

INTRODUCTION TO SINGLE-DNA MICROMECHANICS
(JUNE 2, 2005)

John F. Marko

*Department of Physics, University of Illinois at Chicago ,
Chicago, IL 60607-7059, USA*

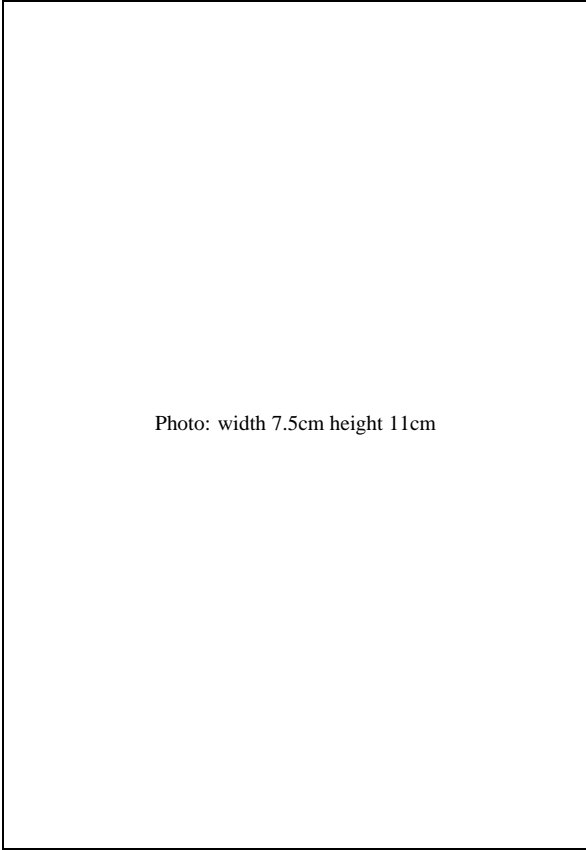


Photo: width 7.5cm height 11cm

Contents

1. Introduction	5
2. The Double Helix is a Semiflexible Polymer	7
2.1. Structure	7
2.2. DNA Bending	9
2.2.1. Discrete-segment model of a semiflexible polymer	10
2.2.2. Bending elasticity and the persistence length	12
2.2.3. End-to-end distance	13
2.2.4. DNA loop bending energies	14
2.2.5. Site-juxtaposition probabilities	15
2.2.6. Permanent sequence-driven bends	16
2.3. Stretching out the double helix	17
2.3.1. Small forces ($< k_B T/A = 0.08$ pN)	18
2.3.2. Larger forces ($> k_B T/A = 0.08$ pN)	18
2.3.3. Free energy of the semiflexible polymer	19
2.3.4. Really large forces (> 10 pN)	21
3. Strand Separation	22
3.1. Free-energy models of strand separation	23
3.1.1. Sequence-dependent models	24
3.1.2. Free energy of internal 'bubbles'	25
3.1.3. Small internal bubbles may facilitate sharp bending	26
3.2. Stretching single-stranded nucleic acids	26
3.3. Unzipping the double helix	29
3.3.1. Effect of torque on dsDNA end	31
3.3.2. Fixed extension versus fixed force for unzipping	32
4. DNA Topology	33
4.1. DNA supercoiling	33
4.1.1. Twist rigidity of the double helix	33
4.1.2. Writhing of the double helix	34
4.1.3. Simple model of plectonemic supercoiling	36
4.2. Twisted DNA under tension	39
4.3. Forces and torques can drive large structural reorganizations of the double helix	41
4.4. DNA knotting	42
4.4.1. Cells contain active machinery for removal of knots and other entanglements of DNA	43
4.4.2. Knotting a molecule is surprisingly unlikely	43
4.4.3. Condensation-resolution mechanism for disentangling long molecules	44
5. DNA-Protein Interactions	45
5.1. How do sequence-specific DNA-binding proteins find their targets?	46
5.1.1. Three-dimensional diffusion to the target	46
5.1.2. Nonspecific interactions can accelerate targeting	47
5.2. Single-molecule study of DNA-binding proteins	48
5.2.1. DNA-looping protein: equilibrium 'length-loss' model	48

5.2.2. Loop formation kinetics	49
5.2.3. DNA-bending proteins	50
5.2.4. Analytical calculation for compaction by DNA-bending proteins	54
5.2.5. Effects of twisting of DNA by proteins	55
5.2.6. Surprising results of experiments	57
6. Conclusion	58
References	59

1. Introduction

Over the past ten years new ‘single-molecule’ techniques to study individual biomolecules have been developed. Many of the new approaches being used are based on micromanipulation of single DNAs, allowing direct study of DNA, and enzymes which interact with it. These lectures focus on mechanical properties of DNA, crucial to the design and interpretation of single-DNA experiments, and to the understanding of how DNA is processed and therefore functions, inside the cell.

A seminal example of a single-DNA experiment was the measurement of the force exerted by RNA polymerase [1], done by Jeff Gelles, Steve Block and co-workers. Gene sequences in DNA are ‘read’ by RNAPol, which synthesizes an RNA copy of a DNA sequence. The experiment (Fig. 1) revealed that as a RNAPol moves along a DNA, it is able to pull with up to 30×10^{-12} Newtons, or 30 piconewtons (pN) of force. In the single-molecule world, this is a hefty force: the motor proteins which generate your muscle contractions, called *myosin* generate only about 5 pN.

RNAPol is an example of a *processive enzyme* which works rather like a macroscopic engine, using stored chemical energy to catalyze not only the synthesis of RNA, but also converting some of that energy to mechanical work. This mechanical work is absolutely necessary for RNAPol’s function: it must move ‘processively’ along the DNA double helix in order to make a faithful copy of DNA. Another important DNA-processing enzyme is *DNA polymerase* which is able to synthesize a copy of a DNA strand; this is important in cell division, since in order to make a copy of itself, a cell must faithfully copy its chromosomal DNAs. Proper understanding of this kind of DNA-processing enzyme machinery requires us to first understand the mechanical properties of DNA itself.

DNA has extremely interesting and unique polymer properties. In double helix form it is a water-soluble, semiflexible polymer which can be obtained in gigantic lengths. We often measure DNA length in ‘bases’ or ‘base pairs’; each DNA base of nm dimensions encodes one of four ‘letters’ (A, T, G or C) in a genetic sequence. A human genome contains 3×10^9 bases divided into 23 chromosomes. Each chromosome therefore contains a DNA roughly 10^8 bases long; chromosomal DNAs are the longest linear polymers known. Furthermore, the

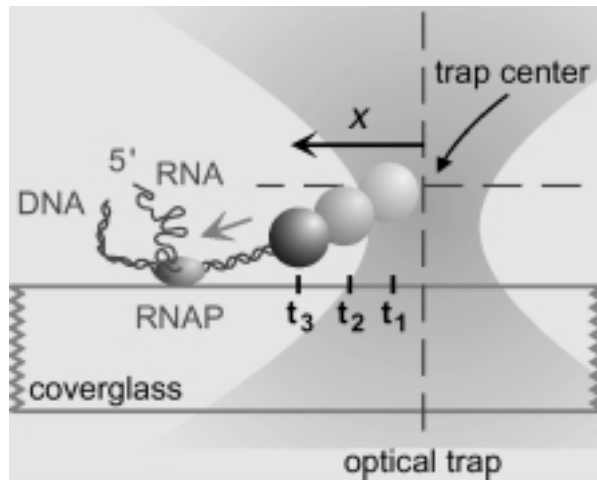


Fig. 1. Sketch of single-DNA experiment of Wang et al to measure force generated by RNA polymerase (reproduced from Ref. [1]). The polymerase is attached to the glass, and the DNA is pulled through it. A bead at the end of the DNA is held in a laser trap; deflection of the bead in the trap indicates the applied force.

base-paired complementary-strand structure of the double helix offers up new types of polymer physics problems, which we will explore in these lectures.

I note some physical scales relevant to these lectures. The fundamental length scale of molecular biology is the nanometer (nm); this is a distance several atoms long, the size of a single nucleic acid (DNA or RNA) base (the basic unit of information in molecular biology), or a single amino acid (the elementary unit of proteins). Cells must maintain their nm-scale organizational structure at room temperature: this requires that components be acted on by forces of roughly $1 k_B T / \text{nm} = 4 \times 10^{-21} \text{ J} / 10^{-9} \text{ m} = 4 \times 10^{-12} \text{ N} = 4 \text{ pN}$.

We can expect the forces generated by single mechanoenzymes to be on the pN scale. If RNAPol generated smaller forces than this, it would get pushed around by thermal forces, and would be unable to read DNA sequence in a processive manner.

Problem 1: Consider a molecule localized by a harmonic force $f = -kx$. What force constant is necessary to have $\langle x^2 \rangle = 1 \text{ nm}^2$? What is the typical (root-mean-squared) force applied to the molecule in this case? Repeat this calculation if the localization is done to 1 \AA (atomic) accuracy.

