposed at the DNA ends must be made up by the remaining length $L - D$ of free DNA, since a length D stuck to the protein has its $d\theta/ds$ fixed at zero. Equilibrium between the bound and unbound states is described by the fixed-twist partition function:

$$Z(\Theta) = \sum_{n=0}^{1} \exp\left[-\left(\frac{C}{2(L-nD)}\Theta^2 - \frac{n\mu}{k_B T}\right)\right] \tag{3.4}$$

where n = 0 stands for the protein removed from the DNA and free to move about in the solution and n = 1 stands for the protein complexed with DNA.

In the absence of imposed torque τ, the binding free energy μ will favor protein attachment to DNA. With external twist, there is competition between the lower twist energy of the protein-off state ($n = 0$) due to the additional length D released by the dissociated protein, versus the

gain of binding free energy μ in the protein-bound state ($n = 1$). The critical torque value τ^* at which these contributions balance can be roughly estimated from (4) to be

$$\tau^* = \sqrt{\frac{2k_BT\mu\ C}{D}} \tag{3.5}$$

which for a μ of 20 k_BT (for a nucleosome) and $D = 150$ bp gives a τ^* of about $9k_BT$.

The probability for the protein to be unbound as a function of imposed twist is :

$$P_{off}(\Theta) = \left[1 + \exp\left(-\frac{\Delta E}{k_BT}\right)\right]^{-1} \tag{3.6}$$

where $\Delta E = E_{on} - E_{off}$ and the torque $\langle\tau\rangle = -k_BT\partial_\Theta \ln Z(\Theta)$ is:

$$\frac{\langle\tau\rangle}{k_BT} = \frac{C\Theta}{L(L-D)}\left[L - DP_{off}(\Theta)\right] \tag{3.7}$$

We want to calculate the critical torque τ^* required to remove a protein that is already bound to DNA. This is easily done in the fixed torque ensemble. The τ-dependent partition function is:

$$\Xi(\tau) = \int_{-\infty}^{+\infty} d\Theta\ Z(\Theta)\ \exp\left(\frac{\tau\Theta}{k_BT}\right) \quad . \tag{3.8}$$

The average twist, $\langle\Theta\rangle = \partial_\tau \ln \Xi(\tau)$, in this ensemble is related to the applied torque by:

$$\langle\Theta\rangle = \frac{\tau}{k_BTC}\left[L - DP_{on}(\tau)\right] \tag{3.9}$$

where $P_{on}(\tau) = \left(\sqrt{2\pi(L-D)/C}\right) \exp\left[\mu/k_BT + (L-D)\tau^2/2C(k_BT)^2\right] / \Xi(\tau)$.

Our theory also applies to the case where torque is applied to one end of a DNA molecule that is pinned to a surface at two places separated by a length D of DNA whose twist is fixed even while the end with no applied torque is free to flip-flop in solution. The DNA-surface connection closer to the source of torsional stress will unbind at a torque described by the above theory.

In Fig. 3.1 we show results for two cases where a DNA-binding protein constrains the twist of some region of the substrate DNA. The square-line curves concern nucleosomes, where the DNA that directly contacts the positively charged histone octamer surface has its twist frozen. Fig. 3.1a (square-line) shows the probability, $P_{on}(\tau)$, for a histone octamer to remain bound to the DNA when a torque τ is applied. In a region of width $\approx 1\ k_BT$ centered around $\tau^* \approx 9\ k_BT$, the protein occupation probability drops from 1 to 0. This is our estimate for the nucleosome destabilization torque, and agrees with (Equation 3.5). Our calculation was done for DNA $\approx$ 200 bp long with 150 bp in direct contact with the octamer, in the form of approximately 1.75 turns of the double helix around the disk shaped histone core (18). We ignore, in our simple model, the protein histone H1 which acts a ramp for the overhanging DNA to pass over when it completes (or enters) its turns over the octameric protein core. The binding free energy of each nucleosome is taken to be $\approx 20\ k_BT$ (roughly the enthalpy of nucleosome binding as measured in physiological 0.15 M univalent salt solution (57)).

The dot-line curves of Fig. 3.1 concern loop-forming proteins. Such proteins can store a lot of DNA with frozen twist without requiring a large μ since the actual protein-DNA interaction

region is only a few to ten bp in length. Typical values of μ can be estimated for say the lac repressor from experimentally determined values of the association constant $k_a = \exp(\epsilon)$ of the lac repressor. Indeed, k_a for the lac repressor-operator system has been measured to be $\approx 10^9\ M^{-1}$ (58) and cro repressor is $10^{12}\ M^{-1}$ (58). This works out to a binding enthalpy $\epsilon \approx 20 k_B T$. Using $\mu = k_B T \log \phi + \epsilon$ and a typical protein concentration $\phi \approx 10^{-9}$ M, we get a $\mu \approx 2 k_B T$. Each protein upon binding typically uses up ≈ 1000 bp of DNA in the loop (53). In comparison to nucleosomes, since much more length D of stored DNA can be released at a lower cost μ, we expect that the protein-DNA complex will become unstable at a lower τ. Indeed, even when μ is kept fixed at 20 $k_B T$ in Fig. 3.1a (dot-line), $P_{on}(\tau)$ jumps from 1 to 0 at around $\tau^* \approx 3\ k_B T$ for $L = 1500$ bp and $D = 1000$ bp.

Corresponding to each of the two cases considered above, in Fig. 3.1b, we show the associated equilibrium twist-torque response curves (square-line for nucleosome and dot-line for loop). Away from τ^*, $\langle \Theta \rangle$ evolves linearly with τ. Close to τ^*, fluctuations of the protein on or off the DNA + protein complex gives rise to a step-like torque response with a width $\Delta\tau = 1 k_B T$. Beyond τ^*, the protein has come off the DNA and we recover the twist elasticity of bare DNA.

The computed critical torque is insensitive to L over a wide range of choices for L: a 10-fold change in L leads to $< 10\%$ shift-up in τ^*. The twist-torque response curves show a more pronounced dependence on the total DNA length reflecting a a change in the distribution of twist with increasing L. For the same 10-fold jump in L, the two segments of the response curve merge with one another approaching the limit of uniform distribution of twist over a linear segment of DNA. The critical torque, τ^*, on the other hand, is still well described by

(Equation 3.5) for long DNA segments so that only small corrections to τ^* are needed when $L \rightarrow \infty$.

3.3 Dynamics of DNA Twisting

The previous section presented an equilibrium calculation which gives an estimate for the torque at which release of a protein which constrains DNA twisting becomes thermodynamically favorable. We now consider the dynamical problem of propagation of twist along a DNA to address the question of whether it is feasible for *transient* torque pulses injected at one point into dsDNA to remove proteins at a second, distant point.

In this section we study the spreading dynamics of such a twist packet and use the result to estimate the range to which sufficient torque is propagated that a bound protein can be knocked off DNA. In Sec. 3.3.1, we discuss twist distortions introduced by two DNA-modifying protein 'machines', RNA polymerase, and DNA gyrase. Then, in Sec. 3.3.2, we describe the model for twist propagation along a DNA. In Secs. 3.3.3 and 3.3.4, we examine simple solutions for time evolution of initially localized twist distortions, where $\Delta Lk = 0$ and $\Delta Lk \neq 0$, relevant to the action of RNA polymerase and gyrase, respectively. In Sec. 3.3.5, we use these results to numerically obtain estimates for R_{max}.

3.3.1 Actions of DNA-Twist-Modifying Enzymes

3.3.1.1 RNA Polymerase

Many of the DNA-processing machines inside cells may be able to generate initially localized twist 'packets'. To consider a concrete example, it is known that RNA polymerase (RNApol) tracks along the DNA double -helix during transcription, producing a net positive linking

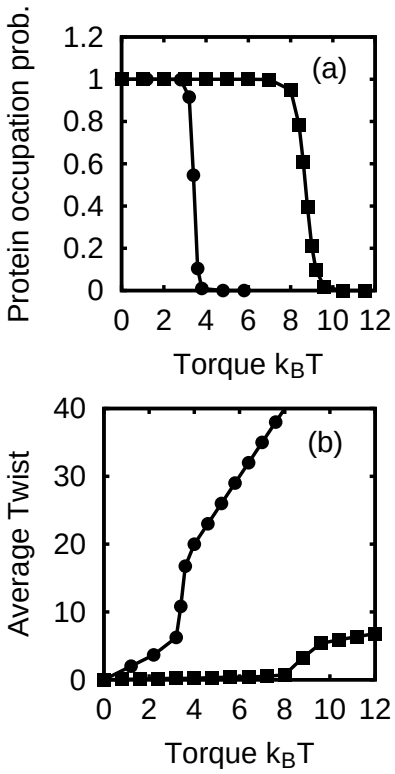


Figure 3.1. The probability for a DNA-bound protein to remain bound as applied torque (in k_BT units) is ramped up in a two-state model of DNA-protein complex (a) and the corresponding twist-torque distribution (b). (a) We consider two cases: a single nucleosome particle (line-square) with D = 150 bp, L = 170 bp, and $\mu = 20\ k_BT$ and a DNA-bound loop forming protein (line-circle) with D = 1000 bp, L = 1500 bp, and $\mu = 20\ k_BT$. The larger stored length of the loop favors dissociation of the complex at lower torques, which happens at $\tau^* \approx 3\ k_BT$. In comparison, the stored length D is much smaller for nucleosomes, implying a larger τ^* of $\approx 9\ k_BT$. In each case, the DNA-protein complex becomes unstable in a narrow region of width $\Delta\tau \approx 1\ k_BT$. (b) The corresponding twist-torque isotherms are linear away from τ^* with different slopes above and below τ^*. Near τ^*, the sudden step from one linear regime to the other corresponds to strong fluctuations between protein-on and protein-off states. Above τ^*, we recover the twist elasticity of bare DNA. We have used a $\mu \approx 20\ k_BT$ for loops to demonstrate the reduction in τ^* for increasing amounts of stored frozen twist. In practice, the loop formation enthalpy will be much lower (see text) so that our τ^* clearly represents an upper bound.

number buildup ahead of the enzyme, and a balancing net negative linking number deficit behind it (51). If we consider a single step of the enzyme, the burst of twisting generated ahead and behind the RNApol will spread out over the length of DNA and eventually settle down to a value consistent with the molecule boundary conditions. The dynamics of the twist packet is determined by balance of the local elastic strain and the dissipative torque of the surrounding medium.

RNA polymerase (RNApol) transcribes the coding strand of DNA into mRNA. Transcription involves processive motion of RNApol. Continuous motion for thousands of bp has been observed (59). If the polymerase is immobilized on a surface the DNA will thread through the polymerase thereby undergoing rotation relative to the fixed transcription machinery (59). In vivo, it is not known with certainty whether the DNA molecule rotates relative to the polymerase or vice-versa. However, since the Stokes radius of RNApol is ≈ 10 nm, it has considerable drag in water. Moreover, the nascent mRNA and in some cases the translation machinery attached to the mRNA increases the molecule's effective hydrodynamic radius thereby suggesting that, in vivo too, the DNA must rotate relative to the transcribing RNApol-complex.

The step size for RNApol is thought to be ≈ 1 bp so that the end of the DNA molecule is rotated by a full 2π in 10.5 steps. One turn is transcribed in 0.1 sec or in other words ≈ 100 bp are transcribed in 1 sec. For every radian transcribed, 1.7 NTP's are hydrolyzed so that a torque of $\approx 10 k_B T$ per radian may well be generated (at present, only the force-generating capability of RNA polymerase is precisely known(60)). Since RNA polymerase melts DNA locally during transcription, some of the torque generated by processive transcription goes into

locally opening the double helix. As discussed in Chapter 2, micromechanical experiments have shown that direct application of unwinding torque $\approx 2k_BT$ is sufficient to separate the two strands (28; 29; 23; 61).

As the DNA threads through RNApol (see Fig. 3.2a), base pairs near the entrance to the RNApol complex are slightly overtwisted (51). Since the net linking number does not change during transcription (the sugar-phosphate backbone remains intact), a compensating undertwisted region develops where the DNA exits the transcribing RNApol molecule. Since the polymerase transcribes processively, it will inject a twist strain with every step along the DNA. Because the time to complete one step is ≈ 0.1 sec while the characteristic time for twist strain decay is $\approx 10^{-6}$ sec (see below), we can consider the evolution of each individual pulse in isolation.

3.3.1.2 DNA Gyrase

The topoisomerase DNA gyrase uses stored energy (ATP) to change the linking number of closed dsDNA, by -2 per enzyme cycle. Gyrase, found in bacterial cells, is thought to be present in order to untwist the double helix so that opening of the double helix for initiation of transcription is made energetically more favorable. The linking number change caused by a single gyrase step initially generates twist strain localized to a region a few tens of bp in length. As for the case of RNApol, the initial twist distribution will then relax to its final state. The total energy expended per cycle of gyrase is $\approx 24\ \mathrm{k}_B\mathrm{T}$ (2 ATPs are hydrolysed), indicating that the torque imparted to the DNA molecule can be as large as $\approx 2k_BT$.

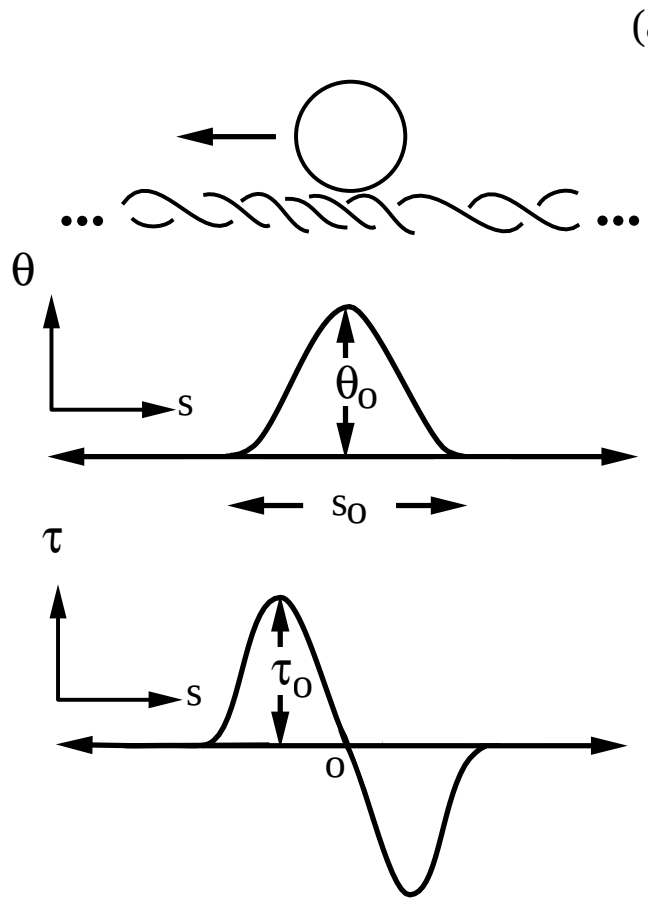


Figure 3.2. Initial twist perturbations introduced by two examples of protein wrenches, in absence of DNA supercoiling. (a) RNA Polymerase. Attachment of RNApol (circle) to DNA deforms DNA producing twist strain. The strain, $\theta(s,0)$, confined to s_0 bp, consists of increasing overtwist as we approach from the left, peaking at strain amplitude θ_0 and then undertwisting back to it's equilibrium value: $\theta(s,0) = 0$. Joint production of under- and overtwisting means $\Delta Lk = 0$, always. Since $\tau(s,t)$ is proportional to strain gradient, two domains of oppositely directed torques arise. Arrow indicates direction of motion of RNApol (continued on next page).