Problem 2: Consider a nanowire made of some elastic material, with circular cross-section of diameter d . In any cross-section of the wire, what will be the

typical elongational stress (force per area) due to thermal fluctuations? What does this suggest about the Young modulus of the material that you might try to use to make a nanowire?

Problem 3: Consider a random sequence of DNA bases 48502 bases long. How many times do you expect to find the sequences AATT, ACTAGT and GGC-CGGCC?

2. The Double Helix is a Semiflexible Polymer

The double helix (sometimes called the ‘B-form’) is taken by DNA most of the time in the cell. This form of DNA has a regular helical structure with remarkably uniform mechanical properties. This section will focus on the bending flexibility of the double helix, which gives rise to polymer elasticity effects which are of biological importance, and accessible in biophysical experiments.

2.1. Structure

The double helix is made of two DNA polymer molecules. Each DNA polymer is a string of four interchangeable types of ‘monomers’, which can be strung together in any sequence. The monomers each carry a *sugar-phosphate backbone* element: these are covalently bound together in the polymer. However, each monomer also carries, attached to the sugar (which is deoxyribose), one of four possible ‘bases’: either adenine (A), thymine (T), guanine (G) or cytosine (C).

The length of each backbone unit is about 0.7 nm when extended. The bases are each about 1 nm wide, and 0.3 nm thick.

The structure of each polymer gives it a definite ‘polarity’. It is conventional to report DNA sequence along each strand in the direction read by RNA polymerase, from 5’ to 3’ (the number refer to carbon atoms in the deoxyriboses). Often people just omit the leading 5’: in this case it is almost always in 5’ to 3’ order.

The bases have shapes and hydrogen-bonding sites which make A-T and G-C bonds favorable, under the condition that the two strands are anti-aligned (see sketch). Such *complementary strands* will bind together, making inter-strand hydrogen bonds, and intra-strand *stacking interactions*. The stacking of the bases drives the two strands to twist around one another to form a helix, since each base is only about 0.3 nm thick while the backbones are roughly 0.7 nm long per base.

We can roughly estimate the helix parameters of the double helix, assuming that the backbones end up tracing out a helical path on the surface of a cylinder of radius 1 nm (the bases are 1 nm across). Since each base is 0.34 nm thick, and

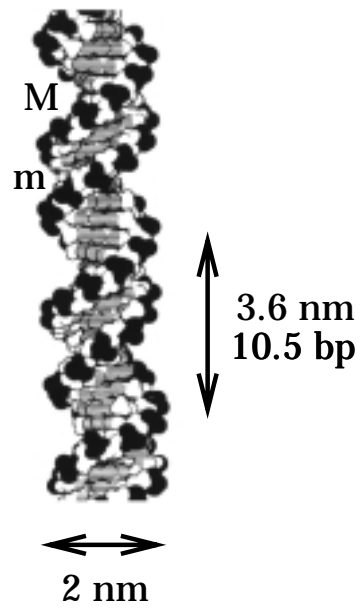


Fig. 2. DNA double helix structure. The two complementary-sequence strands noncovalently bind together, and coil around one another to form a regular helix. The two strands can be seen to have directed chemical structures, and are oppositely directed. Note the different sizes of the major (M) and minor (m) grooves. The helix repeat is 3.6 nm, and the DNA cross-sectional diameter is 2 nm. DNA image reproduced from Ref. [2].

traces a helix contour length of 0.7 nm, the circumference occupied by each base is $\sqrt{0.7^2 - 0.34^2}$ nm = 0.61 nm. Dividing the total circumference (6.3 nm) by this indicates that the double helix contains about 10.3 base pairs (bp) per helical turn. This is very close to the number usually quoted of 10.5 bp/turn; the double helix therefore makes one turn for every $10.5 \times 0.34 = 3.6$ nm.

The B-form double helix is *right-handed*, with the two backbones oriented in opposite directions. This means that there are two types of ‘grooves’ between the backbones: these are in fact rather different in size in B-DNA, and are called the ‘major groove’ and the ‘minor groove’.

We should remember the following conversion factor for the double helix, length: 1 bp = 0.34 nm, thus each micron (1000 nm) worth of DNA contains about 3000 bp = 3 kilobp (kb); one whole human genome is thus close to 10^9 nm = 1 m in length.

Problem 4: Consider a hypothetical form of double helix formed of two *parallel-orientation* strands. Describe the grooves between the backbones.

Problem 5: A student proposes that for two complementary-sequence biological DNA strands, there must be an equivalent form of double helix, of free energy equal to B-DNA, which is instead left-handed. Explain under what circumstances of symmetry of the monomers this conjecture can be expected to be true. Based on textbook pictures of the base and backbone chemical structures, what is your conclusion?

Problem 6: Do you expect the average helix repeat (base-pairs per turn) of the double helix to increase or decrease with increased temperature?

Problem 7: Estimate the ‘Young modulus’ of the double helix, using the assumption that the single-base helix parameters described above apply to room-temperature structure of DNA to roughly 1 Å precision.

2.2. DNA Bending

Although the structure of DNA is often presented in books as if it is static, at room temperature and in solution the double helix undergoes continual thermally excited changes in shape. Per base pair, the fluctuations are usually small displacements (a few degrees of bend, 0.03 nm average separations of the bases) but over long stretches of double helix, they build up to significant, thermally excited random bends. Note that rarely, more profound thermally-excited disturbances of double helix structure (e.g., transient unbinding of base pairs) can be expected to occur.

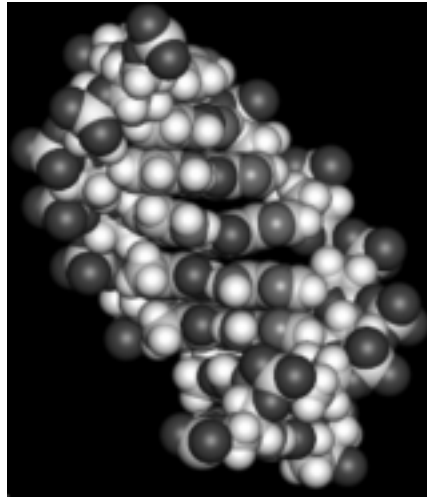


Fig. 3. Molecular-dynamics snapshot of typical DNA conformation for a short 10 bp molecule in solution at room temperature. Reproduced from Ref. [3].

2.2.1. Discrete-segment model of a semiflexible polymer

We can make a simple one-dimensional lattice model of thermally excited bending fluctuations. If we describe our DNA with a series of tangent vectors $\hat{\mathbf{t}}_j$ that indicate the orientation of the *center axis* of the molecule, then the bending energy associated with two adjacent tangents is $E/(k_B T) = -a \hat{\mathbf{t}}_j \cdot \hat{\mathbf{t}}_{j+1}$.

The dimensionless constant a describes the molecule's bending rigidity: $a \gg 1$ means very rigid (adjacent tangent vectors point in nearly the same direction); $a < 1$ means very floppy. We'll talk more about a below, but just to give a rough idea of the stiffness of the DNA double helix, if we consider adjacent base pairs to be described by successive tangents, the value of a to use is about 150.

Problem 8: Estimate the bend between two adjacent tangent vectors excited thermally in the limit $a \gg 1$; your result should be of the form

$$\langle |\hat{\mathbf{t}}_j - \hat{\mathbf{t}}_{j+1}|^2 \rangle \propto a^p$$

where p is a power. Hint: $\frac{1}{2} |\hat{\mathbf{t}}_j - \hat{\mathbf{t}}_{j+1}|^2 = 1 - \hat{\mathbf{t}}_j \cdot \hat{\mathbf{t}}_{j+1}$.

What is the typical single-base bending angle (in degrees) if we take $a = 150$?

We write the (unnormalized) probability distribution of a given conformation of an $N + 1$ -tangent-vector-long chunk of molecule using the Boltzmann distri-

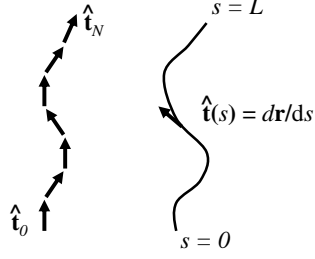


Fig. 4. Discrete-tangent and continuous-tangent models for DNA bending (see text).

bution:

$$P(\hat{\mathbf{t}}_0, \dots, \hat{\mathbf{t}}_N) = \prod_{j=0}^{N-1} e^{a \hat{\mathbf{t}}_j \cdot \hat{\mathbf{t}}_{j+1}} \quad (2.1)$$

Now we compute the thermal correlation of the ends of this segment of polymer:

$$\langle \hat{\mathbf{t}}_0 \cdot \hat{\mathbf{t}}_N \rangle = \frac{\int d^2 t_0 \cdots d^2 t_N \hat{\mathbf{t}}_0 \cdot \hat{\mathbf{t}}_N P(\hat{\mathbf{t}}_0, \dots, \hat{\mathbf{t}}_N)}{\int d^2 t_0 \cdots d^2 t_N P(\hat{\mathbf{t}}_0, \dots, \hat{\mathbf{t}}_N)} \quad (2.2)$$

This calculation is not too hard to do using the formula (recall decomposition of plane waves into spherical waves):

$$e^{a \hat{\mathbf{t}} \cdot \hat{\mathbf{t}}'} = \sum_{l=0}^{\infty} 4\pi i^l j_l(ia) \sum_{m=-l}^l Y_{lm}(\hat{\mathbf{t}}) Y_{lm}^*(\hat{\mathbf{t}}') \quad (2.3)$$

and if you write the dot product $\hat{\mathbf{t}}_0 \cdot \hat{\mathbf{t}}_N$ as an $l = 1$ spherical harmonic, and place $\hat{\mathbf{t}}_N$ along the $\hat{\mathbf{z}}$ axis.

The orthogonality of the spherical harmonics leads to a ‘collapse’ of the many sums over l ’s and m ’s into one sum. In the numerator only the $l = 1$ term (from the dot product) survives; in the denominator only the $l = 0$ term contributes. The result is:

$$\langle \hat{\mathbf{t}}_0 \cdot \hat{\mathbf{t}}_N \rangle = \left(\frac{i j_1(ia)}{j_0(ia)} \right)^N = e^{N \ln[\coth(a) - 1/a]} \quad (2.4)$$

The function $\coth(a) - 1/a$ is less than 1 for positive a . Therefore the correlation of direction falls off simply exponentially with contour distance N along our

polymer. Small local fluctuations of bending of adjacent tangents build up to big bends over the ‘correlation length’ of $-1/\ln[\coth(a) - 1/a]$ segments.

Problem 9: For the $a \gg 1$ limit, how many segments long is the tangent-vector correlation length?

Problem 10: Explain the relation between the discrete-tangent model discussed above and the one-dimensional Heisenberg (continuous-spin) model of classical statistical mechanics. Suppose a magnetic field is added: what would that correspond to in the polymer interpretation?

2.2.2. Bending elasticity and the persistence length

We can connect this discrete model to the continuous model for bending of a thin rod, from the theory of elasticity. We note that the bending energy of two adjacent tangents was, in $k_B T$ units, $-a \hat{\mathbf{t}}_j \cdot \hat{\mathbf{t}}_{j+1}$, which up to a constant is $\frac{a}{2} |\hat{\mathbf{t}} - \hat{\mathbf{t}}'|^2$.

The bending of a thin rod can be described in terms of tangent vectors $\hat{\mathbf{t}}$ distributed continuously along the rod contour. A bent rod has energy which is locally proportional to the square of its bending curvature $d\hat{\mathbf{t}}/ds$ (s is contour length):

$$E = \frac{B}{2} \int_0^L ds \left| \frac{d\hat{\mathbf{t}}}{ds} \right|^2 \quad (2.5)$$

where B is the rod bending modulus. For a rod of circular cross section of radius r made of an isotropic elastic material, $B = \frac{\pi}{4} Y r^4$ where Y is the Young modulus [4].

Problem 11: By considering a simple circular arc, find the contour length along a thin rod for which a one-radian bend has energy cost $k_B T$.

Problem 12: Pretend that dsDNA is made of a plastic material of Young modulus 3×10^8 Pa. Predict the bending constant B .

We can now connect our discrete and continuous models of bending, if we introduce the length b of the segments in our discrete model:

$$\begin{aligned} E &= -k_B T a \sum_{j=1}^N \hat{\mathbf{t}}_j \cdot \hat{\mathbf{t}}_{j+1} = \frac{k_B T a b}{2} \sum_{j=1}^N b \left| \frac{\hat{\mathbf{t}}_j - \hat{\mathbf{t}}_{j+1}}{b} \right|^2 \\ &\rightarrow \frac{B}{2} \int_0^L ds \left| \frac{d\hat{\mathbf{t}}}{ds} \right|^2 \end{aligned} \quad (2.6)$$

where a constant energy shift has been dropped. The final term represents the limit where we make b small, while making a big, keeping the product ab con-

stant. This continuum limit turns the finite difference into a derivative, and the sum into an integral.

The bending elastic constants a and B are related by $k_B T a b = B$, and the rod length corresponds to the number of tangents through $Nb = L$. So, for a rod with bending modulus B , if we wish to use a discrete tangent vector model with segment length b , we need to choose $a = B/(k_B T b)$.

If we now go back to the correlation function (2.4), we can write it in the continuum limit where a becomes large, replacing $\ln[\coth a - 1/a] \rightarrow -1/a$ and obtaining

$$\langle \hat{\mathbf{t}}(s) \cdot \hat{\mathbf{t}}(s') \rangle = e^{-k_B T |s-s'|/B} = e^{-|s-s'|/A} \quad (2.7)$$

The final term introduces the continuum version of the correlation length of (2.4) $A = B/(k_B T)$, called the *persistence length*. For the double helix, a variety of experiments show that $A = 50$ nm (150 bp) in physiological aqueous solution [5] (this term ‘physiological solution’ usually means water containing between 0.01 and 1 M univalent salt, and with pH between 7 and 8, at temperature between 15 and 30 C).

Problem 13: Starting with the persistence length $A = 50$ nm, estimate the bending modulus B , and the ‘effective Young modulus’ Y of the DNA double helix.

Both a and B represent effective elastic constants, and the bending energies being discussed here are really free energies (as in the theory of elasticity, we consider deformations at fixed temperature [4]). The ‘real’, microscopic internal energy must include thermal energy and chemical binding energies of the atoms, but as in many other areas of condensed matter physics we’ll choose to ignore atomic details and use coarse-grained models, since I will focus on phenomena at length scales of nm and larger (double helix deformations, DNA-protein interactions). This is not to say that atomic detail is not important: many important questions about the stability of the double helix and DNA-protein interactions require information at atomic scales, and can be theoretically approached only via numerical simulation of all the atoms involved [6] (Fig. 3).

2.2.3. End-to-end distance

The tangent vector $\hat{\mathbf{t}}(s)$ can be used to compute the distance between two points on our polymer, using the relation $\mathbf{r}(L) - \mathbf{r}(0) = \int_0^L ds \hat{\mathbf{t}}(s)$. This relation can be used to compute the mean-square distance between contour points a distance

L apart:

$$\langle |\mathbf{r}(L) - \mathbf{r}(0)|^2 \rangle = 2AL + 2A^2 (e^{-L/A} - 1) \quad (2.8)$$

In the limit where we look at points closer together than a persistence length, $L/A \ll 1$, we have a mean-square distance $= L^2 + \mathcal{O}(L/A)$; in this limit, the polymer doesn't bend very much, so its average end-to-end distance is just L .

In the opposite limit of a polymer many persistence lengths long, $L/A \gg 1$, we have a mean-square-distance of $2AL$, just the size expected for a random-walk of $L/(2A)$ steps each of length $2A$. We sometimes talk about the *statistical segment length* or *Kuhn segment length* in polymer physics: for the semiflexible polymer this segment length is $2A$. For the double helix, $2A$ is about 100 nm or 300 bp [5].

2.2.4. DNA loop bending energies

We'll hear in Section 5 about proteins which stabilize formation of DNA loops. Often, looping of DNA occurs so that sequences roughly 10 to 1000 bp away from the start of a gene can regulate (repress or enhance) that gene's transcription [7]. Formation of such a loop requires DNA bending, and now we can estimate the associated free energy.

Suppose we form a loop of length L . The simplest model is a circle of circumference L , with radius $L/(2\pi)$ and bending curvature $2\pi/L$, and bending energy

$$\frac{E_{\text{circle}}}{k_B T} = \frac{A}{2} L \left(\frac{2\pi}{L} \right)^2 = 2\pi^2 \frac{A}{L} \quad (2.9)$$

For $L = 300$ bp and $A = 150$ bp, this is a big energy - close to $10 k_B T$. A 100 bp circle would have a bending free energy three times larger than this!

You might be interested in the *lowest* energy necessary to bring two points a contour length L along a rod together. The optimal shape of the rod is of course not circular, but is instead tear-drop-shaped. The exact energy can be computed in terms of elliptic functions to be [8]

$$\frac{E_{\text{teardrop}}}{k_B T} = 14.055 \frac{A}{L} \quad (2.10)$$

about 71% of the energy of the circle. In either teardrop or circle case, the energy of making a loop diverges as $1/L$ for small L .

Problem 14: Carry out an approximate calculation of the tear-drop shape and energy, by using a circular arc combined with two straight segments. Use energy

minimization (with fixed total length) to find the angle at the base of the tear-drop (you should only have one parameter to minimize over) and the tear-drop configuration energy.

Problem 15: In a protein-DNA structure called the nucleosome, 146 bp of DNA make 1.75 helical turns with helix radius of 5 nm, and helical pitch (spacing of turns along the helix axis) of 3 nm. Using the simple models of this section, estimate the bending free energy of the DNA in $k_B T$.