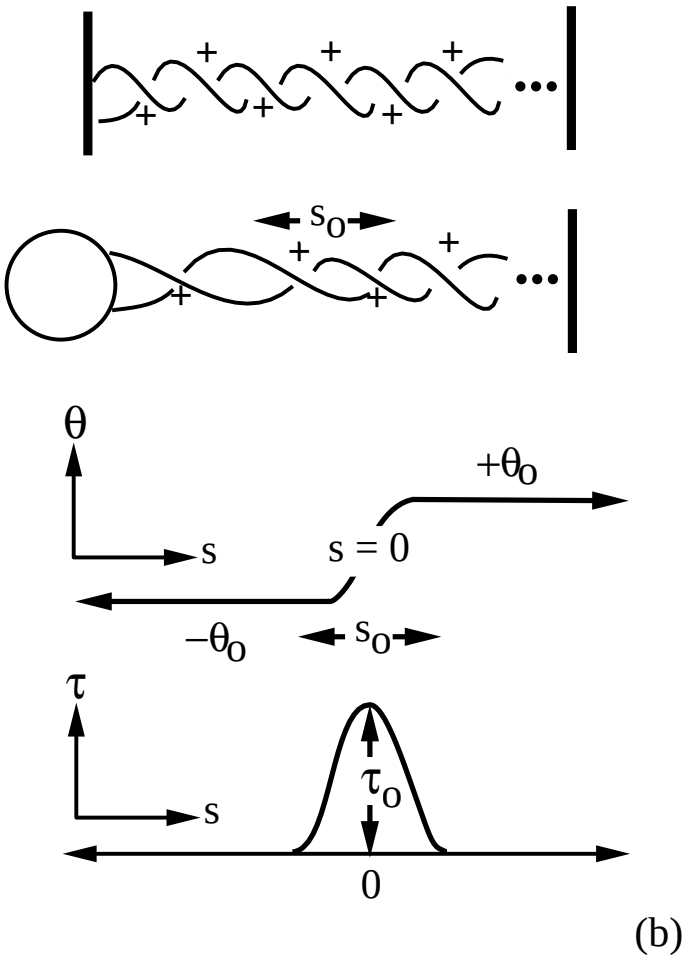


Figure 3.2. (continued) (b) DNA gyrase. In this model of gyrase, the DNA starts off with $Lk = 6$ (six +'s, ignoring the rest of the molecule) When gyrase (circle) binds, two twists get taken out near the binding site (+'s more spread out near gyrase). Without time to adjust to the new equilibrium twist for $Lk = 4$, the DNA is underrotated by $-\theta_0$ near the gyrase and overrotated, by the same amount, close to the other end (where twist rate is still consistent with $Lk = 6$). The strain switches from under- to overrotation in a region of width s_0 bp. For both RNA Polymerase and DNA gyrase, our dynamics will quickly forget the exact shape of the initial condition.

The structural mechanism by which gyrase catalyzes the topology-changing reaction is not understood. It is thought that a sequence of DNA-gyrase binding events is needed to complete the reaction. It has been proposed (62) that after gyrase binds to the DNA, it first waits for a DNA conformational fluctuation which bends a segment of the polymer over itself in such a manner that a part of the DNA gets stuck to a temporarily open domain or gate of gyrase multimer; then a second fluctuation similarly constrains another DNA segment. This segment is then allowed to spontaneously dissociate from the gyrase followed by the closing of the gate thereby preventing thermally-driven unbinding of first trapped segment. Next, gyrase introduces two nicks into the DNA to which it is bound and passes the closed trapped segment through the double-nicked DNA. This is followed by resealing of the breaks. Then the writhe introduced by gyrase gets converted into twist and propagates out through the first bent region of the DNA. The initial state that we will consider is therefore one where the DNA twist angle 'steps' by 4π over a few tens of base pairs. Below, we assume full interconversion of writhe into twist once either of these enzymes catalyzes one reaction step.

Thus, RNA polymerase is likely able to produce large torques in DNA (i.e. more than enough to denature it), while DNA gyrase is likely generating much lower peak torques, near to the $\approx 2k_BT$ threshold for separating DNA strands. However, gyrase expends up to twice the total stored energy as does RNA polymerase per reaction step (i.e. 2 ATPs vs. 1 ATP, respectively).

3.3.2 Spreading of Twist Distortions

An initially localized twist distortion in dsDNA will spread along the length of DNA and induce twisting in regions both upstream and downstream from the point where it is inserted. This spreading of a twist packet is governed by the local balance of torque produced by internal elastic strain and the dissipative torque that the medium exerts on the twisted part of DNA. Following (47), consider DNA to be an elastic cylinder so that the internal elastic torque at position s due to a given twist strain is proportional to the strain at s. If $\theta(s,t)$ is the *excess* angle, in excess of B-DNA twist, between a base pair and its nearest neighbor at position s and time t then the net torque exerted on a short segment of DNA of length Δs by internal elastic restoring forces is:

$$\Delta\tau_{internal} = k_B T C \frac{\partial^2\theta}{\partial s^2}\Delta s \tag{3.10}$$

while the torque exerted by the drag force on the same cross-section is:

$$\Delta\tau_{viscous} = \zeta_R \frac{\partial\theta}{\partial t}\Delta s \ . \tag{3.11}$$

Here ζ_R is the rotational drag coefficient per length of the cylindrical dsDNA. The rotational drag coefficient per length for a cylinder of radius a is $\zeta_R = 4\pi\eta a^2$ (47), where η is the viscosity of the surrounding fluid.

In equilibrium, $\Delta\tau_{internal} = \Delta\tau_{viscous}$, which gives the equation of motion for twist:

$$\frac{\partial\theta}{\partial t} = D\ \frac{\partial^2\theta}{\partial s^2} \tag{3.12}$$

The twist transport coefficient is $D = Ck_BT/\zeta_R$. D has the dimensions of a diffusion constant, and Eq. Equation 3.12 has the form of a diffusion equation, but note that the dynamics described above are deterministic. Plugging in the DNA radius of $a = 1$ nm, the twist persistence length $C \approx 100$ nm, the viscosity of water $\eta = 10^{-3}$ Pa·sec, and assuming room temperature, we obtain a twist transport coefficient of $D \approx 3 \times 10^{-8}$ m^2/sec. In base pair units, this is about 2.5×10^{11} bp^2/sec. Therefore, in a few microseconds, a twist pulse spreads about a kilobase along a double helix.

We can estimate the typical range R_{max} to which we expect the decaying twist strain to spread and produce torques τ large enough to destabilize a protein-DNA complex by $R_{max} \sim \sqrt{k_BTC\theta_0 s_0/\tau}$, where s_0 is the length over which the disturbance is initially spread over and θ_0 is the starting strain magnitude. From Sec. 3.2, a $\tau \approx 5k_BT$ is what is needed to take apart a protein-DNA complex. If we take $\theta_0 \approx 1$ radian, $C \approx 100$nm, $\theta_0 \approx 1$ radian and $s_0 \approx 1$nm we get an R $\approx$ 10nm.

We have ignored the three-dimensional shape of the DNA molecule in arriving at (Equation 3.12). Other authors have constructed theories including writhe dynamics(42; 63), and considered effects of permanent bends in dsDNA (50); both these effects will both play a role in propagation of twist over long distances along DNA. Below, we analyze the dynamics of temporary twist perturbations in short (at most a few persistence lengths) DNA segments for which we can ignore the contribution of writhe (the 3D conformation) of the DNA molecule to its global topology.

3.3.3 Dynamics of Twist Packet With $\Delta Lk = 0$

Twist strains with net $\Delta Lk = 0$ arise during the transcription of DNA by RNA polymerase, or in general by any DNA binding protein that locally distorts the double helix locally, without introducing breaks in the sugar-phosphate backbone. As a model of these kinds of distortions, we consider initial conditions where $\theta = 0$ far away from the region where the twist is initially distorted. We consider the interior of a DNA sufficiently long that boundary conditions do not need to be specified.

3.3.3.1 Spreading of an Angle Pulse

We consider a simple case of the dissipative spreading of a twist distortion consisting of small region of a DNA molecule which has its twisting shifted from the relaxed B-DNA structure. The base pairs can be brought back into register after the protein either unbinds from the DNA or releases the distortion into the polymer without necessarily coming off the DNA. We imagine that the starting pulse has a Gaussian shape (Fig. 3.2a):

$$\theta(s,0) = \theta_0 e^{-s^2/(2s_0^2)} \tag{3.13}$$

This initial condition has the portion of the molecule of width s_0 centered around $s = 0$ twisted by about θ_0. DNA-binding proteins routinely apply sufficient forces to the double helix to pull $s_0 \approx 10$ bp regions out of joint by ≈ 1 radian.

The elastic torque is just $k_B T C \partial\theta/\partial s$, and thus the initial condition has opposite elastic torques on opposite sides of $s = 0$. Therefore this simple initial condition is roughly applicable to

the initial twist pulse introduced by a RNApol step. The initial elastic torques have maximum amplitude $\approx \pm 0.7 k_B T C \theta_0 / s_0$. For one step of RNA polymerase, we consider an initial state with $s_0 = 10$ nm (30 bp) and $\theta_0 = 1$ rad, giving peak torques $\approx 10 k_B T$.

Given the initial condition (Equation 3.13), the angle time evolution generated by (Equation 3.12) is just

$$\theta(s,t) = \frac{\theta_0 s_0}{\sqrt{s_0^2 + 2Dt}} e^{-s^2/(2[s_0^2+2Dt])} \tag{3.14}$$

The corresponding elastic torque in the DNA is

$$k_B T C \frac{\partial \theta}{\partial s} = \frac{k_B T C \theta_0 s_0}{[s_0^2 + 2Dt]^{3/2}} s e^{-s^2/(2[s_0^2+2Dt])} \tag{3.15}$$

Fig. 3.3a shows a plot of (Equation 3.14) for different fixed times, given our proposed RNA polymerase pulse initial condition, with $s_0 = 30$ bp (10 nm), $\theta_0 = 1$ rad, $C = 100$ nm and $D = 3 \times 10^{-8}$ m^2/sec. We note that the energy stored in the initial twist distortion is $\sqrt{\pi} k_B T C \theta_0^2 / 4 s_0 = 4.4\ k_B T$, a fraction of the total energy released in an RNApol reaction step. The width increases as $\sqrt{t}$ while the height falls off as $1/\sqrt{t}$.

In Fig. 3.3b we plot the torque $\tau(s,t) = k_B T C \partial\theta / \partial s$ at fixed times. The torque drops off as the pulse spreads, on a short timescale ≈ 1 nm. In Sec. 3.3e, we will discuss how we estimate R_{max}, the maximum distance to which the torque can propagate with sufficient amplitude to destabilize a bound protein, from our torque calculation.

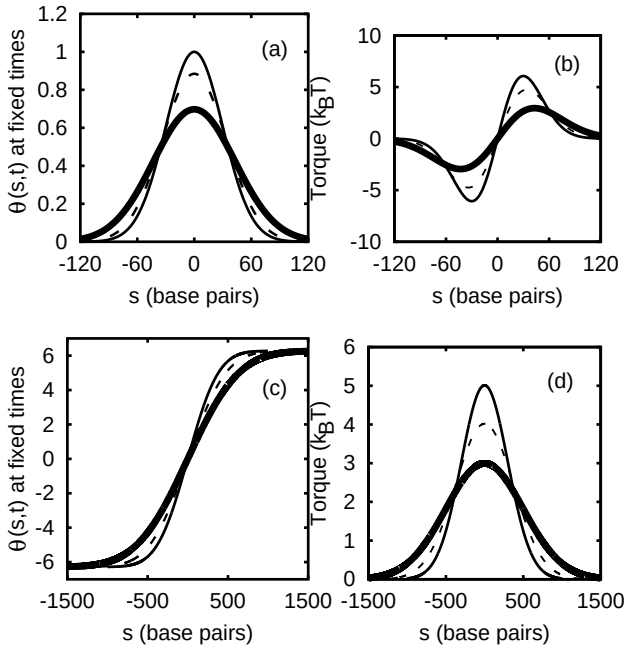


Figure 3.3. Dynamical twist and torque dissipation for $\Delta Lk = 0$ (a) and (b) and $\Delta Lk \neq 0$ (c) and (d), respectively. The molecule is 1000 bp long with twist persistence length C = 300 bp. Twist transport coefficient D = 2.5×10^{10} bp^2/sec. (a) Initial Gaussian angle pulse (solid-thin; $s_0 = 30$ bp and $\theta_0 = 1$ radian) decays symmetrically about the origin. By 0.5 ns (dashed), the pulse can deliver a torque of 3 k_BT no farther than 59 bp. Beyond 1.9 ns (solid-thick), strain pulse amplitude can no longer generate $\tau = 3\ k_BT$ anywhere on the DNA. Strain eventually decays to zero throughout the molecule. (b) Torque produced by angle pulse as a function of position along DNA (in bp) for t = 0 ns (solid-thin), 0.5 ns (dashed), and 1.9 ns (solid-thick); τ is proportional to the strain gradient. $\tau(s,0)$ has equal amounts of under- (left) and overtwisting (right) torque about $s = 0$, with peak $\tau \approx \pm 6k_BT$. At $\approx$ 0.5 ns (dashed), the torque front has simultaneously broadened and lost amplitude, fixing the range to which $\tau = 3\ k_BT$ can be delivered 59 bp. By $t = 1.9$ ns, torque front amplitude has everywhere slipped below $\tau = 3$ k_BT. (c) At $t = 0$ (solid-thin), the step-like strain on DNA starts from underrotated by $\theta_0 = 2\pi$ radians on the left to overrotated by 2π on the right, in a region $s_0 = 300$ bp. By t = 100 ns (dashed), the traveling strain has broadened but also lost amplitude, being able to deliver a torque of 3 $k_BT \approx 280$ bp away from the origin. After $\approx$ 300 ns (solid-thick) from the start, the strain cannot generate $\geq 3k_BT$ of torsion anywhere on the molecule. (d) Torque corresponding to (c) at t = 0 ns (solid-thin), 100 ns (dashed), and 300ns (solid-thick). Initially, the torque is Gaussian with peak of $\approx 5\ k_BT$ and width $\approx$ 300 bp. By 100 ns (dashed), the applied torque $\geq 3\ k_BT$ as far as 280 bp from the origin, in either direction. At $t = 300$ ns (solid-thick), the τ-amplitude is everywhere below 3 k_BT (dashed).

3.3.4 Dynamics of Twist Packet With $\Delta Lk \neq 0$

Twist strains may arise from a topology-changing event, such as the action of DNA gyrase, which changes the linking number of DNA. We consider this topology change to be concentrated in DNA twisting, so that the DNA twist angle θ jumps from $-\theta_0$ to $+\theta_0$ over a contour length of about s_0:

$$\theta(s,0) = \theta_0 \left[\left(\frac{1}{s_0\sqrt{2\pi}} \right) \int_{-\infty}^{0} dx \left(e^{-(s+x)^2/2{s_0}^2} - e^{-(s-x)^2/2{s_0}^2} \right) - 1 \right] \tag{3.16}$$

This expression may be written in terms of error functions, but keeping it in this integral form makes it more clearly a solution of (Equation 3.12). For the case of DNA gyrase, the linking number change of -2 corresponds to $\theta_0 = 2\pi$. Fig. 3.2b sketches this initial condition. The elastic energy of this initial state is $k_BTC\theta_0^2/\sqrt{\pi}s_0 \approx 22\ k_BT$ (for $s_0 = 300$ bp, $C = 300$ bp), again less than the total stored energy released in the enzyme catalytic cycle.