2.2.5. Site-juxtaposition probabilities

These bending energies are not by themselves enough to accurately predict the probability that a DNA segment of length L forms a loop; we must also sum over bending fluctuations, thermally excited changes in shape. For the simple bending model described above, sophisticated calculations have been done for the probability of forming a loop.

Calculations of Stockmayer, Shimada and Yamakawa [8, 9] tell us the probability density for finding the two ends of a semiflexible polymer brought smoothly together (with the same orientation):

$$J_{\text{circle}} = \frac{\pi^2}{(2A)^3} \left(\frac{2A}{L} \right)^6 e^{-E_{\text{circle}}/k_B T + 0.257L/A} \quad (2.11)$$

If the condition that the ends come together smoothly is relaxed, the same authors found

$$J_{\text{teardrop}} = \frac{28.01}{(2A)^3} \left(\frac{2A}{L} \right)^5 e^{-E_{\text{teardrop}}/k_B T + 0.246L/A} \quad (2.12)$$

The units of these expressions are density (inverse volume), i.e., concentration of one end, at the position of the other.

For DNA, since the double helix can bend only over roughly 100 nm, the natural scale for J is very roughly $J \approx (100 \text{ nm})^{-3} = 10^{-6} \text{ nm}^{-3} \approx 10^{-6} \text{ Mol/litre}$ (1 Mol/litre, or M, is 0.6 nm^{-3}).

The empirical results provide an accurate interpolation between the two limits where bending energy ($L/A < 1$), and entropy ($L/A > 1$) dominate, including the experimentally- and numerically-established result that the *peak* probability of juxtaposition occurs for molecules about $L = 170 \text{ nm}$ (500 bp) long [5, 11].

For $L \gg A$ we reach the long-distance limit, where we may estimate the probability of finding the two ends of a long DNA close together, using the average end-to-end distance (2.8), which is $\approx \sqrt{2AL}$ in this limit. For $L \gg A$, the two ends are somewhere in a volume $\approx (AL)^{3/2}$. Therefore, the probability of finding the two ends together for $L/A \gg 1$ decays as $J \approx 1/(AL)^{3/2}$.

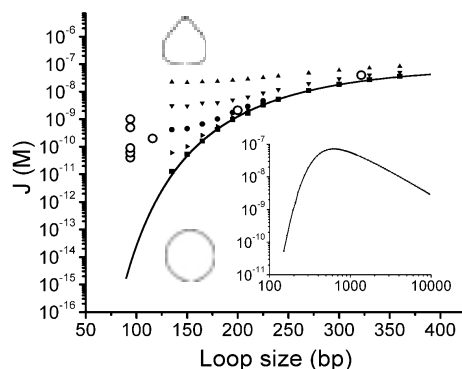


Fig. 5. Juxtaposition probability (J_{circle} in Mol/litre) for double helix. Solid lines show theoretical result for simple semiflexible polymer model of DNA double helix. Inset shows theoretical J_{circle} for large distances, showing peak near 500 bp and $L^{-3/2}$ decay. Main figure focuses on energetically-dominated small- L behavior, showing strong suppression of probability in the simple semiflexible polymer model (solid line). Open circles show recent experimental data of Cloutier and Widom for short DNAs [12]; there is an anomalously large probability of juxtaposition for 94 bp. Filled symbols correspond to a theory of DNA site juxtaposition including the effect of thermally excited ‘hinges’ [13]; we’ll hear more about this in Sec. 3.1.3.

This formula does not account for self-avoidance, but because the double helix has a segment length $2A$ so much longer than its diameter (only 3 nm even when electrostatic repulsion in physiological solution is taken into account) self-avoidance effects can be neglected for molecules as large as 10^4 bp in length.

2.2.6. Permanent sequence-driven bends

We’ve focused on thermally excited bends, using a model which has as its ‘ground state’ a perfectly straight conformation. The average shape of any DNA molecule depends on its sequence: different sequences have slightly different average distortions. A remarkable discovery is that it is possible, by ‘phasing’ sequences that generate kinks, one can obtain DNAs with strong permanent bends along them [14]. Some of these strong permanent bends are implicated in biological processes, for example facilitation of the binding of proteins that bend or wrap DNA.

2.3. Stretching out the double helix

One type of single-molecule experiment which has become widely studied is the stretching of DNAs using precisely calibrated forces. Early experiments showed that the double helix displayed polymer stretching elasticity of exactly what was expected from the semiflexible polymer model introduced above. This has been important to the design of many experiments focusing on the effects of proteins or other molecules binding to, or moving along DNA. This subsection reviews the basic polymer stretching elasticity of a long ($L \gg A$) double helix DNA.

A force f applied to a single DNA molecule of length L appears in the Boltzmann factor coupled to the end-to-end vector along the force direction (which we take to be z). Our energy becomes:

$$E = \frac{k_B T A}{2} \int_0^L ds \left| \frac{d\hat{\mathbf{t}}}{ds} \right|^2 - f \hat{\mathbf{z}} \cdot [\mathbf{r}(L) - \mathbf{r}(0)] \quad (2.13)$$

We can turn the end-to-end vector into an integral over $\hat{\mathbf{t}}$ as before, giving

$$\beta E = \int_0^L ds \left[\frac{A}{2} \left| \frac{d\hat{\mathbf{t}}}{ds} \right|^2 - \beta f \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} \right] \quad (2.14)$$

A single parameter $\beta A f$ controls this energy (to see this, write Eq. 2.14 using contour length in units of A). We therefore have two regimes to worry about: forces below, and above the characteristic force $k_B T/A$.

For the double helix, $A = 50$ nm, so $k_B T/A = 0.02 k_B T/\text{nm} = 0.08$ pN. This is a low force due to the long persistence length of the double helix.

Ideally, we want to calculate the partition function

$$Z(\beta A f) = \int \mathcal{D}\hat{\mathbf{t}} e^{-\beta E} \quad (2.15)$$

and then calculate the end-to-end extension, using

$$\langle \hat{\mathbf{z}} \cdot [\mathbf{r}(L) - \mathbf{r}(0)] \rangle = \frac{\partial \ln Z}{\partial \beta f} \quad (2.16)$$

This can be done in general numerically, but we can find the low- and high-force limits analytically.

2.3.1. Small forces ($< k_B T/A = 0.08 \text{ pN}$)

For small forces, we can calculate the end-to-end extension using linear response, since we know the zero-force fluctuation of the mean-square end-to-end distance: recall that this was $2AL$. This counted three components; by symmetry we have

$$\left\langle (\hat{\mathbf{z}} \cdot [\mathbf{r}(L) - \mathbf{r}(0)])^2 \right\rangle = \frac{2AL}{3} \quad (2.17)$$

The linear force constant will be $k_B T$ divided by this fluctuation, giving a small-extension force law:

$$f = \frac{3k_B T}{2AL} z + \dots \quad (2.18)$$

where we use the shorthand $z = \langle \hat{\mathbf{z}} \cdot [\mathbf{r}(L) - \mathbf{r}(0)] \rangle$ to indicate the average end-to-end extension in the force direction.

This is just the usual ideal (Gaussian) low-extension force law familiar from polymer physics. The spring constant of the polymer is inversely proportional to the persistence length, and to the total chain length.

2.3.2. Larger forces ($> k_B T/A = 0.08 \text{ pN}$)

The linear force law shows that our ideal DNA will start to stretch out when forces of $\approx k_B T/A$ are applied to it. We can also calculate the very nonlinear elasticity associated with the nearly fully stretched polymer, using an expansion in $1/\sqrt{f}$.

Suppose that the polymer is quite stretched out, so that $\hat{\mathbf{t}}(s) = \hat{\mathbf{z}} t_{\parallel} + \mathbf{u}$, where \mathbf{u} is in the xy plane, and has magnitude $\ll 1$. Since $\hat{\mathbf{t}}^2 = 1$, $t_{\parallel} = \sqrt{1 - |\mathbf{u}|^2} = 1 - \frac{1}{2}|\mathbf{u}|^2 + \dots$. Plugging this into the Hamiltonian (2.14) and expanding to leading order in $|\mathbf{u}|^2$ gives:

$$\beta E = -\beta f L + \frac{1}{2} \int_0^L ds \left[A \left| \frac{d\mathbf{u}}{ds} \right|^2 + \beta f |\mathbf{u}|^2 \right] \quad (2.19)$$

In this limit, the fluctuations can be seen to slightly reduce the length, generating the final energy cost term.

Introducing Fourier modes $\mathbf{u}_q = \int_0^L ds e^{iqs} \mathbf{u}(s)$ diagonalizes the Hamiltonian:

$$\beta E = -\beta f L + \frac{1}{2L} \sum_q (Aq^2 + \beta f) |\mathbf{u}_q|^2 \quad (2.20)$$

where $q = \pm 2\pi n/L$ for $n = 0, \pm 1, \pm 2, \dots$. The fluctuation amplitude of each mode is therefore

$$\langle |\mathbf{u}_q|^2 \rangle = \frac{2L}{Aq^2 + \beta f} \quad (2.21)$$

where the leading 2 comes from the two components (x and y) of \mathbf{u} . Now we can compute the real-space amplitude:

$$\langle |\mathbf{u}(s)|^2 \rangle = 2 \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{1}{(Aq^2 + \beta f)} = \frac{1}{\sqrt{4\beta A f}} \quad (2.22)$$

and finally the extension in the force direction

$$z = L \langle t_{\parallel} \rangle = L \left(1 - \frac{1}{2} \langle |\mathbf{u}|^2 \rangle + \dots \right) = L \left(1 - \frac{1}{\sqrt{4\beta A f}} + \dots \right) \quad (2.23)$$

The semiflexible polymer shows a distinct $1/\sqrt{f}$ behavior as it is stretched out. Also note that the energy expressed in wavenumbers shows that there is a force-dependent correlation length for the bending fluctuations, given by $\xi = \sqrt{k_B T A / f}$. Experiments on double-helix DNAs show this relation [15–20].

The asymptotic linear relation between z and $1/\sqrt{f}$ is quite useful. It turns out this holds well theoretically for the exact solution of the semiflexible polymer model under tension, for $z/L > 0.5$. If you have experimental data for stretching a semiflexible polymer, you can plot z versus $1/\sqrt{f}$ and fit a line to the $z/L > 0.5$, the z -intercept of the linear fit estimates the molecular length L , and the $1/\sqrt{f}$ intercept gives an estimate of $\sqrt{4\beta A}$, i.e., a measurement of persistence length. The agreement between different kinds of single-DNA experiments gives strong evidence that for long molecules, most of the elastic response comes from thermal bending fluctuations.

2.3.3. Free energy of the semiflexible polymer

It is useful to compute the free energy difference between unstretched and stretched polymer from the extension in the force direction, by integrating (2.16):

$$\ln Z(f) = \beta \int_0^f df' z(f') + \ln Z(0) \quad (2.24)$$

We'll drop the constant $\ln Z(0)$, which amounts to taking the relaxed random coil as a 'reference state' with free energy defined to be zero. For the semiflexible polymer the free energy takes the form

$$\ln Z(f) = \frac{L}{A} \gamma(\beta A f) \quad (2.25)$$

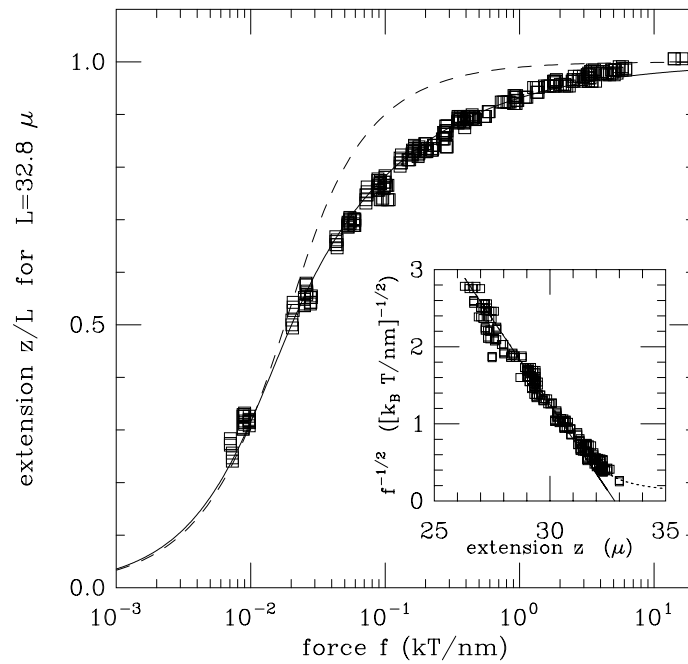


Fig. 6. Experimental data and models for stretching of the double helix, from Ref. [20]. Main figure shows experimental data (squares) of Smith et al [15] and a fit to the semiflexible-polymer model (solid line), for a persistence length $A = 53$ nm. The units of force are $k_B T/\text{nm}$; recall $1k_B T/\text{nm} = 4.1$ pN, for $T = 300$ K. Inset shows a plot of extension versus inverse square root of force, showing the linear relation between these two quantities. Dashed line shows result for freely-jointed polymer model with segment length 100 nm; this model describes the low-force polymer elasticity, but fails to describe the high-force regime of the experiment.

which is the scaling form of the partition function for the semiflexible polymer in the limit $L/A \gg 1$.

The dimensionless function $\gamma(x)$ can be computed numerically [20] for the semiflexible polymer (or for many variations of it [21]), but for us it will be sufficient to consider the limits:

$$\gamma(x) = \begin{cases} x - \sqrt{x} + \dots & x > 1 \\ 3x^2/4 + \dots & x < 1 \end{cases} \quad (2.26)$$

The free energy we are computing here, the log of the partition function at fixed force, can be converted to the work done extending the polymer to a given extension, $W(z)$, by the Legendre transformation $W(z) = -k_B T \ln Z + fz$.

2.3.4. Really large forces (> 10 pN)

For forces in the range 10 to 40 pN, the double helix starts to stretch elastically. This stretching can be described by adding an term to the result above [18, 20]:

$$\frac{z}{L} = 1 - \frac{1}{\sqrt{4\beta A f}} + \frac{f}{f_0} \quad (2.27)$$

The constant f_0 has dimensions of a force, and represents the stretching elastic constant of the double helix. In terms of the Young modulus, this elastic constant for a rod of circular cross-section of radius r is $f_0 = \pi r^2 Y$. Experimental data indicate $f_0 \approx 1000$ pN [22, 23].

Finally, at about 60 to 65 pN, depending a bit on salt concentration, there is an abrupt transition to a new double helix state about 1.7 times longer than B-form. This is sometimes called the S-form of DNA; there is at present some controversy over whether this form is base-paired or not [24].

Fig. 7 shows some experimental data for the high-force response of dsDNA (squares and diamonds) from two groups. Note the elongation of the double helix above the fully extended double helix value of 0.34 nm/bp, and the sharp ‘overstretching’ transition force ‘plateau’ near 63 pN.

Problem 16: Above we saw that $B = (\pi/4)Yr^4$ where r is the cross-sectional radius of an elastic rod. Compare the Y values inferred from B and f_0 . Are they consistent?

Problem 17: Consider longitudinal stretching fluctuations of adjacent base pairs. Compute the energy of a fluctuation of amplitude (length) δ : what is the root-mean-square value of the single-base-pair longitudinal fluctuation $\sqrt{\langle \delta^2 \rangle}$?

Problem 18: Under some conditions, a *single strand* of DNA will behave like a flexible polymer of persistence length $A_{ss} \approx 1$ nm. Find the characteristic force at which you might expect a single-stranded DNA to become 50% extended.

Problem 19: Consider the Hamiltonian (2.14) generalized so that it contains a *vector* force \mathbf{f} coupled via dot product to end-to-end extension.

Show $\partial_{\beta f_i} \ln Z = \langle x_i(L) - x_i(0) \rangle$ and $\partial_{\beta f_i}^2 \ln Z = \langle [x_i(L) - x_i(0)]^2 \rangle$ where the indices i label the three spatial coordinates.

Now, assuming the force to be in the z direction, verify the following formula relating the average extension and the ‘transverse’ end-to-end vector fluctuations:

$$\frac{\langle z(L) - z(0) \rangle}{\langle [x(L) - x(0)]^2 \rangle} = \frac{f}{k_B T} \quad (2.28)$$

Hint: use the fact that the partition function is a function of only the *magnitude* of \mathbf{f} .

This exact, nonperturbative relation is used in magnetic tweezer experiments to infer forces applied to single DNA molecules [25]. This does not depend on the details of the polymer part of the Hamiltonian - even if it contains long-ranged interactions - as long as it is invariant under space rotation.

Problem 20: For the semiflexible polymer, consider the approximate force-extension relation $\beta A f = z/L + 1/[4(1 - z/L)^2] - 1/4$. Show that this function reproduces the high- and low-force limiting behaviors derived above (it is not a terribly accurate representation for the exact behavior of $f(z)$). Compute the free energy $W(z)$ using this relation. Hint: integrate (2.16).

Problem 21: Consider the ‘freely jointed chain’ obtained by setting $a = 0$ in the segment model. Calculate the extension, and free energy ($\ln Z$) as a function of force. Also calculate the transverse mean-squared fluctuations as a function of force, and verify Eq. 2.28 for this model.

3. Strand Separation

In the previous section we didn’t say much about a feature of the double helix of paramount biological and biophysical importance: it consists of two covalently bonded *single-stranded* DNAs (ssDNAs) which are relatively weakly stuck to one another. The weakness of the binding of the two strands makes it possible for the two strands of a double helix to be separated from one another, either permanently as occurs in vivo during DNA replication, or transiently as occurs during DNA transcription (reading of DNA by RNAPol) and DNA repair.