The initial twist state given above has a simple time evolution:

$$\theta(s,t) = \theta_0 \left[\left(\frac{2}{\pi} \right)^{1/2} \frac{1}{\sqrt{s_0^2 + 2Dt}} \int_{-\infty}^{s} dx\ e^{-x^2/(2[{s_0}^2+2Dt])} - 1 \right] \tag{3.17}$$

The elastic torque is also easily computed to be:

$$k_BTC\frac{\partial\theta}{\partial s} = \left(\sqrt{\frac{2}{\pi}} \right) \frac{k_BTC\theta_0}{\sqrt{s_0^2 + 2Dt}} e^{-s^2/(2[{s_0}^2+2Dt])} \tag{3.18}$$

Fig. 3.3c and 3.3d show $\theta(s,t)$ and $\tau(s,t)$ for the parameters $\theta_0 = 2\pi$ and $s_0 = 300$ bp, suitable to describe DNA gyrase. The initially confined twist 'step' broadens $\propto \sqrt{t}$.

3.3.5 Maximum Distance for Twist-Pulse-Induced Protein Removal

To understand what biological implications of propagating twist are, we compute, in this section, the maximum distance, R_{max} to which twist can spread out and still produce torque levels sufficient to dislodge bound proteins. We are interested in the distant production of torsion similar in magnitude to, for instance, the equilibrium torque needed to straighten out a looped domain in the DNA ($\approx 3\ k_BT$) or, perhaps, torques large enough to remove the histone core proteins from the nucleosome ($\approx 9\ k_BT$). We solve for R_{max} from (15) or (18) with τ set to, for instance, $3\ k_BT$.

It must be noted that to obtain R_{max} we compare an equilibrium torque estimate with dynamical torque (15) or (18). We therefore assume that the bound structure is able to equilibrate on the time scale of the pulse dwell time at $s = R_{max}$. For tightly bound structures, it is possible that even very large transient torques at R_{max} will not be able to effect unbinding since the torque will not persist there long enough for dissociation to occur. Our R_{max} estimates are therefore upper bounds.

In Figs. 3.3b and 3.3d, we plot the induced torque front generated by propagating twist for $\Delta Lk = 0$ and $\Delta Lk = -2$, respectively. The starting $\Delta Lk = 0$ and $\Delta Lk = -2$ pulses have peak torques $\approx 6.0\ k_BT$ and $\approx 5\ k_BT$. Since the twist evolution is dissipative, these torques present upper bounds to how much torsion the twist pulses can later generate. For both RNApol and gyrase, these torques are insufficient to disrupt nucleosome structure. We therefore consider

R_{max} for the lower threshold $\tau^* \approx 3\ k_BT$, our estimate of the torque needed to open DNA loops. Since melting DNA requires torques $\approx 2\ k_BT$, we also consider how far propagating twist can produce sustained torques large enough to open single-stranded 'bubbles' in the DNA. Since, for a given protein, the binding enthalpy to ssDNA may be lower than dsDNA, formation of single-stranded domains in DNA may locally destabilize bound proteins.

For $\Delta Lk = -2$, the strain-induced torque profile is Gaussian (Fig. 3.3d), yielding an exact expression for R_{max} as a function of the critical torque τ^* and time:

$$R_{max} = ({s_0}^2 + 2Dt)^{1/2} \left[\ln \left(\frac{2\theta_0^2 C^2}{\pi \tau^{*2} ({s_0}^2 + 2Dt)} \right) \right]^{1/2}. \tag{3.19}$$

We plot (19) as a function of $\log_{10}$(time in ns) for the unlooping (Fig. 3.4, solid-thick) and melting torques (Fig. 3.4, dashed-thick). The starting Gaussian pulse can deliver $\tau^* = 3\ k_BT$ $\approx$ 300 bp away. As the strain spreads out, the torque pulse broadens; in $\approx$ 100 ns R_{max} falls to 280 bp. Thereafter, in just another 200 ns, $R_{max} = 0$, indicating that the torque amplitude is $< 3\ k_BT$ at all points on the DNA. As the pulse broadens, its outer envelope becomes more horizontal, allowing larger torques to be delivered farther but, at the same time, the traveling pulse loses strength, thus tending to reduce R_{max}. For $\tau^* = 3\ k_BT$, R_{max} steadily declines with time indicating that pulse height dissipates so fast that even with greater spread the pulse cannot deliver torques larger than $3\ k_BT$ any farther than at $t = 0$. As a result, the torque pulse (Fig. 3.3d) produces $\tau \leq \tau^*$ in an ever shrinking interval around the origin. On the other hand, when $\tau^* = 2\ k_BT$, R_{max} is unimodal. At the start, torques $\geq \tau^*$ are produced up to $\approx$

400 bp away. As the strain relaxes through the molecule, R_{max} increases to a maximum of $\approx$ 450 bp in 300 ns and then rapidly falls to 0 in another 700 ns.

For angle pulse evolution ($\Delta Lk = 0$), R_{max} satisfies :

$$\log R_{max} - \left(\frac{{R_{max}}^2}{2({s_0}^2 + 2Dt)} \right) = \log \left(\frac{\tau^* ({s_0}^2 + 2Dt)^{3/2}}{C\theta_0 s_0} \right) . \tag{3.20}$$

This is used to plot R_{max} against $\log_{10}$(time in ns) for $\tau^* = 3\ k_BT$ (Fig. 3.4, solid-thin) and $\tau^* = 2\ k_BT$ (Fig. 3.4 dashed-thin). For $\tau^* = 3\ k_BT$, R_{max} is initially $\approx$ 55 bp, reaching its peak of $\approx$ 60 bp in about 0.5 ns, thereafter decaying to 0 in another 0.5 ns. For the lower melting torque, $R_{max} = 65$ bp at $t = 0$, rises to $\approx$ 72 bp in $\approx$ 1.7 ns and , in another 2 ns, falls to $R_{max} = 0$ bp.

Our results indicate that, with no barriers to the twist relaxation considered here, RNA polymerase will generate transient twist pulses which decay below the levels necessary to alter even loosely-bound proteins over $\approx$ 1 ns timescales. This is probably too short for even weakly bound proteins to respond. On the other hand, we find that DNA gyrase may be able to remove weakly bound proteins with torque pulses lasting up to $\approx$ 100 ns, and acting up to roughly 300 bp away. This difference stems from the larger and more extended initial angle pulse, and of course the larger initial stored energy, associated with the cycle of DNA gyrase.

3.4 Conclusion

In this chapter we have considered the stability of DNA-protein complexes in the presence of torsional stress. Our equilibrium calculations have the simple result that the characteristic

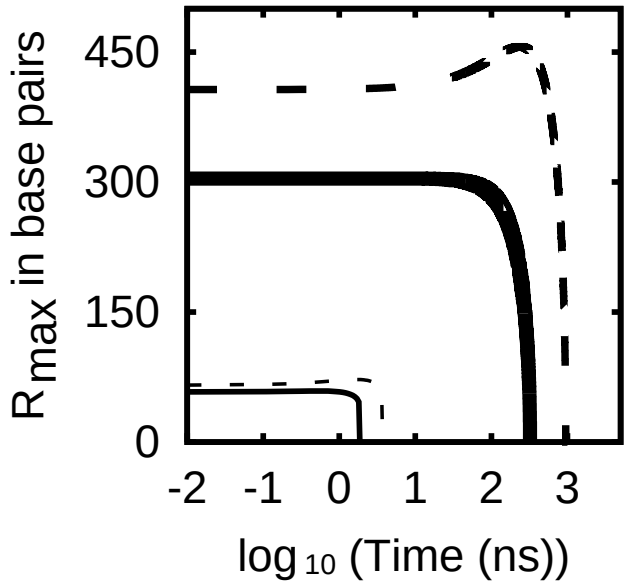


Figure 3.4. R_{max} vs. $\log_{10}$(t in ns) for critical torque $\tau^* = 3\ k_BT$ when $\Delta Lk = 0$ (solid-thin) and $\Delta Lk \neq 0$ (solid-thick) and $\tau^* = 2\ k_BT$ when $\Delta Lk = 0$ (dashed-thin) and $\Delta Lk = -2$ (dashed-thick). All parameter choices as in Fig. 4. For RNApol ($\tau^* = 3\ k_BT$), the strain ($s_0 = 30bp, \theta_0 = 1$ radian) initially builds up a torque $\geq \tau^*$ up to 57 bp but as it spreads and loses height it cannot create such torques beyond 59 bp. The pulse needs ≈ 0.5 ns to generate torque this far. In ≈ 2 ns, torque is below τ^* at all points on the DNA so that $R_{max} = 0$. When $\tau^* = 2\ k_BT$, R_{max} at t =0 is 65 bp, peaking at 72.3 bp in 1.7ns and decaying to 0 in another ≈ 2 ns. Gyrase ($\tau^* = 3\ k_BT$, $s_0 = 300$ bp, $\theta_0 = 2\pi$ radians) at first generates torsion $\geq \tau^*$ up to ≈ 300 bp, asymmetrically about the origin. But it takes only another 200 ns for the maximum torque pulse height to slip below $\tau^* = 3\ k_BT$ i. e. $R_{max} = 0$. For $\tau^* = 2\ k_BT$, initial R_{max} is ≈ 400 bp. With increasing pulse width R_{max}begins to rise eventually reaching a maximum of ≈ 450 bp in ≈ 300 ns. In the next 700 ns, the torque envelope becomes flatter, tending to push torque farther, but not fast enough to offset the rapid loss of pulse amplitude, the net effect being a steady decrease in R_{max}. It takes longer for $R_{max} \to 0$ when $\Delta LK = -2$ versus $\Delta Lk = 0$ because of the greater starting amplitude and spread of the gyrase induced strain compared to RNA Polymerase. For fixed ΔLk, the peak R_{max} is larger for $\tau^* = 2k_BT$ than $3k_BT$ as expected from our scaling relation for R_{max} (see text).

torque associated with removal of a protein which allows contour length D of DNA to absorb twisting strain is $\tau^* \approx \sqrt{k_B T \mu C/D}$, where μ is the free energy difference associated with the protein-DNA interactions holding the complex together. This formula is the torque analog of the characteristic force needed to open a protein-DNA complex, $f^* \approx \mu/D$ (55). Strongly-bound compact structures like the nucleosome ($\mu \approx 20 k_B T$, $D \approx 50$ nm) require large torques $\approx 10 k_B T$ to disrupt them. On the other hand, more weakly bound structures, or DNA loops with large D values, likely will require torques in the range 2 to 3 $k_B T$.

Torques in the few $k_B T$ range may be applied by RNA polymerase and DNA gyrase, according to experimental data and rough theoretical estimates (50; 42; 59; 60). Thus, torques applied by RNA polymerase and gyrase to an anchored DNA segment can reach those necessary to open relatively weakly-bound DNA-protein complexes. On the other hand, nucleosomes should be relatively stable to the direct effect of the torques applied by these enzymes.

These torques are close to those needed to disrupt DNA secondary structure. Double-stranded DNA is now known to be unstable against such strong torques, and converts to 'melted' forms for left-handed (unwinding) torques larger than $2 k_B T$, and for overwinding torques larger than $6 k_B T$ (28; 29; 61; 30). This raises the interesting question of the interplay between torque-generated denaturation and protein dissociation; one can easily imagine the unbinding of a protein in response to local DNA melting. The coupling of DNA melting to protein dissociation could be described using an elaboration of the equilibrium model of the present chapter.

It is important to realize that our equilibrium theory applies most strictly to equilibrium experiments, i.e. where the time scale for the experiment greatly exceeds the on and off times for the DNA-protein complex. Particularly in the case of large structures like nucleosomes, there may be large free energy barriers to overcome to transit between bound and unbound states (64).

We have also analyzed the dynamics of spreading of single twist pulses and steps according to the simple dynamics used to describe twist relaxation of short DNA segments in free solution(47). The small diameter of the DNA double helix results in rapid spreading of twist pulses. In turn, this makes an initial torque pulse spread and decay rapidly. The result is that single pulses of torque injected by enzymes can affect bound proteins only at short (< 100 bp) distances.

Our dynamical model supposes that any net linking number injected into the DNA is eventually removed, e.g. at free ends of a linear DNA. In the case where one has a small circular or otherwise topologically constrained domain, torsional stress will of course build up cumulatively with each enzyme cycle. Usually a steady state will be reached, defined in the test tube by the enzyme stall point or, in vivo, by the action of counteracting enzymes (e.g. topoisomerase I, which allows linking number to be removed from a point along a dsDNA by generation of a transient break in one DNA strand). In the case where an appreciable steady-state torque is reached, our pulse-propagation model should not be used, and instead the 'equilibrium' theory will be relevant.

Some enzymes may inject twist strain continously into DNA over some time (e.g. RNApol). It is then possible that far away from the strain source, approximately steady-state torques may build up cumulatively. For instance, if the enzyme cycle time is comparable to the typical propagation time of the twist pulse then successively injected twist strains can ramp up the torque through small increments. For the simple twist propagation considered here, this is not relevant, since the RNApol/gyrase cycle times (≈ 0.1 s) are long relative to twist propagation times on few kb scales (recall that a twist pulse spreads over 1 kb in $\approx 10^{-6}$ sec). However, we note that P. Nelson has suggested that intrinsic bends may greatly slow the propagation of twist along long DNAs, perhaps to the point that a steady-state torque may be applied by RNA polymerase over ≈ 1000 bp distances even on linear DNA (50). If such a steady state of torsional stress occurs, the 'equilibrium' theory is best applied.

We have not considered the complete dynamics of linking number, including the interconversion of twist to writhe (63). This is complicated since the subsequent relaxation of writhe must take into account hydrodynamic interactions. We have simplified the problem by considering only short (< 1000 bp) regions of DNA for which the writhing dynamics can plausibly be ignored. However, on large (few kb) molecules, the twist relaxation time will become long enough for writhing to occur (42).

Enzyme-driven twist propagation has been proposed as a possible chromatin "remodeling" mechanism for a class of recently discovered, highly conserved, ATP-consuming proteins (67). These enzymes facilitate in vivo transcription of genes, by allowing regulatory proteins to bind to DNA domains that are otherwise inaccessible because they are confined to the nucleosome.

Increasing access to bound regions of the DNA could involve: (a) sliding or linearly displacing the histone protein core along the DNA, (b) creating a large transient DNA loop by partially unbinding a segment of the wrapped DNA from the nucleosome, thereby exposing the looped DNA (c) temporarily ejecting the histone from the chromatin fiber, or (d) first disassembling and then later reassembling the histone octamer.

Currently, transient unbinding of a large region of nucleosome surface-bound DNA is the favored way of thinking about remodeling-enzyme facilitated gene regulation(52). Since remodeling activity generates torque it has been suggested that this may be used to drive twist strain through the DNA, and causing the elastic twist energy of deformed DNA to tear off bound regions from the nucleosome surface. The binding energy per length of DNA bound to the nucleosome is $\approx 0.15\ k_BT$/bp (assuming that 146 bp of DNA has $\mu = 20\ k_BT$)(57; 68). Our equilibrium estimate for the torque needed to unbind a pinned domain of DNA, $\tau^* \approx \sqrt{2k_BTC(\mu/D)}$ where $C = 100$ nm is the twist persistence length, predicts that remodeling complexes have to produce sustained torques $\approx 9\ k_BT$ to unbind a region of the surface-bound DNA. Since the torque depends only on the binding energy per unit length of DNA, disrupting the initial DNA-histone contacts, even in the presence of thermally-induced unpinning of contacts near the DNA overhangs, requires as much torque as disrupting later contacts. Therefore, the nucleosome will be stable in the presence of gradually increased torque, until a threshold torque is reached, at which point the histones will dissociate. Of course, the large total $\mu \approx 20\ k_BT$ suggests that an even larger torque barrier must be overcome for histone removal to occur.

We conclude with the observation that our mechanisms for torque-propagation and subsequent torque-based bound-protein ejection depend on the rate at which protein motors acting on DNA produce torque. Although this question is highly relevant to our work, it is also of great significance to other areas of biological micromechanics and is ultimately related to the actual force-generation mechanism used by protein motors and wrenches. In spite of their great importance, these questions remain unresolved to date. Moreover, there are no direct measurements of this time scale for any motor. Future experimental work in this direction would be of great importance in suggesting whether the mechanisms we have described are plausible.

CHAPTER 4

UNZIPPING DNA WITH BUILT-IN SEQUENCE INHOMOGENEITIES AND BOUND PROTEINS

4.1 Introduction

The two strands of the DNA double helix can be 'unzipped' by application of force. Since inhomogeneities in sequence generate inhomogeneities in strand binding, force variations occur, correlated with DNA sequence. A few groups have carried out single-molecule studies of DNA unzipping by force. Essevez-Roulet, Bockelmann and Heslot have studied the $\approx$ 15 pN forces encountered during unzipping of 50 kb λ DNA, and that group has been able to resolve the variation in force with sequence (70; 71; 72; 73). GC-rich regions were observed to require higher force for their unzipping than AT-rich regions, in accord with the generally higher base-pairing free energy associated with GC-rich sequences.

The basic unzipping force follows from the simple argument that unzipping will be thermodynamically favorable when the free energy required for unzipping one base, $g_0 \approx 2.0\ k_BT$, is equal to the mechanical work done, $f\ell$, where $\ell \approx 0.5$ nm is the projected length of ssDNA liberated during unzipping of one base pair. The resulting force $f \approx g_0/\ell$ is on the order of the typical force of 15 pN observed experimentally. A more quantitative calculation will be introduced below, based on known elastic properties of ssDNA.

In this simplest picture, the variation of strand-separation free energy along the molecule gives rise to the variations in unzipping force observed experimentally. However, experiments observe a characteristic 'sawtooth' force pattern similar to stick-slip motion (70; 71; 72). Below we carry out a simple analytical treatment of a single inhomogeneity during unzipping of a dsDNA, characterized by a distinct unpairing free energy per base $g_0 + g_1$. This inhomogeneity will be Δ bases in length, and located N bases from the beginning of the unzipping.

Our calculations are carried out in the ensemble of fixed extension between the single-strand ends, corresponding to experiments done with stiff cantilevers or stiff optical traps. For the case where the binding in the inhomogeneous region exceeds that upstream and downstream ($g_1 > g_0$), we find that there are two types of force traces generically encountered. The first is a sawtooth force trace; in this case the unzipping fork is never at equilibrium inside the inhomogeneity. At the discontinuous force at the sawtooth terminus, the two 'coexisting' positions of the unzipping fork have the same energy; neither is inside the inhomogeneity. Near this value of extension, one should see dynamic 'switching' between the two coexisting states.

The second type of force trace, obtained for sufficiently long regions of inhomogeneity, is a 'plateau' trace, where the force reaches a value characteristic of the inhomogeneity strength g_1. The force plateau is terminated by a force jump between two 'coexisting' states; unlike the sawtooth, the higher-force state is inside the inhomogeneity region.

We analyze the equilibrium statistical mechanics of our model using mean-field theory, and with exact numerical calculations. We then discuss the barrier-crossing kinetics of two-state switching expected at the force discontinuities. Finally, we generalize our inhomogeneity model

to analyze the force signal expected from a DNA-binding protein which binds to a specific dsDNA site. As the unzipping fork passes the binding site, protein dissociation can occur, accompanied by a force signal. We find that even moderately-strongly-binding proteins generate signals easily detected by present experiments.

4.2 Unzipping DNA With Sequence Inhomogeneity

4.2.1 Model

The free-energy, g_0, to separate (i.e. remove to infinity) two paired complementary bases in dsDNA in physiological salt conditions (say 150 nM NaCl) is ≈ 1 to 4 k_BT (69) depending on sequence; we shall throughout this chapter use the average value of $g_0 = 2.5\ k_BT$. The energetic component comes from (1) complementary hydrogen bonding, (2) electrostatic repulsion between phosphate groups on sugar-phosphate backbones and (3) counter-ion binding. The entropic component comes from (1) segregation of bases into interior of the double-helix (i.e. base-stacking interaction) and (2) local ordering of water molecules around the DNA (DNA excluded-volume interactions can be ignored here)(8).

In certain DNA sequences, the measured per basepair stabilization energy is known to be larger or smaller than this baseline. For instance, basepairs in a GC-rich region inserted into an otherwise random-sequence DNA molecule have a higher basepair disruption energy. Similarly, certain known sequences of basepairs optimize hydrophobic stacking interactions leading to locally stronger or weaker basepairing. More generally, by modulating base-stacking and local density of hydrogen-bonding through variation of sequence, 'abnormally' stable or weak DNA oligomers can be created.

The typical force scale for unzipping a single basepair is set by g_0 and the ssDNA elasticity. Since it turns out that $\approx$ 2 k_BT of work is done over a distance of $\approx$ 0.5 nm to unzip a basepair, a force of $\approx$ 15 pN is expected. Imagine, now, unzipping a DNA molecule of length L bp (Fig. 4.1a) by applying equal and opposite forces to two beads ligated to the starts of the two backbones at one end of that molecule. We suppose that the polymer has a random sequence (binding energy per bp g_0) save for a Δ bp long stretch starting at basepair N and terminating at $N + \Delta$ where, to pry apart a basepair, an appreciable amount of *additional* work per base g_1 ($|g_1| > 0$) needs to be done. Suppose we use a stiff pipette so that its thermal-bending fluctuations are negligible and we can reliably measure the inter-bead distance, $2X$, which we use as our control parameter. At each value of X, some number n basepairs will be unzipped on average.

In our model, we assume no *intermittent* basepairing between parts of the two complementary ssDNAs that result from unzipping n basepairs, so that n also denotes the position, along the DNA, of the tip of the unzipping fork. (Fig. 4.1a). For a given X, n and applied force fluctuate in our ensemble; n fluctuates only through *sequential* closing or opening of basepairs at the tip of the fork. For convenience, we define $\gamma = g_1/g_0$ and $\delta = \Delta/N$.

A state of our molecule is specified by n i.e. by the number of basepairs disrupted. The corresponding internal energy of such a state is

$$\frac{E(n)}{k_BT} = \frac{3X^2}{b^2 n} + ng_0 + \Delta g_1 h\left(\frac{n-N}{\Delta}\right). \tag{4.1}$$

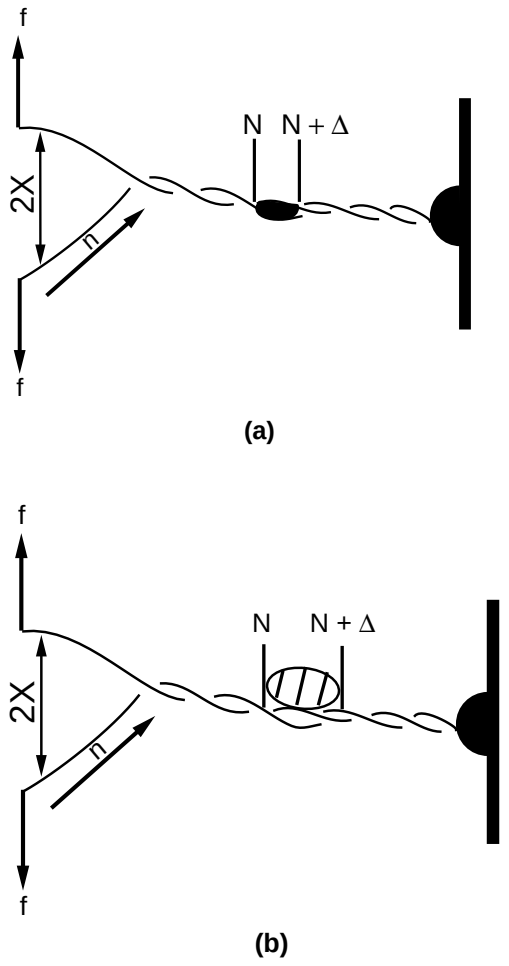


Figure 4.1. Unzipping DNA with sequence inhomogeneity (a) and bound proteins (b). The inhomogeneity (dark patch) is located at basepair N and is Δ bp long. The experimentalist controls the distance, $2X$, between the twos ssDNA arms, while n denotes the position of the unzipping fork for some X. Note that although fluctuations in the two arms may result in reannealing of some of the broken complementary basepairs, we neglect this possibility, because of the very short persistence length of ssDNA. While each basepair is bound with an average of g_0 energy, basepairs inside the inhomogeneity have a further excess, g_1 of cohesive energy. A variation on the theme is (b). A protein (stripped oval) binds at N and uses up Δ basepairs. Binding to dsDNA is energetically favored by an amount μ_{ds}. Binding to ssDNA is not favored or only very weakly stabilized; the difference between the free energies of protein binding to dsDNA and ssDNA is μ_{ds-ss}.

Here the function $h(x)$ is

$$h(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x \leq 1 \\ 1 & x > 1 \end{cases} \tag{4.2}$$

For a given X, if n basepairs are unzipped, we treat the two ssDNA components as a single Gaussian polymer of $2n$ units, with $2X$ as its end-to-end distance. We ignore the possibility that fluctuations in our Gaussian polymer may lead to formation of denaturation bubbles; this is a good approximation because at the force scales relevant to mechanical unzipping of DNA, the two unzipped DNA-arms are quite stretched, and so unlikely to undergo large spatial fluctuations which could bring complementary bases on opposite ssDNA-arms close enough to form basepairs.The first term in eq. 1, therefore, represents the entropic elasticity of a phantom polymer with persistence length b. If $n < N$, we do work ng_0 to unzip n basepairs. But once we begin unzipping basepairs inside the inhomogeneity, i.e. when $n \geq N$, additional work proportional to g_1 is done. The term, $\Delta h(\frac{n-N}{\Delta})$, is 0 if $n < N$; when $n \geq N$, it is simply the number of basepairs disrupted inside the inhomogeneity, i.e. $n - N$.

The partition function is

$$Z(X) = \sum_{n=1}^{\infty} \exp\left[-\frac{3X^2}{b^2 n} - ng_0 - \Delta g_1 h\left(\frac{n-N}{\Delta}\right)\right]. \tag{4.3}$$

The average force, $\langle f \rangle$, in $k_B T$ units is

$$\langle f \rangle = -k_B T \frac{\partial \ln Z(X)}{\partial X}. \tag{4.4}$$

Our theory is a modified version of Bockelmann *et al*'s model presented in (72). In their work, they explicitly introduce sequence-dependent basepair opening energies, which they treat as fitting parameters. The polymer elasticity of the two unzipped ssDNA arms is treated via the freely-jointed chain (FJC) model, for which an exact expression for the force vs. distance response is known (11). The persistence length, the contour length of ssDNA consisting of a given number of bases, and the elastic modulus of ssDNA also appear in their work as free parameters. Finally, an explicit description of the elastic stretching of the DNA linkers, used to connect the main λ-DNA molecule to the protein coated beads, is introduced.

Throughout this chapter, we neglect g_0's baseline genomic sequence dependence; designed sequence inhomogeneities or bound proteins are taken to interact with DNA with energies well separated from genomic variations in g_0. Outside the inhomogeneity or protein binding site, DNA is taken to have a uniform cohesive energy of g_0 while inside the obstruction its binding energy is modified to $g_0 + g_1$. It is straightforward in our model to include basepair-level variations in g_0 but we do not do so. We also neglect the elastic response of the linkers and use a simple Gaussian polymer model, instead of the FJC model, to describe the entropic elasticity of ssDNA. These ommisions are readily rectified and do not lead to different physics.

4.2.2 Mean-Field Analysis

If we rescale $\xi = (3/g_0)^{1/2} X/(Nb)$, $\delta = \Delta/N$, $\nu = n/N$, and $\gamma = g_1/g_0$, and scale the energy (Equation 4.1) in units of Ng_0k_BT, we obtain:

$$\frac{E}{k_BTNg_0} = \frac{\xi^2}{\nu} + \nu + \delta\gamma h\left(\frac{\nu-1}{\delta}\right). \tag{4.5}$$

The mean-field (minimum-energy) solution given the rescaled extension ξ results from minimizing Equation 4.5 with respect to the rescaled number of open bases ν. The mean-field solution depends only on the rescaled parameters γ and δ.

The mean-field 'equation of state', or relation between ν and ξ, is obtained by differentiating Eq. Equation 4.5 with respect to ν:

$$\frac{\xi^2}{\nu^2} = 1 + \gamma h' \left(\frac{\nu - 1}{\delta} \right) \tag{4.6}$$

Since h' is either 1 or 0, depending on whether $(\nu - 1)/\delta$ is inside or outside the interval $[0, 1]$, Eq. (Equation 4.6) can be restated as:

$$\xi = \begin{cases} \nu & \nu < 1 \quad or \quad \nu > 1 + \delta \quad (I) \\ (1+\gamma)^{1/2}\nu & 1 < \nu < \delta \quad (II) \end{cases} \tag{4.7}$$

At all times the force is given by $f/f_0 = \xi/\nu$, where $f_0 = (3g_0)^{1/2}/b = 15$ pN. In region I, $f/f_0 = 1$, corresponding to the constant force of unzipping encountered outside the inhomogeneity. In region II, the force is $f/f_0 = (1+\gamma)^{1/2}$, i.e. higher for $\gamma > 0$ due to the extra strand binding energy that must be overcome to make unzipping proceed.

The slanted lines of Fig. 4.2 show the equation of state (Equation 4.7), assuming $\gamma > 0$ (i.e. higher strand binding energy in the inhomogeneity). As ξ is increased from zero, the equilibrium ν increases initially linearly and the unzipping force is constant. However, when ξ reaches $\xi_1 = 1$ (corresponding to the unzipping fork reaching the inhomogeneity), we reach the end of the first segment of (Equation 4.7). For ξ between ξ_1 and the lesser of $\xi_2 = 1 + \delta$ and

$\xi_3 = (1+\gamma)^{1/2}$, there is a 'gap' in Eq. Equation 4.7. When the unzipping fork reaches this gap, there will be a 'stick-slip' force signal, which we find to occur with two forms. The first is a sawtooth force signal, and the second is a 'ramp-plateau' signal.