Conversion of dsDNA to ssDNA can be accomplished in a few ways:

Elevated temperature: The double helix is stable in ‘physiological’ buffer (pH near 7, univalent salt in the 10 mM to 1 M range) for temperatures below about

50 C. Over the range 50 to 80 C, the double helix ‘melts’, with AT-rich sequences falling apart at the low end of this range, and highly GC-rich sequences holding together until the high end of this temperature range.

Denaturing solution conditions: Too little salt (< 1 mM NaCl), which increases electrostatic repulsion of the negatively charged strands, or pH too far from 7, destabilizes the double helix, lowering its melting temperature.

Sufficient ‘unzipping force’ applied to the two strands: If you pull the two strands apart, they will separate at forces in the 10 to 20 pN range, with force variations reflecting the sequence composition.

Below I will discuss the last of these three modes of strand separation, unzipping by force. A process similar to idealized ‘forced-unzipping’ is carried out in the cell to generate single-stranded DNA for DNA repair and replication; specialized motor enzymes called *DNA helicases* track along the double helix, pushing the two strands apart. The function of helicases can be precisely studied using single-DNA methods [26–28], and models for their activity require us to understand unzipping by force. To do that, we’ll need to learn about the strength of the base-pairing interactions, and the polymer elasticity of ssDNA.

3.1. Free-energy models of strand separation

In the simplest picture of DNA melting, we ignore base sequence entirely, and consider simply the average free energy difference g per base pair between isolated, relaxed ssDNAs and dsDNA, at room temperature and in physiological solution conditions. Then, for an N -base-pair-long molecule, the free energy difference between ssDNAs and dsDNAs would be just $G_{\text{ssDNAs}} - G_{\text{dsDNA}} = Ng$. For random DNA sequences, this $g \approx 2.5k_B T$; its positive value reflects the fact that the double helix is more stable than isolated single strands: very roughly, the probability of observing melted single strands is $e^{-\beta Ng}$.

Thermal melting can be most simply thought about by considering the temperature dependence of the base-pairing free energy, breaking it into ‘enthalpy’ h and ‘entropy’ s per base pair, i.e., $g = h - sT$. At the melting temperature $T_m = h/s$, the free energy of isolated ssDNAs is equal to free energy of double helix, making these two states equally probable.

Problem 22: Random-sequence DNA has $g \approx 2.5k_B T$ at 25 C, and melts near 70 C ($T = 343$ K). Estimate h and sT at 25 C ($T = 298$ K).

Problem 23: For the simple model where the strand separation free energy per base is a constant $g = h - sT$, calculate the probability of finding separated single strands as a function of temperature (you may consider this to be a two-state system). How does the *width* of the melting transition as a function of temperature scale with N ? You may want to plug in some numbers from the

previous problem.

3.1.1. Sequence-dependent models

A number of groups are working on accurate algorithms to predict the melting temperatures of dsDNAs as a function of sequence. One of these classes of models assign a contribution to base-pairing free energy for each *pair* of bases, the idea being that stacking interactions of adjacent base pairs play an important role in determining the stability of the double helix. The raw data behind such models are melting temperature data for a set of different-sequence, short (10 to 20 bp) dsDNAs.

Table 3.1.1 lists a set of free energies due to Santalucia [10] for the ten different oriented pairs of bases that occur along a DNA strand. All remaining pairs of bases can be obtained from considering the complementary sequence on the adjacent strand, e.g., the contribution of 5'-GA is the same as that of 5'-TC found in the table. The free energy of strand separation for a long N -bp molecule is obtained by adding the $N - 1$ adjacent-base contributions together. In addition, there are contributions for the ends which we won't discuss - all though they are significant when considering melting of short molecules.

The key point of Table 3.1.1 is that AT-rich sequences are lower in strand-separation free energy (the values for AT, AA and TA are all less than $1.7k_B T$), while GC-rich sequences are higher (GG, GC and CG are $3k_B T$ or more). Models of this type are not infallible - in reality, double-helix structure and energy depends on longer than nearest-neighbor sequence correlations - but they do give some idea of sequence dependence of base-pairing free energy.

The data of Table 3.1.1 are for the physiological ionic strength of 150 mM NaCl; lower ionic strengths reduce the base-pairing free energy. An ionic-strength correction for the base-pairing free energy has been given by Ref. [10]: $\Delta g_i = 0.2 \ln(M/0.150)$ where M is the molarity of NaCl.

Problem 24: Calculate the free energy differences between separated ssDNAs and double helices, for the following sequences: 5'-AATTAATTAATT, 5'-GCGCGCGCCGG, 5'-AGCTCCAAGGCT. You may want to consult reference [10] to include the end effects.

Problem 25: In Table 3.1.1 you can see that AT-rich sequences have roughly $2k_B T$ less free energy holding them together than do GC-rich sequences. For a random N -base sequence, there will therefore be a mean free energy of strand separation, and fluctuations of that free energy. Calculate the mean free energy per base pair, and estimate the fluctuations.

3.1.2. Free energy of internal 'bubbles'

The above discussion suggests that thermal melting might be described by a one-dimensional Ising model with sequence-dependent interactions, i.e., with some quenched 'randomness'. However, this would ignore an important physical effect that acts to suppress opening of bubbles in the interior of a long double helix. This effect is the *entropic cost* of forcing an internal 'bubble' to close [29]. This cost is not included in the strand separation free energy models described above which are fit to data obtained from melting of short double helices.

This loop free energy is easy to roughly understand - we have already discussed it above indirectly in our discussion of juxtaposition of DNA sequences. We mentioned that the long-molecule limit for DNA juxtaposition probability should be $J \approx N^{-3/2}$ simply from considering the fact that the two molecule ends should be found in a volume of radius $R \approx N^{1/2}$. If we think about this probability in terms of a free energy cost of constraining the ends to be near one another, we obtain the loop free energy cost

$$\Delta G_{\text{loop}} = \frac{3}{2} k_B T \ln N \quad (3.1)$$

Since ssDNA has a persistence length of roughly one base (0.7 nm), the N relevant here is simply the number of bases in the loop. For an internal ssDNA bubble formed by opening N base pairs, we should use $2N$ as the loop length.

This additional free energy discourages opening of internal bubbles, eliminating the use of the simple Ising model with short-ranged interactions to describe DNA melting. In fact, the logarithmic interaction of (3.1) is sufficiently long-ranged to kill the usual argument against a phase transition in a 1d system. A real phase transition occurs in the 'pure' DNA melting model including the logarithmic loop effect; however, variations in local melting temperatures due to sequence variations along long real DNAs wash out a sharp phase transition [29].

We can estimate the total free energy cost of an N -base-pair internal bubble, adding the base-pairing/stacking free energy to the loop free energy cost:

$$\sum_{i=1}^N g_i + \frac{3}{2} k_B T \ln(2N) \quad (3.2)$$

The sequence-dependent term ranges from about $N k_B T$ to $4N k_B T$, making the price of a large, 10 bp bubble from roughly 20 to 45 $k_B T$: i.e., very rare except for the most AT-rich sequences. Larger bubbles are even more costly, making them exceedingly rare excitations.

3.1.3. Small internal bubbles may facilitate sharp bending

Small internal bubbles are not impossibly costly excitations: a 3 bp bubble costs 8 to 15 $k_B T$. Short AT-rich 3 bp sequences are by this reckoning, open roughly 0.1% of the time, and can be expected every few hundred base pairs (e.g., the particularly weak sequence TATA appears once every 256 base pairs in random-sequence DNA).

These small, thermally excited bubbles suggest an explanation for the recent results of Cloutier and Widom [12] (see Fig. 5) showing that the cyclization (loop-formation) probabilities of dsDNAs less than 300 bp long are far larger than we would expect from the simple elastic bending model 2.5. The experimental data indicate that tight bends of the double helix can occur via an alternative, lower-free-energy mechanism. One possibility is that via separation of a few base pairs, a ‘flexible joint’ might appear that could reduce the bending energy of formation of a loop. Although the free energy cost of generating a few-base-pair ‘joint’ is roughly $10k_B T$, for short DNAs this becomes similar to the bending free energy saved by concentrating much of the bending into a localized, highly distorted defect in the double helix.

Problem 26: Consider Fig. 5, which shows experimental data indicating that circular closure of 94 bp DNAs occurs with probability far above the expected value J_{circle} . Suppose that for some free energy ϵ we can form a small bubble, and *kink* the DNA so that it can still close smoothly, but now if one bubble is excited, with the tear-drop shape which minimizes the bending energy. If *two* bubbles are excited, no bending is required. Estimate the probabilities of the zero-bubble, one-bubble and two-bubble closure states (use the Yamakawa-Stockmayer-Shimada loop formation probabilities 2.11 and 2.12; don’t forget that the kink can appear at any base pair position along the molecule). Estimate what ϵ should be to explain the 94 bp data.

Fig. 5 shows how the juxtaposition probability is affected by the inclusion of flexible joints with energy cost 9, 10, 11 and 12 $k_B T$, via a detailed calculation [13]. The experimental data are described well by joints which cost 11 $k_B T$, close to the value expected for localized strand separation of a few base pairs.

3.2. Stretching single-stranded nucleic acids

Single-stranded DNA has also been studied in single-molecule stretching experiments, and shows polymer elasticity distinct from that of double-stranded DNA: *ssDNA has twice the contour length per base of the double helix* since the helical backbones of the double helix contain about 0.7 nm per base, about half of the double helix contour length of 0.34 nm per base pair,

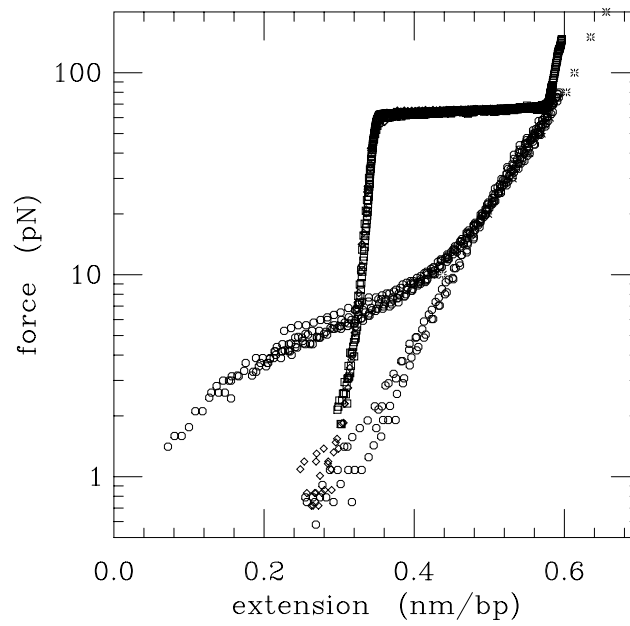


Fig. 7. Force versus extension of double helix and ssDNA. Squares show experimental dsDNA data of Léger et al [30, 31] for 500 mM NaCl buffer, diamonds show experimental dsDNA data of Smith et al [23] for 1 M NaCl buffer. Data for physiological salinity (150 mM NaCl) are similar, but have a plateau shifted a few pN below the 500 mM result, see Refs. [24, 31]. Circles show experimental data of Bustamante et al [32] for ssDNA; stars show high-force ssDNA data of Rief et al [33]. The left, lower-extension curve is for 150 mM NaCl, while the right, higher-extension curve is for 2.5 mM NaCl. The two ssDNA datasets converge at high force, to the behavior $x \approx \ln f$.

ssDNA has a persistence length of roughly a nanometer since the stiffness of the double helix is generated by the base pairing and stacking; once isolated, the ssDNA backbone is very flexible,

ssDNA can stick to itself by base-pairing and stacking interactions between bases along the same molecule.

These features are illustrated in Fig. 7 which plots experimental data for double helix and ssDNA side by side. The double helix, with a persistence length of 50 nm, is extended to its full contour length of about 0.34 nm/bp by forces of a few pN, and then shows a stiff force response, and finally the ≈ 60 pN force plateau. By comparison, ssDNA (open circles) only gradually stretches out, showing no stiff response near 0.34 nm/bp, and no force plateau. The force required to half-extend ssDNA is more than 3 pN; this reflects its short persistence length ≈ 1 nm (recall that the force needed to stretch out a polymer is roughly $k_B T/A$).

Fig. 7 also shows the strong dependence of ssDNA on salt concentration (open circles, left and right branches). At 150 mM NaCl ('physiological' salt, left set of data), ssDNA sticks to itself at low extensions, leading to an ≈ 1 pN force threshold to start opening the molecule. At low salt concentration (10 mM NaCl, right set of data) electrostatic self-repulsion eliminates this sticking effect, and the force threshold for initial extension.

For low salt concentration, the extension is well described by a logarithmic dependence on force, $\ln f/f_0$. This behavior can be understood in terms of a scale-dependent persistence length resulting from electrostatic effects [20,34,36]. At low forces, electrostatic self-repulsion effectively stiffens the polymer, helping to stretch it out; at higher forces, this effect is less pronounced (the monomers are farther away from one another) and the chain becomes harder to stretch. This effect is much more pronounced for ssDNA than for dsDNA since the backbone persistence length ≈ 1 nm is comparable to, or even less than, the screening length for electrostatic interactions (recall the Debye screening length is $\lambda_D = 0.3 \text{ nm} / \sqrt{M}$ for NaCl at M Mol/litre).

Problem 27: Force-extension data of Fig. 7 at low ionic strength are described by $x(f) \approx x_0 \ln(f/f_0)$ where x_0 and f_0 are constants. Compute the force-extension response in the high-force limit using Eq. 2.22, given the scale-dependent persistence length

$$A(q) = \begin{cases} A_0 q_0/q & q < q_0 \\ A_0 & q > q_0 \end{cases} \quad (3.3)$$

A more realistic model of scale-dependence of persistence length, based on Coulomb self-interactions, gives rise to similar behavior; see Refs. [20, 34–36]. A recent experiment by Visscher et al [37] on a poly-U RNA, eliminating base-

pairing, shows scale-dependent persistence length behavior rather clearly.

3.3. Unzipping the double helix

We now have all the pieces to analyze unzipping of the double helix by a force which pulls the two strands apart (Fig. 8). We will compare the free energy of two paired bases, g , to the free energy at constant force for two unpaired and extended bases. The free energy per base can be found from the experimental elasticity data via 2.24:

$$\gamma(f) = \int_0^f df' x(f') \quad (3.4)$$

where $x(f')$ is the length per base of the ssDNA data of Fig. 7. The function $\gamma(f)$ increases with f . The threshold for unzipping occurs when this two times this free energy – for the two bases – equals the base-pairing energy g :

$$2\gamma = g \quad (3.5)$$

Treating the ssDNA as a harmonic ‘spring’ we can write $\gamma(f) \approx (\ell f)^2 / (2k_B T)$ where $\ell \approx 0.4$ nm (this roughly matches the integral of the 150 mM force curve of Fig. 7 for forces below 20 pN). This gives an unzipping force:

$$f = \frac{\sqrt{k_B T g}}{\ell} \quad (3.6)$$

Plugging in g from 1 to $4 k_B T$, we see that the unzipping force varies from 10 to 20 pN, depending on sequence. Experiments of Bockelmann and Heslot on genomic molecules find fluctuations around 15 pN, the average of this range (see Fig. 9) [38]. The full range of unzipping forces from 10 to 20 pN has been observed by Rief et al [33, 39] in experiments on pure AT and GC DNAs.

Problem 28: For the harmonic model of ssDNA extensibility, calculate the force-extension relation. Compare the results for forces between 1 and 20 pN with the 150 mM NaCl ssDNA data in Fig. 7.

Problem 29: We can alternately describe unzipping using *extension* as a control parameter. Suppose one has a partially unzipped dsDNA, where n base pairs have been separated. The free energy is made up of elastic stretching energy, and base pairing energy:

$$F = \frac{k_B T (2x)^2}{2(2n)b^2} + ng \quad (3.7)$$

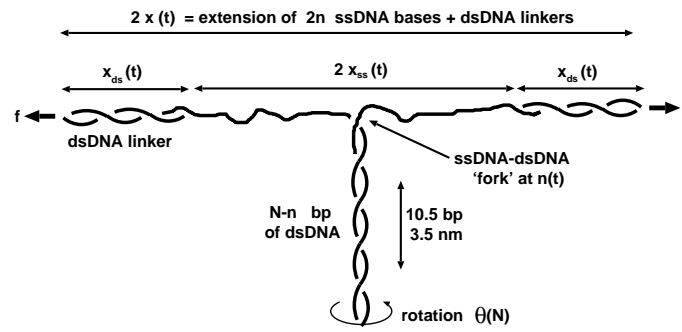


Fig. 8. Unzipping of DNA by force. Note that a torque can be applied to the end of the dsDNA region, coupled to the rotational angle θ .

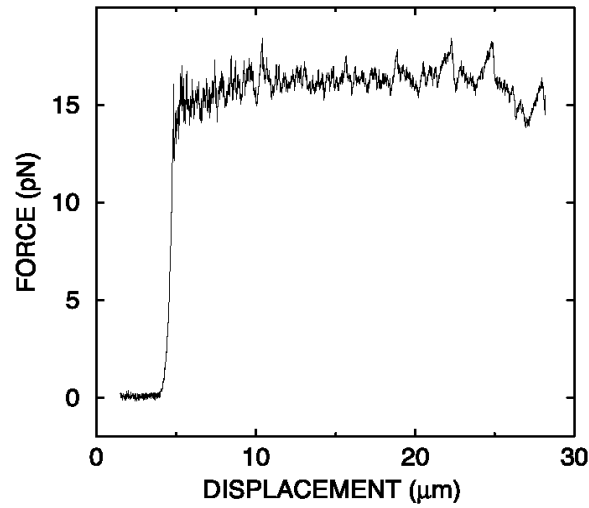


Fig. 9. Experimental data of Bockelmann et al [38] for unzipping of DNA at $0.02 \mu\text{m}/\text{sec}$. Sequence-dependent variations in force occur, around an average force of about 15 pN.