4.2.2.1 Sawtooth Force Signal

Fig. 4.2a shows the case where $\xi_2 < \xi_3$, which occurs for $\delta < (1+\gamma)^{1/2} - 1$ (i.e. for a sufficiently short region of inhomogeneity, or for sufficiently large inhomogeneity binding). For $\xi_1 < \xi < \xi_2$, ν must remain constant at 1, as indicated in Fig. 4.2, since there is no mean-field equilibrium state with ν between 1 and $1+\delta$.

At some point ξ^* the energy of the stretched configuration with $\nu = 1$ will be equal to the energy of one of the states given by (Equation 4.7), and ν will jump to a new, larger value. This may occur for $\xi^* < \xi_3$, as shown in Fig. 4.2a; in this case, the force signal will be a sawtooth. To be more precise, for $\xi < \xi_1 = 1$, $f/f_0 = 1$ (unzipping of homogeneous region); for $1 < \xi < \xi^*$, $f/f_0 = \xi$ (stretching of ssDNA from 'stuck' fork); then, for $\xi > \xi^*$, we return to $f/f_0 = 1$. The result is a sawtooth terminated by a force jump.

To find the value ξ^* at which the sawtooth jump occurs, one must find where the energy of the stuck fork ($\beta E/[Ng_0] = \xi^2+1$) equals the energy of the opened region ($\beta E/[Ng_0] = 2\xi+\gamma\delta$). These energies are equal when $\xi^* = 1+(\gamma\delta)^{1/2}$. Of course, to obtain a sawtooth, we must make sure that $\xi^* < \xi_3$. This condition holds for

$$\delta < 1 - \frac{2}{\gamma}([1+\gamma]^{1/2} - 1) \tag{4.8}$$

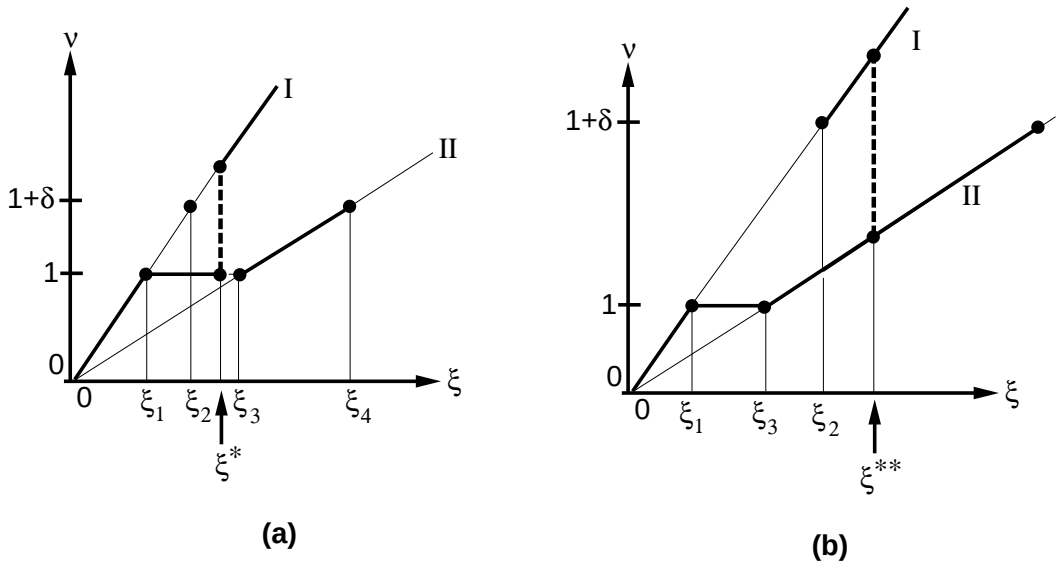


Figure 4.2. Illustration of mean field analysis of when DNA's unzipping response is sawtooth (a) and when ramp-plateau (b). Along the x-axis is plotted $\xi = \sqrt{3/g_0}X/Nb$ and along the y-axis $\nu = n/N$. These are rescaled versions of X and n. $\delta = \Delta/N$. The two lines I and II are the mean-field plots of $\langle n(X)\rangle$ (dropping rescaled units for the moment) when $n < N$ (line I), i.e. before the unzipping fork has entered the inhomogeneity, and $n > N + \Delta$ (line II), i.e. after the unzipping fork has exited the inhomogeneity. When $1 \leq \nu \leq 1 + \delta$, the molecule cannot lie on branch I. (a) When $\xi_2 = (1 + \delta) < \xi_3 = \sqrt{(}1 + \gamma)$ and $\xi_=(1 + \delta)\sqrt{(}1 + \gamma)$, the molecule will never lie on branch II since at some $\xi^* < \xi_3$ there will be a state on branch I with the same energy as $\nu = 1$. This gives the sawtooth pattern. (b) When $\xi_3 < \xi_2$, the molecule travels along branch II until ξ^{**}, where it finds a state on branch I with equal energy; then it jumps to branch I. This gives rise to the ramp-plateau pattern.

i.e. for sufficiently small δ corresponding to a short inhomogeneity. For the sawtooth force signal, the mean-field fork position is never inside the inhomogeneous region; the fork 'sticks' at the beginning of the inhomogeneity, and then 'slips' through the entire inhomogeneous region. Such stick-slip events are clear in the experimental data of Bockelmann *et al* (70; 71; 72).

For the sawtooth force signal, the maximum force occurs just before the force jump. The amplitude of the force jump in units of f_0 is just $(\gamma\delta)^{1/2}$, and therefore can be increased either by increasing the width or binding strength of the inhomogeneous region. At ξ^*, the two states separated by the force jump have equal energy, and are in 'coexistence', with an energy barrier separating them. The height of this energy barrier is

$$\frac{\beta E_{barrier}}{N g_0} = \frac{(1 + [\gamma\delta]^{1/2})^2}{1+\delta} + \delta - 2(\gamma\delta)^{1/2} - 1 \tag{4.9}$$

For small δ, the barrier $\approx \gamma\delta$; in terms of actual energy this gives $E_{barrier} \approx k_B T g_1 \Delta$, essentially the excess work done opening the inhomogeneous region. Given that g_1 might be 0.5, for $\Delta = 30$ bp, we obtain a $15 k_B T$ barrier. Such a large barrier should require a long time for thermal crossing. In fact, recent experiments on RNA helix-loop unzipping show slow (≈ 1 sec) two-state switching corresponding to crossing of barriers of roughly this amplitude (76). We will discuss the barrier-crossing dynamics in more detail below.

4.2.2.2 Ramp-Plateau Force Signal

Fig. 4.2b (which is identical to Fig. 4.2a except for the order of ξ_2 and ξ_3) illustrates the case $\xi_3 < \xi_2$, which occurs for $\delta > (1+\gamma)^{1/2} - 1$. Analogous to the sawtooth response analyzed

above, when ξ lies in $[\xi_1, \xi_3]$, $\nu = 1$. In this regime, the force ramps-up from $f/f_0 = 1$ at ξ_1 to $f/f_0 = (1+\gamma)^{1/2}$ at ξ_3. For $\xi \leq \xi_1$, $f/f_0 = 1$.

When ξ exits this interval on the right, ν evolves along branch II of Fig. 4.2b instead of branch I. By construction, this happens because the only available mean-field solutions lie on branch II or energy is lower here than states on branch I (see Fig. 4.2b). In any case, the energy minima will continue to lie on this line until ξ reaches $\xi^{**} > \xi_3$, where the energy on branch I for ξ^{**} is equal to the energy on branch II for ξ^{**}. Therefore, as long as $\xi_3 \leq \xi < \xi^{**}$, the tension in on each single-stranded arm of the DNA molecule will be constant at $f/f_0 = (1+\gamma)^{1/2}$, which is just the force at ξ_3.

Beyond ξ^{**}, the energy minima will switch from lying on branch II back to on branch I since the energy will always be lower on this branch (because $\gamma = g_1/g_0 > 0$). As a result, the force at ξ^{**} abruptly drops from $f/f_0 = (1+\gamma)^{1/2}$ to $f/f_0 = 1$, remaining constant thereafter.

The switch from branch I to branch II occurs when

$$\xi^{**} = \frac{d}{2}\left(1 + \sqrt{1+\gamma}\right), \quad d < \xi^{**} < d\sqrt{1+\gamma}. \tag{4.10}$$

Restricting ξ^{**} to $\left]d, d\sqrt{1+\gamma}\right[$ or, equivalently, the condition $1 < (1+\sqrt{\gamma+1})/2 < \sqrt{1+\gamma}$, is what leads to the ramp-plateau instead of the sawtooth force profile.

When the force signal has a ramp-plateau profile, the barrier height at the force jump is given by

$$\frac{\beta E_{barrier}}{N g_0} = \left(\frac{d}{2}\right)\left[1 + \frac{\gamma}{2} - \sqrt{1+\gamma}\right]. \tag{4.11}$$

This gives barrier heights comparable to those from the sawtooth case.

To summarize, as we increase ξ from 0, the force is constant until ξ reaches ξ_1. Then, the force builds-up from $f/f_0 = 1$ to $f/f_0 = (1+\gamma)^{1/2}$ at ξ_3 (ramp). Beyond ξ_3, the force is once again constant but bigger than at ξ_1 (elevated plateau). Past ξ_3, when ξ is ξ^{**}, the force rapidly drops from $(1+\gamma)^{1/2}$ to $f/f_0 = 1$, its value at ξ_1, completing the ramp-plateau force profile. By adjusting Δ or g_1, we can transition between two distinct force responses. As discussed below, this is also true when we take into account fluctuations.

4.2.3 Evaluation of Z(X) With Fluctuations

4.2.3.1 Force-Distance Response

We incorporate fluctuations around the saddle-point into the partition function by explicitly performing the sum in (Equation 4.3) on a computer; our results are therefore exact within our single-domain-wall approximation. This allows us to calculate, for instance, the force-extension curve shown in Fig. 4.3, by computing the derivative in (Equation 4.4). In Fig. 4.3, we show the result of a calculation for the DNA's force-distance response, where we have taken $g_1 = 0.7$ k_BT and $N = 40$ bp. Throughout this chapter, all numerical results, unless specified otherwise, are obtained by setting $g_0 = 2.5$ k_BT and $b = 0.7$ nm. Three curves are shown, corresponding to the molecule's behavior when Δ (the length of the inhomogeneity) is 12 bp (solid line), 24 bp (dashed line), and 36 bp (solid line with solid circles), respectively.

There is some disagreement in the literature about the correct value of ssDNA's persistence length. In fact, there is even debate whether a persistence length can be defined at all for ssDNA (75). These problems arise from the fact that, unlike dsDNA, ssDNA sticks to itself

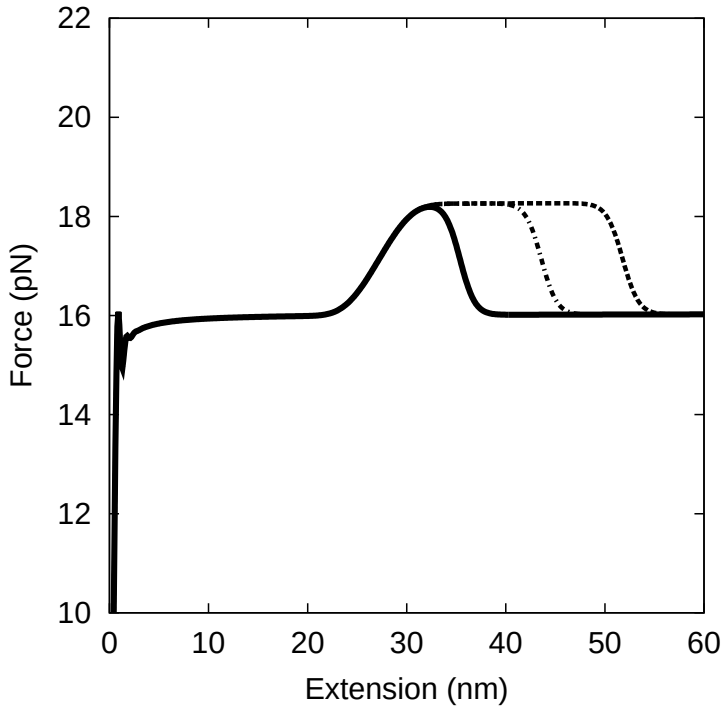


Figure 4.3. Unzipping DNA with sequence inhomogeneity. Plotted along the x-axis is one-half end-to-end extension i.e. X. Force in picoNewtons (pN) is plot along y-axis. The three curves correspond to $\Delta/N = 0.3$ (solid), =0.6 (uneven dashes), and =0.9 (dashes). $N = 40$ bp, $g_0 = 2.5\ k_BT$, $b = 0.7$ nm and $g_1/g_0 = 0.3$ giving $g_1 = 0.7$. When Δ is 12 bp long (solid), the force-response is sawtooth-like but as we increase Δ to 24 bp (uneven dashes) or 36 bp (dashes), the force kink stretches out into a ramp-plateau shape. Note that the peak force is well above the constant unzipping force, so that, experimentally, it can be discerned from the background signal.

nonspecifically under physiological conditions. While at large forces, say $f > 10$ pN, ssDNA shows elastic behavior similar to a flexible polymer, at low forces, $f < 10$ pN, self-interactions, including complementary basepairing and hydrophobic aggregation, lead to a more complicated and as yet not fully understood elastic response. Specifically, the problem is that numerous self-adhesion interactions lead to a ssDNA secondary structure that is not understood. So, the persistence length b enters our model as a phenomenological parameter which summarizes number of different effects that bear upon ssDNA's persistence length, each of which is incompletely understood. As a parameter, we find that its present value gives a good fit of theory to experimental data.

From Fig. 4.3 we see that as Δ is increased, the force-profile transitions from sawtooth to ramp-plateau shape. For instance, when $\Delta = 12$ bp, the force profile is a sawtooth, while for $\Delta = 24$ bp, it has a ramp-plateau profile. Increasing Δ beyond this transition increases the 'width' of the plateau but does not change the peak force, i.e. the height of the plateau, or the shape of the force pulse. Indeed, the slope of the force decay at the end of the plateau is parallel for all $\Delta \geq 12$ bp; in other words, beyond the transition to ramp-plateau, all we see is a stretching of the force signal.

Switching between sawtooth and ramp-plateau profiles with changing Δ can be understood qualitatively as follows. Imagine sitting atop the tip of the unzipping fork. For a fixed value of force, its position keeps fluctuating about some average. Away from the inhomogeneity, the tip's average position proportionally tracks changes in applied force, so that $\langle n \rangle / N$ vs. f is linear and the force-extension curve is flat.

Since inside the inhomogeneity the unzipping force threshold is higher than usual, as the tip gets closer to N, the observed force begins to rise. The rate of fork advancement with extension begins to decrease and the constant linear relation between $\langle n \rangle / N$ and f is lost. Eventually, as we drive the tip closer to N and correspondingly the applied force peaks, one of two things happen: the whole inhomogeneity melts all at once (small Δ) or the tip breaks through into the inhomogeneity (large Δ).

If Δ is sufficiently small (sawtooth case), the peak force that develops before the tip reaches N may be sufficient to drive the whole inhomogeneity to open simultaneously. This corresponds to the sawtooth force pattern, with the force growing linearly as the tip approaches N and then a sudden drop in the force when all Δ basepairs *cooperatively* open. When Δ is sufficiently large (ramp-plateau), the maximum force generated when the tip is just next to N (this force depends only on g_1) is not sufficient to produce concerted destabilization of Δ; instead the tip punctures into Δ and starts to *sequentially* unzip basepairs inside. This gives rise to force plateau inside the inhomogeneity. But, when a sufficient number of basepairs within N have been unzipped, the remaining Δ-domain *cooperatively* melts, producing the sudden drop in force marking the end of the distinct force signal.

4.2.3.2 Sawtooth vs. Ramp-Plateau "Phase Diagram"

Above, we discussed how the DNA's force response can be modulated by changing the length of the sequence inhomogeneity in the molecule. We could have kept Δ fixed and changed g_1, the binding strength within the inhomogeneity. In order to show how both Δ and g_1 alter the molecule's force behavior, we plot, in Fig. 4.4, a unzipping-'phase diagram', showing the

boundary between regions in the $\delta - \gamma$ plane where the force response signal is sawtooth or ramp-plateau.

The solid curve, running from top to bottom, in Fig. 4.4 is the mean-field 'phase-boundary,' separating one force-response from another, which is nothing but (4.8). Plotted alongside it are the exact, i.e. numerically computed, boundaries (dashed, line-dots, dots) for N=40, 200, and 1000. We calculate these lines by, first, fixing γ and then scanning across δ. For each δ, we calculate the molecule's force-response profile. When the molecule shows the sawtooth response, we find that increasing δ results in increase of the peak force of the force-kink. At the threshold between sawtooth and ramp-plateau response, the peak force becomes stationary and then, for further increases in δ, the force-kink begins to stretch resulting in a force-plateau. Therefore, we use the δ, where the force maximum becomes stationary with respect to changes in δ, to define the boundary between the two force traces. These stationary points are difficult to locate unambiguously, because the 'transition' between the two generic force traces occurred slowly over a wide range of δ, particularly for $\gamma \leq 0.2$. We have determined the transition between responses by hand just to establish the plausibility of experimentally observing it.

As the three successive exact transition boundaries show, with increasing N, these calculations converge to the mean field result. Indeed, the mean field results are the $N \to \infty$ limit of the numerical theory. As mentioned above, the situation for $\gamma \leq 0.2$ becomes problematic; it is difficult to pick out where the evolution of the peak force becomes stalled due to the weak force signal. We do not plot the lines for $\gamma \leq 0.2$; in this regime the force signal will likely be experimentally unobservable, being washed out by thermal fluctuations.

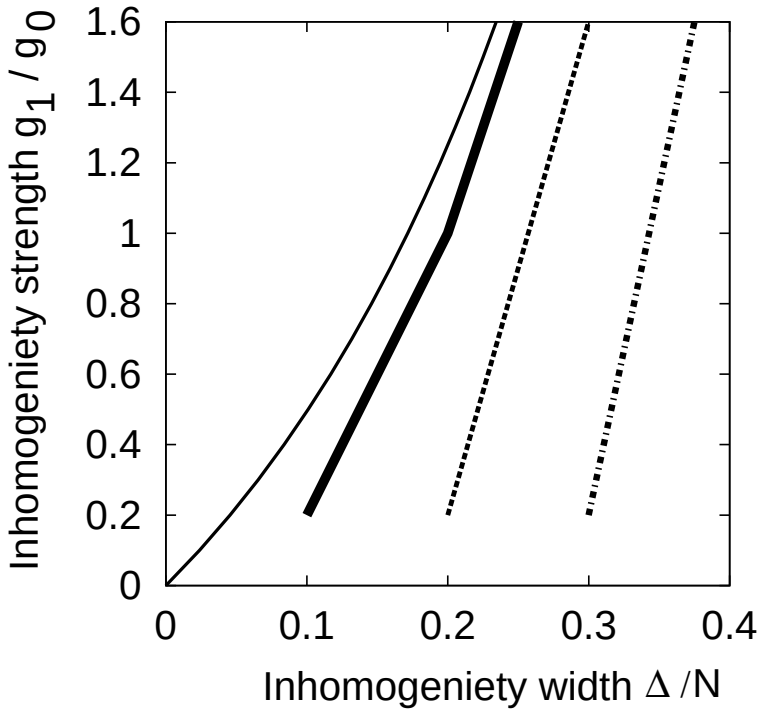


Figure 4.4. Plots indicating transitions from sawtooth (to the left of the boundary lines) to ramp-plateau (to the right of the boundary lines). The four curves correspond to the analytic mean field result (thin solid) and numerical calculations for $N = 1000$ (thick solid), $N = 200$ (dashed) and $n = 40$ (dot-dashed). With increasing N, the boundary lines approach the mean field result, consistent with the fact our mean field theory is the $N \to \infty$ limit of the exact theory. The numerical results are not shown below $g_1/g_0 = 0.2$ because the transition becomes ambiguous for values less than that.

4.3 Unzipping DNA With Bound Proteins

4.3.1 Model

In this section, we consider the unzipping force behavior of DNA with a protein bound to a specific position along it, instead of a sequence-inhomogeneity. We suppose that the protein contacts Δ bp when it binds to dsDNA. Binding of the protein to the dsDNA is thermodynamically favored. Let μ_{ds} denote the dsDNA-protein-binding enthalpy. We take μ_{ds} to be > 0 so that binding of the protein to dsDNA is thermodynamically favored, We also suppose that our protein can bind to ssDNA but does so with different (far lower) enthalpy. We denote by μ_{ds-ss}, also > 0, the *difference* between the dsDNA-protein-binding enthalpy and the ssDNA-protein-binding enthalpy. Since, in thermal equilibrium, the protein can spontaneously unbind from the DNA altogether, we use the binary variable p to denote whether the protein is bound ($p = 1$) or unbound ($p = 0$). And finally, N stands for the start, in basepairs, of the protein-attachment region, while n and X are defined as in the inhomogeneity model.

The energy of the DNA-protein complex is given by:

$$\frac{E(n,p)}{k_B T} = \frac{3X^2}{b^2 n} + n g_0 - p \left[\mu_{ds} - \mu_{ds-ss}\, h\left(\frac{n-N}{\Delta} \right) \right] \tag{4.12}$$

where $h(y)$ is the function defined in Eq. 4.2.

The first term is just the work done in maintaining an end-to-end distance of $2X$ of an entropic polymer made up of $2n$ repeat units. As before, this is just the two single-stranded arms of the DNA double-helix, considered as one polymer, obtained from sequentially unzipping

n basepairs. The second term is the energy necessary to break n complementary basepairs. The third and final term, which depends on p, is zero when the protein is not bound to the DNA. When the protein is bound to dsDNA, i.e. when the unzipping fork has not yet reached into the DNA domain that contacts the protein, the function $h(y) = 0$ and the DNA-protein complex is stabilized by μ_{ds}. But when the fork unzips basepairs past N, the stabilization enthalpy is reduced by $\mu_{ds-ss} h\left([n-N]/\Delta\right)$.

For large enough μ_{ds-ss}, the protein may come-off the DNA as the unzipping fork makes its way through the DNA in contact with the protein. Doing so results in the loss of the extra stabilization enthalpy gained by the system when the protein remains bound to ssDNA, but this might not be large enough to compensate for decrease in the elastic energy needed to unzip the Δ basepairs 'sheltered' by the bound protein.

For large enough μ_{ds-ss}, it becomes favorable for the protein to come off the DNA altogether when sufficient number of bases have been disrupted inside $[N, N+\Delta]$. On the other hand, if $\mu_{ds-ss} \approx \mu_{ds}$, we would expect the protein to remain stuck to the ssDNA even after the unzipping fork has disrupted almost all basepairs in the protein-binding domain of the DNA.

The partition function now involves summation over both p and n:

$$Z(X) = \sum_{p=0}^{1} \sum_{n=1}^{\infty} \exp\left(-\frac{3X^2}{b^2 n} - n g_0 + p\left[\mu_{ds} - \mu_{ds-ss}\; h\left(\frac{n-N}{\Delta} \right) \right] \right). \tag{4.13}$$

which, summing over p, becomes

$$Z(X) = \sum_{n=1}^{\infty} \exp\left[-\frac{3X^2}{b^2 n} - n g_0 + \log\left(1 + \exp\left[\mu_{ds} - \mu_{ds-ss}\; h(\frac{n-N}{\Delta})\right]\right)\right]. \tag{4.14}$$

The average force, $\langle f \rangle$, in $k_B T$ units is still

$$\langle f \rangle = -k_B T \frac{\partial \ln Z(X)}{\partial X}, \tag{4.15}$$

while the equilibrium position of the unzipping fork is

$$\langle n \rangle = Z^{-1} \sum_{n=1}^{\infty} n \exp\left[-\frac{3X^2}{b^2 n} - n g_0 + \log\left(1 + \exp\left[\mu_{ds} - \mu_{ds-ss}\; h(\frac{n-N}{\Delta})\right]\right)\right]. \tag{4.16}$$

and, finally, the probability, $P(X)$, that the protein is bound to the DNA is given by

$$P(X) = \langle p \rangle = Z^{-1} \sum_{n=1}^{\infty} \exp\left(-\frac{3X^2}{b^2 n} - n g_0 + \mu_{ds} - \mu_{ds-ss}\; h(\frac{n-N}{\Delta})\right). \tag{4.17}$$

4.3.2 Evaluation of Z(X) With Fluctuations

We evaluate $Z(X)$ by performing the sums indicated above. We take $b = 0.7$ nm (with the caveats of (75)), $\mu_{ds} = 5.0\ k_B T$, $\mu_{ds-ss} = 15.0\ k_B T$, $\Delta = 5$ bp, $g_0 = 2.5 k_B T$, and N = 40 bp. The values of μ_{ds} and Δ are representative of a wide assortment of DNA-binding proteins, while $\mu_{ds-ss} = 15.0\ k_B T$ is also a plausible estimate since most DNA-binding proteins tend to bind exclusively to either dsDNA or ssDNA.

Our results are presented in Fig. 4.5. We see that the force extension plot is very similar to results from inhomogeneity unzipping, studied in Section 4.2. Away from where the protein is bound (N=40 bp), the force required to unzip DNA is constant, $\approx 15pN$. But close to the protein, there is a force-kink, similar to sawtooth response of Section 4.2. Here, the presence of the bound protein obstructs the sequential passage of the unzipping fork through the basepairs in contact with the protein. As long as the protein is bound, the unzipping fork is stalled outside the region starting at N, corresponding to a build up of the force.

For sufficiently large force, one of two things can happen. The protein-binding domain of DNA can spontaneously unzip, with the protein still bound, but now to ssDNA instead to dsDNA. Since it is reasonable to assume that proteins that bind strongly to dsDNA will bind only weakly to ssDNA (if they bind at all), we have chosen the large value of $\mu_{ds-ss} = 15.0\ k_BT$. The second possibility is that the protein simply dissociates from the DNA, leaving the underlying DNA free to unzip under the imposed force. This is in fact what happens in our model, as can be seen from Fig. 4.5 inset where we show the progress of the unzipping fork, $\langle n \rangle / N$, with X and the probability, $P(X)$, for the protein to remain bound for a given X. Examining the results for $P(X)$, we see that in a narrow interval around $X \approx 30$ nm $P(X)$ drops sharply from near 1 to near 0. The dip in $\langle n \rangle / N$ vs. X, starting off where $P(X)$ just begins to drop-off, shows the fork stalled just outside N, thus allowing the force to build up to a level large enough to drive off the bound protein. The peak force to which the force builds up to is ≈ 18 pN. This is just large enough for the protein to unbind and, as we see from $\langle n \rangle / N$, the fork quickly reverts to linearly tracking X.

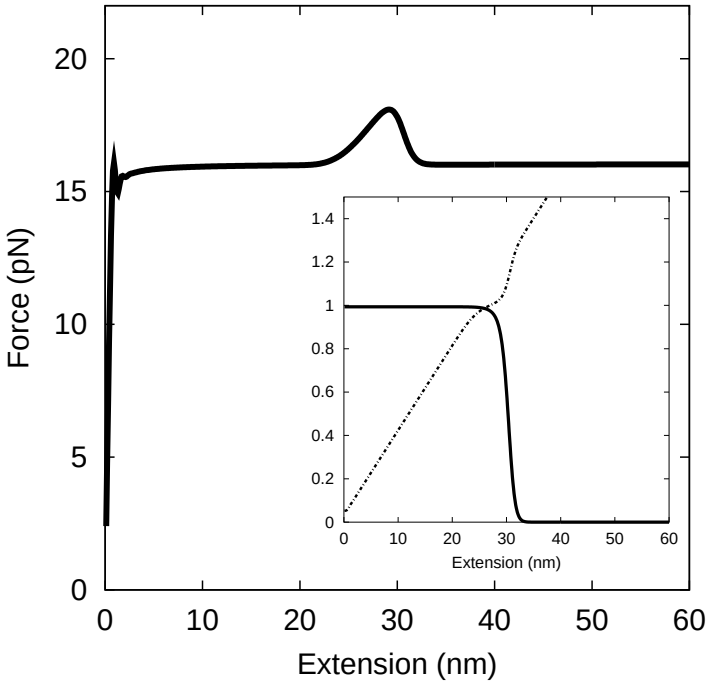


Figure 4.5. Force response from unzipping DNA with bound protein. (Inset) Probability of protein binding and mean position of fork vs. X. The protein binds to $\Delta = 5$ bp, $N = 40$ bp, $b = 0.7$ nm, $g_0 = 2.5\ k_BT$, $\mu_{ds} = 5.0\ k_BT$, and $\mu_{ds-ss} = 15\ k_BT$. The signal can be distinguished from the baseline force response for these choices of μ_{ds} and μ_{ds-ss}, which reflect typical protein binding free energies. The shape of the kink is less like a sawtooth than inhomogeneity unbinding case. (Inset) The probability (thick solid) drops from 1 to 0 where the fork (uneven dashed) deviates from linearly tracking X.

One marked difference between our results in this section and those of Section 4.2 is that while the force humps in the latter ended quite abruptly, with a steep decrease in the force after the unzipping fork successfully penetrated and thoroughly disrupted the basepairing inside the inhomogeneity, the force humps in the protein-binding case are quite smooth, both in the excess-force buildup phase and excess-force relaxation phase.

Our main aim in this section was to show how a propagating unzipping fork can drive off a bound protein. Moreover, a distinct and substantial (relative to noise) force signal, over and above what is seen in unzipping naked DNA, is obtained. By using reasonable values of protein-binding free energies and other physical parameters, we have shown that it is plausible that both these phenomena can be obtained in both in vivo and in vitro.

4.4 Unzipping Free Energy Barrier and Two-State Kinetics

We now return to the case of a simple sequence-inhomogeneity (Sec. 4.2). In both the sawtooth and ramp-plateau cases we saw that the final opening event generates a force jump. In the mean-field theory (Sec. 4.2.2) this force jump is a true discontinuity, but even when fluctuations are included (Sec. 4.2.3) a rather sharp final step occurs. At the extension where the force jump occurs, two well-separated values of unzipped bases are in equilibrium (i.e. have the same value of free energy). Since these states are locally stable, and have the same free energy, there must be a free energy barrier separating them.

This barrier height was explicitly calculated for the mean-field sawtooth (Eq. 4.9, Sec. 4.2.2.1) and was seen to scale with Ng_0. Fig. 4.6 shows the free energy as a function of number of unzipped bases, at such a sawtooth force-jump point, for a realistic case, $\Delta = 40$, $g_0 = 2.5$,

$g_1 = 1.25$, $N = 400$, for extension $X = 313.2$ nm. The figure shows that the barrier separating the 'closed' (left) and 'open' (right) states is 15 k_BT in height. By making N large, while keeping Δ/N fixed, this barrier can be made indefinitely large.

For large N, the barrier crossing time can be enormous. Since the time per barrier-crossing attempt will be on the order of the viscous time associated with a few-nm-long object (a few nsec), the time needed to transit the barrier of Fig. 4.6 at the coexistence extension will be on the order of e^{15} times the attempt time, on the order of 0.1 sec. Indeed, two-state switching between 25 bp open and closed RNA helix-loop structures has been observed to occur on a 1 sec timescale (76). Recent results of Bockelmann *et al* have observed similar switching during DNA unzipping (73).

4.5 Conclusions

Unzipping DNA with designed sequence inhomogeneities and bound proteins results in distinctive force signals which correspond to the disruption of the inhomogeneity and dissociation of the bound protein from the DNA, respectively. The force signal is a kink superposed on the baseline DNA unzipping response but discernible from it. For sequence inhomogeneities, the force kink is either sawtooth or ramp-plateau in shape, depending on g_1 and Δ for g_0 and N fixed. Adjusting g_1 or Δ, one can smoothly interconvert between the two responses. The parameter values where one force trace changes into another are characterized exactly in the mean field approximation and computed numerically when fluctuations are included. The barrier at the edge of the two force traces is shown to be $\sim 20\ k_BT$, indicating the fluctuations between the two free energy minima at the edge of force trace will be on the 1 sec time scale.

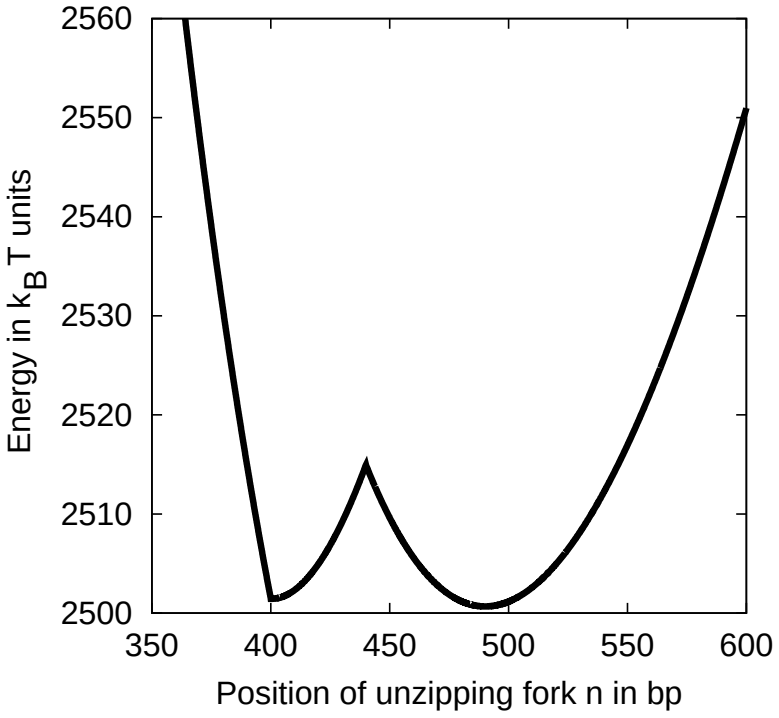


Figure 4.6. Free energy barrier for the sawtooth force response with X chosen so that the molecule is right at the end of the sawtooth. $X = 313.2$ nm, $N = 400$ bp, $\gamma = 0.5$, $\delta = 0.1$, $g_0 = 2.5\ k_BT$, $b = 0.7$ nm. The length, in basepairs, of the inhomogeneity $\Delta = 40$ bp. The barrier height is $\approx 15\ k_BT$, which implies that switching time between the two states ~ 0.1 sec. The magnitude of the calculate barrier height is close to what is expected from the analytic mean field formula for the barrier height.

A distinct force signal also results when we unzip a dsDNA molecule with a protein bound somewhere along it. We assume protein to bind to dsDNA much more strongly than ssDNA. As a result, when the unzipping fork approaches the protein-binding domain of the DNA, a large force has to be applied to drive the unzipping fork through the domain. This large force will make it favorable for the protein to dissociate from the DNA altogether, rather than remain bound to ssDNA (this binding, by assumption, is very weakly stabilized or not stable at all). The build up to this large force before the protein unbinds and the drop in force after the protein unbinds gives the force signal the shape of a kink, similar to the sawtooth trace of the sequence-inhomogeneity case. Using plausible numbers for the relevant theory parameters (μ_{ds}, μ_{ds-ss} and Δ), a large force kink over and above the usual unzipping response results, suggesting that experimentalists may be able to discern a protein unbinding event when unzipping DNA with bound proteins. In summary, our modeling suggests that the distinctive effects we have described, for both sequence inhomogeneities and bound proteins, should be observable in experiments. Existing protocols and experimental techniques are well-suited to directly testing our proposals.

CHAPTER 5

CONCLUSION

5.1 Introduction

The purpose of this chapter is to place this thesis into the broader landscape of scientific activity pursued by biophysicists, biologists and biochemists. We begin, in Section 5.2, by summarizing, in general terms, the principal results of the last three chapters. In Section 5.3, we indicate what experiments may shed further light on our theoretical proposals and identify promising work being undertaken in this direction. In Section 5.4, we address the important question whether biological micromechanics in general and our research in particular can speak to hypotheses about the working of cells and their constituents that interest biologists. In Section 5.5, we suggest future directions for the application of statistical mechanics to biologically motivated problems.

5.2 Summary of the Research Presented in Preceding Chapters

5.2.1 Mechanically Stressing DNA With Twist Constraints

With the ability to apply picoNewtons of force in a controlled manner to individual DNA (and now even certain protein) molecules, experimentalists have studied extensively how mechanical stress in the form of a stretching force or torque affects DNA structure. In Chapter 2, we theoretically analyzed how DNA's high-force i.e. $f > 10$ pN elasticity is modulated when we compel DNA to stretch but not untwist. Depending on how many extra turns were put into

or taken out of the DNA before pulling, a rich force response involving three new force plateaus ensues. In order to understand the effect of the twist constraint, we proposed a model of DNA where the native B-DNA structure can transform to a mixture of four other novel structural states (in general, the combinations involve only 2 states out of 4). We called these states S-DNA, which is 70% extended relative to B-DNA contour length and undertwisted, P-DNA which is also $\approx 70\%$ extended but overtwisted, sc-P-DNA which is a coiled-up (globular) version of P-DNA so has $\approx 0\%$ extension compared to B-DNA and is highly overtwisted, and 'Z'-DNA, which has an extension similar to B-DNA but with opposite twist. By forming combinations of these structural states, DNA can satisfy the imposed twist constraint and relax to a lower free energy. For instance, if DNA is stretched with no turns initially taken out but also not allowed to untwist ($\sigma = 0$), DNA transforms to a combination of S-DNA and P-DNA, since the under- and overtwists of each state, in the right ratio, will preserve the initial number of turns.

Using fits to the experimental data, we were able to determine the mean extension and twist of the 4 states. In fact, we were quite constrained in the choice of values of some the parameters by other experiments that independently estimated them. Our results allowed a precise determination of the mean twist rate of S-DNA.

Although the interpretation of DNA force plateaus as transitions between distinct structural states is generally accepted, the question of what is the secondary structure of S-DNA has been more controversial. Indeed, Rouzina *et al* (38) have argued that S-DNA is essentially single-stranded DNA (ssDNA). Our theoretical analysis provides further evidence that S-DNA is in fact distinct from single-stranded DNA. But beyond ascertaining mean extensions and twists,

single molecule experiments do not allow more refined knowledge of the the secondary structure of stress-transformed states. With the model calibrated, we were able to plot which DNA state or combination of states would be dominant for given (f, τ) and (f, σ) pairs. Some regions of the phase diagram have not yet been explored and this remains an active enterprise for some experimental groups.

5.2.2 Twisting Proteins Off DNA Using Transient Torque Propagation

In Chapter 3, we developed a model to study how torque pulses get propagated along DNA in order to explore the possibility that the torque produced by certain protein motors when they act on DNA can disturb the binding of distant proteins to DNA. The first step in the argument was an estimate of the torques that would be required to drive dissociation of a bound protein. Using a simple two-state equilibrium binding-unbinding model, we estimated this torque. Then, we argued for a particular form of the equation of motion. Amongst other approximations that went into the equation, the most significant was that we ignored hydrodynamic interactions, i.e. the non-local interaction between different regions of the DNA mediated by the 'wake' of the propagating twist pulse. With the equation of motion at hand, we specified the initial conditions for the two types of protein wrenches we considered in detail, RNApol and gyrase. We solved the equation of motion to determine the pulse shapes and torque profiles as a function of time position along the molecule. Using these results we estimated how far the broadening twist pulse can deliver sufficient torque to dislodge a bound protein.

Based on our results, we concluded that propagating twist pulses do not affect the binding of proteins that more more that ≈ 100 bp away. We point out that torque can have other uses, like partially stripping DNA off the surface of histone octamers.

5.2.3 Unzipping DNA With Sequence Inhomogeneities and Bound Proteins

In Chapter 4, we studied unzipping DNA molecules with built-in sequence inhomogeneities or bound proteins. These present an obstruction to the uniform unzipping of DNA because they locally affect the cohesiveness of the double-helix. Our goal in this research was to explore whether these obstruction would generate distinctive force signals through which their presence could be detected in the laboratory. We did this by constructing a simple model and then solved it numerically and in the mean-field limit. Using typical values for the extra cohesiveness of a designed sequence inhomogeneity or the extra binding free energy of a protein to dsDNA compared to ssDNA, we concluded that a significant force signal does in fact result. Moreover, in case of the sequence inhomogeneity, we found that when its length is varied, two different and experimentally distinguishable force signals result. We computed the parameter sets where we expect transformation from one force signal to another.

5.3 Experiments Relevant to Our Research

In Chapter 2, we computed the force-torque and force-twist phase diagrams for DNA. This corresponds to an experimental design that can simultaneously stretch DNA and independently apply controlled torques or twists. While the case of fixed twist has been experimentally realized, that of fixed torque has not. Our work suggests that the development of apparatus to apply controlled amounts of torque to DNA will allow for a complete exploration of DNA's varied

structural states. An experimental determination of this phase diagram will help confirm or suggest changes to the model's fitting parameters and computed phase diagram. One approach to this problem is to use drag forces on beads to apply controlled amounts of torque. This line of research is being pursued by the Bustamante and Cozzarelli groups (John Marko, private communication). Another possibility is to use observation of twist fluctuations to deduce the torque. Magnetic tweezers are well-suited to this task, although accurately measuring the fluctuations is a difficult technical problem.

Single molecule experiments have been quite successful in discovering novel structural states of DNA but have not provided much information about their secondary structure. Solving this problem remains an outstanding task. It is unlikely that mechanical stress-strain experiments can be modified to reveal this information although combination of mechanical experiments with spectroscopic measurements (e.g. using fluorescence transfer techniques to monitor the proximity of fluorophores with Angstrom precision) represents a promising direction for future research.

Although in Chapter 3 we did not find an appreciable effect on the binding of distant proteins from transiently generated torque pulses, we did suggest that this mechanism could be of relevance to other biological phenomena. Again, the development of micromanipulation techniques for controlling torque at the macromolecular scale will allow direct measurements to be made on how DNA+protein systems are effected by dynamically generated torque. Since the twist of DNA can be controlled, at least one experimental group, that of Joel Stavans, is pursuing how twisting DNA affects the binding of proteins (B. Schnurr, private communication).

The model of Chapter 4 was specifically conceived to test whether the unzipping of DNA with a segment more cohesively bound than average would display an extraordinary force signal from the disruption of that inhomogeneity. We indeed found that to be the case. All the experimental prerequisites already exist to perform a direct test of our results and Carlos Bustamante's group has expressed interest in pursuing the experiments (C. Bustamante, private communication). The same holds true for unzipping DNA with proteins bound to specific sequences. A distinct force signal, corresponding to the dissociation of the protein from the DNA should be observable.

5.4 Biological Relevance of Our Research

The study of high-force elasticity of DNA may, at first, seem irrelevant to the workings of the cell. This is not the case: as Leger *et al* (20) found, the protein RecA applies enormous forces $\approx$ 100 pN (relative to 0.1 pN, the force scale associated with DNA entropic elasticity) as it polymerizes on DNA. Other examples of large-force generating proteins are RNApol ($\approx$ 40 pN) (60) and DNApol ($\approx$ 20 pN) (75). Chromatin assembly also involves forces $\sim$ 10 pN (79; 65). Understanding the structural states accessible to DNA under these forces will contribute to a better appreciation of the biochemical mechanisms underlying transcription (RNApol), replication (DNApol), and mitosis (chromatin assembly).

Recently, two papers (52; 80) have addressed the problem of how RNApol transcribes through the DNA that is stuck to the surface of the histone octamer. One mechanism suggested has been that RNApol uses the torque it generates to locally drive unpinning of the DNA from

the nucleosome (DNA+histone octamer). As we discuss in the conclusion to Chapter 3, our model of torque propagation is relevant to this mechanism.

Other papers have addressed the problem of chromosome remodeling (52). This refers to the dynamic and local unfolding of the chromosome to allow transcription of the DNA that was previously folded into the chromosome and thus inaccessible to reading by RNApol. Again, we have suggested that torque generated by known chromsome remodellers may provide at least a partial explanation of this phenomenon (52).

Our work on unzipping sequence inhomogeneities is less relevant to biological activity in cells but that of driving proteins off DNA by unzipping DNA is quite relevant to the understanding of the mechanism of a number of interesting proteins. For instance, RNApol is known to locally unzip DNA where it is transcribing it. Binding of nearby proteins may be destabilized and in this way RNApol may clear a path for itself on the DNA. Many other proteins, including DNA helicases and the chromatin remodellers mentioned above, are known to create ssDNA where they act, which could be a generic way of nonlocally affecting the binding of other proteins.

5.5 Future Directions

At the level of DNA, life can be reduced to transcription, translation, repair, regulation, recombination, and replication. Transcription is the production of mRNA from a stretch of DNA that codes for a protein. Translation is the use of mRNA to construct the corresponding protein. This involves reading the mRNA and arranging the protein monomers (the amino acids) into the corresponding sequence. DNA can develop spontaneous breaks or can become otherwise deformed after some genetic event (like transcription). These deformations have to

be ironed out before the next round of transcription or some other process can properly begin; collectively, this activity is called DNA repair. Regulation refers to modulation of protein production in cells in response to internal clocks or external stimuli. Not all regulatory activity occurs on DNA (more generally, DNA and RNA, i.e. nucleic acids) but ultimately, they can be traced back to signals being read off DNA by regulatory proteins. Recombination is a way for an organism to augment its DNA by incorporating non-native DNA. Finally, replication is the production of an identical copy of the cell's DNA. When the cell divides in two, each daughter cell receives a copy of the parent cell's DNA.

All these processes are studied intensively by molecular biologists. A higher level of organization is the cell, which results from the proper working of genetic processes described above. Cell biologists address questions which pertain to processes at the cellular level; the emphasis shifts from studying solitary genetic events at the DNA level to studying complexes of proteins carrying out tasks needed for simple cellular functions.

Condensed matter physics can contribute significantly to the growth of knowledge at both these levels of description. While the number of experimental studies using single molecule techniques and other methods from physics has grown explosively, we shall only focus on theoretical problems that arise from operation of cells at both levels of description. Here, our goal is simply to suggest to the reader the breadth of problems in this field.

Single-stranded DNA arises frequently in various genetic processes. The lack of detailed knowledge about the low-force elasticity of ssDNA is a serious deficiency. Numerous self-interactions between mismatched basepairs is responsible for the complicated stretching be-

havior of ssDNA (75). Understanding ssDNA's low-force mechanical response is an urgent requirement.

Certain protein molecules called restriction enzymes cut DNA at specific sites flagged by a particular sequence. Since there are no long-range interactions between the protein and this site, how does the protein search for its binding site? This is a specific example of more general search problems encountered at the DNA level of description. Many proteins, e.g. regulatory proteins, bind to particular sites along the DNA specified only by its sequence. Is there a general strategy for solving this problem?

The search problem in restriction enzymes has been the subject of extensive research by biochemists (81; 82). Three possibilities have been suggested. (1) The enzyme randomly binds to the DNA and executes a 1D random walk search. (2) The protein binds at random and if it is not already at its binding site, detaches from the DNA, executes a 3D random walk in the solution, and binds somewhere else. It does this until it finds the binding site. (3) The protein attaches to a random site on the DNA, then executes a short 1D random search; if it doesn't find its target, it detaches and then quickly rebinds somewhere close by (hopping). Biochemists have devised ingenious indirect experiments to deduce the mechanism but theoretical analysis is necessary to rule out alternatives.

Most proteins are present in the cell in very low concentrations, ~ 1 nM. Are there significant deviations from Boltzmann statistics in the kinetics of their reactions? Statistical fluctuations in genetic processes may require novel probabilistic description. Simulations and models of

low-reactant-concentration reactions may help shed light on whether deviations from canonical kinetics are significant (83).

In eukaryote cells, the mRNA is produced in the nucleus, a compartment in the cell with small holes on its surface that selectively allow certain chemicals to enter or leave the nucleus, but translation occurs outside the nucleus. A question of great interest to biologists in whether the mRNA diffuses out of the nucleus or is carried through the nuclear pores by an active transport mechanism (84). Clearly, if the mRNA were to diffuse out, there would be a time scale associated with it arrival at the inner nuclear membrane, search for a suitable pore to exit out of, and the reorientation necessary for the mRNA polymer to strain itself through the pore. If the time scale is too long, then diffusion can be ruled out as a candidate process.

Our research also points to other paths that can be explored. There is inadequate understanding of the secondary structure of the novel structural states discovered by micromanipulation studies. For example, nothing close to atomic-level resolution of B-DNA structure exists for S-DNA. Devising experimental means to gauge the true secondary structures of these new states is an interesting problem. On unzipping DNA, a crucial question is whether we can neglect the buildup of torque that results when we unzip it. As we unzip from one end, the DNA will tend to spin about its helical axis. If it is prevented from doing so, say be sticking both 3' and 5' strands at the other end to an immovable surface, tension will buildup in the chain. How this tension affects further unzipping of the DNA needs to be explored (74). Since DNA in cells has numerous obstructions to free rotation and given that transiently generated ssDNA is ubiquitous, this problem has great biological relevance (50).

The prospects for applying theoretical physics to molecular and cellular biology appear promising. The advent of experimental techniques that allow mechanical and other measurements to be made on isolated single- or few-molecule systems and their increasing adoption means that the task of formulating quantitative models that can explain the experimental data and guide future experimental work is just beginning.

CITED LITERATURE

1. O. T. Avery, C. M. Macleod, and M. McCarty J. Exp. Med. **79**, 137 (1944).

2. E. W. Chargaff, Experientia **6**: 201 (1950).

3. L. Pauling and R. B. Corey, Nature **171**, 346 (1953).

4. F. H. C. Crick and J. D. Watson, Nature **171**, 737 (1953).

5. M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson, Nature **171**, 738 (1953); R. E. Franklin and R. G. Gosling Nature **171**, 740 (1953); R. E. Franklin and R. G. Gosling Nature **172**, 156 (1953).

6. F. H. C. Crick, G. Gamow, J. D. Watson, and S. Brenner, Nature **192**, 1227 (1961).

7. J. Darnell, H. Lodish, and D. Baltimore *Molecular Cell Biology* (Scientific American Books, New York, 1990), pp. 68-72.

8. K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, Proc. Natl. Acad. Sci. USA **83**, 3746 (1986).

9. M. H. F. Wilkins, R. G. Gosling and W. E. Seeds, Nature **167**, 759 (1951).

10. S. B. Smith and A. J. Bendich, Biopolymers **29**, 1167 (1990).

11. J. F. Marko and E. D. Siggia, Macromolecules **28**, 209 (1995).

12. P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron, Science **271**, 792 (1996).

13. S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).

14. D. Bensimon, A. J. Simon, V. Croquette, and A. Bensimon, Phys. Rev. Lett. **74**, 4754 (1995).

15. H. Clausen-Schaumann, M. Rief, C. Tolksdorf, and H. E. Gaub, Biophys. J. **78**, 1997, (2000).

16. M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Nat. Struct. Biol. **6**, 345 (1999).

17. D. B. Nikolov, H. Chen, E. Halay, A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley, Nature **377**, 119 (1995).

18. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. Richmond, Nature **389**, 251 (1997).

19. A. Stasiak, E. di Capua, and Th. Koller, J. Mol. Biol. **151**, 557 (1981).

20. J. -F. Léger, J. Bourdieu, D. Chatenay, and J. F. Marko, Proc. Natl. Acad. Sci. USA **95**, 12295 (1998).

21. T. R. Strick, J. -F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, Science **271**, 1835 (1996).

22. P. Cluzel, Ph.D. thesis, Université Pierre et Marie Curie, Paris, 1996.

23. J. -F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, Phys. Rev. Lett. **83**, 1066 (1999)

24. J. F. Marko and E. D. Siggia, Science **265**, 506 (1994).

25. D. Moroz and P. Nelson, Proc. Natl. Acad. Sci. USA **94**, 14418, (1997).

26. C. Bouchiat and M. Mezard, Phys. Rev. Lett. **88**, 1556 (1998).

27. A. V. Vologodskii and J. F. Marko, Biophys. J. **73**, 123 (1997).

28. T. Strick, V. Croquette, and D. Bensimon, Proc. Natl. Acd. Sci. U.S.A. **95**, 10579 (1998).

29. S. Cocco and R. Monasson, Phys. Rev. Lett. **83**, 5178 (1999).

30. J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette, Proc. Natl. Acad. Sci. U.S.A. **95**, 14152 (1998).

31. S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).

32. T. T. Perkins, S. R. Quake, D. E. Smith, and S. Chu, Science **264**, 822 (1994).

33. T. T. Perkins, D. E. Smith, R. G. Larson, and S. Chu, *ibid.* **268**, 83 (1995).

34. C. Bustamante, J. F. Marko, S. B. Smith, and E. D. Siggia, Science **265**, 1599 (1994).

35. A. Vologodskii, Macromolecules **27**, 5623 (1994).

36. A. Lebrun and R. Lavery, Nucl. Acid Res. **24**, 2260 (1996).

37. A. Herbert and A. Rich, J. Biol. Chem. **271**, 11595 (1996).

38. I. Rouzina and V. Bloomfeld, Biophys. J. **80**, 882 (2001).

39. F. B. Fuller, Proc. Natl. Acad. Sci. U.S.A. **68**, 815 (1971).

40. N. R. Cozzarelli, T. C. Boles, and J. White, in *DNA Topology and its Biological Effects*, edited by N. R. Cozzarelli and J. C. Wang (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1990), Chap. 4.

41. S. B. Smith, Y.Cui, A. C. Hausrath, and C. Bustamante, Biophys. J. **68**, A250 (1995).

42. J. F. Marko, Phys. Rev. E **57**, 2134 (1998).

43. A. V. Vologodskii *Topology and Physics of Circular DNA* (CRC Press, Boca Raton, Florida, 1992).

44. T.R. Strick, D. Bensimon, V. Croquette, *Micro-mechanical measurement of the torsional modulus of DNA*, in "Structural Biology and Functional Genomics", E.M. Bradbury, S. Pongor (eds), NATO Science Series 3, Kluwer Academic Publishers, Boston (1999).

45. A. Sarkar and J. F. Marko, Phys. Rev. E **63**, 61909 (2001).

46. J. F. Marko and E. D. Siggia, Phys. Rev. E **52**, 2912 (1995).

47. M. D. Barkley and B. H. Zimm, J. Chem. Phys. **70**, 2991 (1979).

48. S. A. Allison and J. M. Schurr, Chem. Phys. **41**, 35 (1979).

49. C. Levinthal and H. R. Crane, Proc. Natl. Acad. Sci. USA **42**, 436 (1956).

50. P. Nelson, Proc. Natl. Acad. Sci. USA **96**, 14342 (1999).

51. L. F. Liu and J. C. Wang, Proc. Natl. Acad. Sci. USA **84**, 7024 (1987); J. C. Wang and A. S. Lynch, Curr. Opin. Genes Dev. **3**, 746 (1993).

52. A. Flaus and T. Owen-Hughes, Curr. Opin. Genes Dev. **11**, 148 (2001).

53. R. Schleif, Ann. Rev. Biochem. **61**, 199 (1992).

54. T. Hirano, Ann. Rev. Biochem. **69**, 115 (2000).

55. J. F. Marko and E. D. Siggia, Biophys. J. **73**, 2173 (1997).

56. T. Strick, J. Allemand, D. Bensimon and V. Croquette, Biophys. J. **74**, 2016 (1998).

57. R. W. Cotton and B. A. Hamkalo, Nucleic Acids Res. **9**, 445 (1981); H. Ausio, N. Borochov, D. Seger, and H. Eiseberg, J. Mol. Biol. **176**, 77 (1984a); H. Ausio, N. Borochov, D. Seger, and H. Eiseberg, J. Mol. Biol. **177**, 373 (1984b).

58. W. T. Hsieh, P. A. Whitson, K. S. Matthews, and R. D. Wells, J. Biol. Chem **262**, 14583 (1987); L. Finzi and J. Gelles, Science **267** 378 (1995); J. G. Kim, Y. Takeda, B. W. Matthews, and W. F. Anderson, J. Mol. Biol. **196**, 149 (1987).

59. Y. Hirada, O. Ohara, A. Takatsuki, H. Itoh, N. Shimamoto, and K. Kinoshita Jr., Nature **409**, 113 (2001).

60. H. Yin, M.D. Wang, K. Svovoda, R. Landick, S. Block, J. Gelles, Science **270**, 1653 (1995).

61. A. Sarkar, J. F. L éger, D. Chatenay and J. F. Marko, Phys. Rev. E **63**, 51903 (2001).

62. S. C. Kampranis, A. D. Bates, A. Maxwell, Proc. Natl. Acad. Sci. **96** 8414 (1999); also see http://www.cid.csic.es/homes/roca/ and references therein.

63. R. Kamien, Eur. Phys. J. B**1**. 1 (1998); C. Woglemuth, T. Powers and R. Goldstein, Phys. Rev. Lett. **84**, 1623 (2000).

64. Y. Cui and C. Bustamante, Proc. Natl. Acad. Sci. **97**, 127 (2000).

65. M. L. Bennink, S. H. Leuba, G. H. Leno, J. Zlatanova, B. G. de Grooth, J. Greve, Nat. Struct. Biol. **8**, 606 (2000).

66. B. M. Ali, R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi, J. Stavans, Proc. Natl. Acad. Sci. **98**, 10658 (2001).

67. M. Vignali, A. H. Hasssan, K. E. Neely, J. L. Workman, Mol. Cell. Biol. **20**, 1899 (2000); R. E. Kingston and G. J. Narlikar, Genes. Dev. **13**, 2339 (1999).

68. K. Polach and J. Widom, J. Mol. Biol. **254**, 130 (1995).

69. T. R. Strick, D. Bensimon, and V. Croquette, Genetica **106**, 57 (1999).

70. B. Essevaz-Roulet, U. Bockelmann and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11935 (1997).

71. U. Bockelmann, B. Essevaz-Roulet and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).

72. U. Bockelmann, B. Essevaz-Roulet and F. Heslot, Phys. Rev. E **58**, 2386 (1998).

73. U. Bockelmann, P. Thomen, B. Essevaz-Roulet, V. Viasnoff and F. Heslot, Biophys. J. **82**, 1537 (2002).

74. P. Thomen, U. Bockelmann, and F. Heslot, Phys. Rev. Lett. **88**, 248102 (2002).

75. B. Maier, D. Bensimon, and V. Croquette, Proc. Natl. Acad. Sci. USA **97**, 12002 (2000).

76. J. Liphardt, B. Onoa, S. B. Smith, I. Tonoco, C. Bustamante, Science **292**, 733 (2001).

77. S. Cocco, R. Monasson and J. F. Marko, Proc. Natl. Acad. Sci. U.S.A. **98**, 8608 (2001).

78. S. Cocco, R. Monasson and J. F. Marko, Phys. Rev. E **65**, 041907 (2002).

79. B. Ladoux, J. P. Quivy, P. Doyle, O. du Roure, G. Almouzni, and J. L. Viovy, Proc. Natl. Acad. Sci. U.S.A. **97**, 14251 (2000).

80. G. Felsenfeld G, D. Clark, and V. Studitsky, Biophys. Chem. **86** 231 (2000).

81. N. Stanford, M. D. Szczelkun, J. F. Marko, and S. E. Halford, EMBO J. **19**, 6546 (2000).

82. O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6929 (1981).

83. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002); J. M. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **99**, 5988 (2002).

84. J. C. Politz and T. Pederson, J. Struct. Biol. **129**, 252 (2000).

VITA

Name: Abhijit Sarkar

Email: asarkar@callan.phy.uic.edu

Education: B.Sc. in Physics from Jamia Millia Islamia,
New Delhi, India 1993
M.Sc. in Physics from Jamia Millia Islamia,
New Delhi, India 1995

Teaching Experience: Physics Department Teaching Assistant
for undergraduate courses including molecular
biophysics, also graduate statistical mechanics

Professional Membership: American Society for Cell Biology 2000-present
American Physical Society 2001-present

Conference Presentations: APS March Meeting '00,'01 & '02: contributed talks
ASCB 2000: poster
PMMB VII 2002: poster

Publications: Structural transitions of a twisted and stretched DNA molecule,

J. F. Leger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu,

D. Chatenay and J. F. Marko,

Phys. Rev. Lett. **83**, 1066 (1999).

Structural transitions in DNA driven by force and torque,

A. Sarkar, J. F. Leger, D. Chatenay, J. F. Marko,

Phys. Rev. E **63**, 51903 (2001).

Removal of DNA-bound proteins by DNA twisting,

A. Sarkar and J. F. Marko,

Phys. Rev. E **63**, 61909 (2001).

Dynamics of chromosome compaction during mitosis,

A. Sarkar, S. Eroglu, M. Poirier, A. Nemani,

P. Gupta, and J. F. Marko,

Expt. Cell Res. **277**, 48 (2002).

Unzipping DNA with built-in sequence inhomogeneities and bound proteins

proteins,

A. Sarkar and J. F. Marko,

Phys. Rev. E, Submitted (2002).