Note that opening n base pairs results in a $2n$ -base-long ssDNA (see Fig. 8). Note also that $n \geq 0$. Find the equilibrium number of base pairs unzipped, as a function of extension. For a partially unzipped molecule, also calculate the *fluctuation* in the number of bases that are unzipped. What are the corresponding *extension* fluctuations?

3.3.1. Effect of torque on dsDNA end

As unzipping proceeds, the dsDNA region must rotate to allow the two ssDNAs to be pulled out. If a torque is applied at the end of the dsDNA region, it can affect the unzipping force. This rotation is $\theta_0 = 2\pi/10.5 = 0.60$ radians per base pair unzipped. Adding the work $\tau\theta_0$ that must be done against the torque for each unzipped base pair, the equation for unzipping becomes

$$2\gamma = g - \tau\theta_0 \quad (3.8)$$

For the sign convention of Fig. 8, right-handed torque reduces the stability of the double helix, while left-handed torque acts to stabilize it. Using our harmonic approximation, we can obtain a torque-dependent unzipping force [40]:

$$f = \frac{\sqrt{k_B T (g - \tau\theta_0)}}{\ell} \quad (3.9)$$

As torque becomes more positive, the unzipping force threshold is decreased. When the torque becomes positive enough to unwind the DNA on its own, the unzipping force threshold becomes zero: this point is given by $\tau = g/\theta_0$, which ranges from $1.6k_B T$ for weakly bound (AT-rich) sequences, to $7k_B T$ for the most strongly bound (GC-rich) sequences.

If unzipping is done rapidly, the rotation of the dsDNA will generate a drag torque. In the simplest model for this where the DNA is supposed to spin around its axis, the drag torque is roughly

$$\tau = -4\pi\eta r^2 L_{ds} \frac{d\theta}{dt} \quad (3.10)$$

where L_{ds} is the length of the dsDNA region, $r \approx 1$ nm is the dsDNA cross-section hydrodynamic radius, and viscosity $\eta = 10^{-3}$ Pa·sec for water and most buffers. Effects of the drag associated with dsDNA rotation have been observed in experiments of Bockelmann and Heslot (see Fig. 10) [43]; the above model is in fair agreement with the experiment [41]. Note that P. Nelson has argued that there is an additional and large contribution to the rotational drag by permanent bends along the DNA contour [42]. The shape of the DNA gives rise

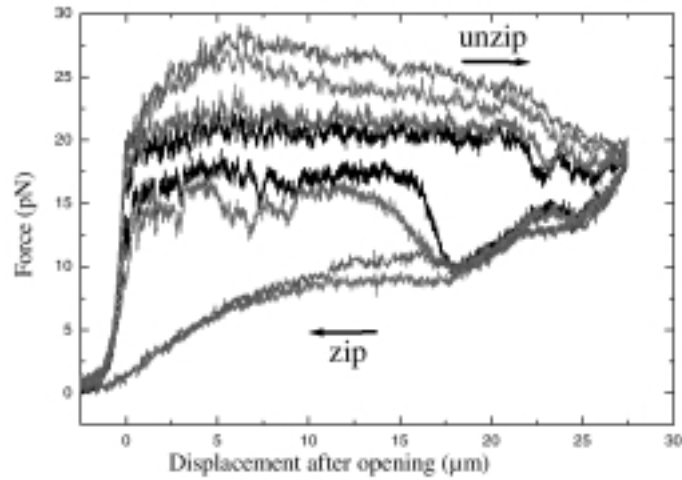


Fig. 10. Experimental results of Ref. [43] showing unzipping force rate-dependence. The two ssDNA ends are forced apart at velocities of 4, 8, 16 and 20 $\mu\text{m}/\text{sec}$. Force versus ssDNA extension (see Fig. 8) is plotted. During unzipping, higher velocities generate higher unzipping forces.

to an effective increase in its cross-section radius r and thus the rotational drag coefficient.

Problem 30: Estimate the number of base-pairs per second that should be unzipped in order that rotational drag can push the unzipping force up by 5 pN (assume a uniform molecule with $g = 2.5k_B T$).

3.3.2. Fixed extension versus fixed force for unzipping

In single-molecule stretching experiments, like any experiment on a small system, choosing whether force or extension are controlled can be critical to the results. For example, laser tweezers and atomic force microscopes essentially control the position of the end of a molecule; magnetic tweezer setups by contrast provide fixed force. Unzipping of DNA provides a very good example of how these two types of experiments give different kinds of data. Fixed-extension unzipping experiments push the ssDNA-dsDNA ‘fork’ along, and observe jagged force ‘stick-slip’ events. Each stick event corresponds to the momentary stalling of the fork at a GC-rich ‘barrier’: the force then increases to a level where a ‘slip’, or barrier-crossing event occurs (see Fig. 9).

Conversely, in a fixed-force experiment, one observes the increase of extension as a function of time. For unzipping, this typically takes the form of a series of extension *plateaus*. These plateaus again correspond to the stalling of the fork

at a GC-rich barrier region; however, now the force is constant, and one must wait for a thermal fluctuation for unzipping to proceed. If one is well below the maximum unzipping force for GC-rich sequences (see 3.6), the barriers can be immense: even a fraction of a $k_B T$ per base pair required to cross a long, slightly GC-rich region can give rise to an immense barrier. This effect has been theoretically emphasized by D. Lubensky and D. Nelson [44] and the constant-force extension plateaus have been observed in experiments by the group of Danilowicz *et al* [45].

Recent experiments by the same group have studied unzipping as a function of temperature [46]. The results are in surprisingly discord with predictions based on the temperature dependence in the ‘standard models’ of DNA strand separation free energy [10].

4. DNA Topology

The topological properties of DNA molecules are important biologically. The linking number of the two strands in the double helix is particularly important to DNA structure in bacterial cells, and controls ‘supercoiling’, or wrapping of the double helix around itself. The entanglement of the double helix with itself (knotting), and with other molecules (braiding) is also important since DNA molecules (chromosomes) must be separated from one another during cell division.

4.1. DNA supercoiling

The phenomenon of supercoiling is familiar from dealing with twisted strings or wires: twist strain in a string can be relaxed by allowing the string to wrap around itself. For DNA molecules, description of this behavior requires one more ingredient, thermal fluctuation of the molecule conformation.

The physical feature of the double helix that gives rise to supercoiling is the wrapping of the two strands around one another. Neglecting bending for the moment, the relaxed double helix has one link between strands for each 10.5 bp along the molecule. This ‘relaxed linking number’ can be expressed as $Lk_0 = N/10.5$ bp for an N -bp double helix. The relaxed helix repeat of 10.5 bp can be expressed as a length $h = 3.6$ nm, allowing us to also write $Lk_0 = L/h$.

4.1.1. Twist rigidity of the double helix

Still avoiding bending, if we twist the double helix so that one end is rotated by an angle Θ relative to the other, the number of links between the strands will be changed by an amount $\Theta/(2\pi)$. In this case where there is no bending, the

change in linking number of the double helix, ΔLk , equals the change in twist, ΔTw .

It costs some energy for this twist distortion: a simple harmonic model is

$$\frac{E}{k_B T} = \frac{C}{2L} \Theta^2 = \frac{2\pi^2 C}{L} (\Delta Tw)^2 \quad (4.1)$$

This ‘twist’ energy is controlled by an elastic constant C with dimensions of length. This *twist persistence length* is about 100 nm for double helix DNA based on recent single-molecule experiments [47]; note that this is appreciably larger than the estimate of ≈ 75 nm that is the result of a number of solution-phase experiments. We’ll see a possible explanation for this disagreement later when we discuss twist rigidity of DNA.

Problem 31: Consider the harmonic twist energy. Calculate the thermal expectation value of Θ^2 : your result will depend on the molecule length L . Why is C called the twist persistence length?

Problem 32: Assuming the double helix to be composed of a uniform isotropic elastic medium, use A and C to determine the two Lamé coefficients, and equivalently the Young modulus and the Poisson ratio (you will likely want to review Landau and Lifshitz’ *Theory of Elasticity* [4] unless you are really an expert in elasticity theory; also recall that we have already figured out the Young modulus from both the bending persistence length *and*, independently, from the stretching force constant).

Problem 33: What torque is necessary to twist a DNA of length L by angle Θ ? For left-handed twisting, for what angle Θ will the twisting build up enough torque to start unwinding AT-rich sequences (see Sec. 3.3.1)?

4.1.2. Writhing of the double helix

When we allow bending of the double helix to occur, the linking number is no longer equal to the twisting number. However, as long as the bending radius is large compared to the radius of the double helix, there is a simple relation between twisting and bending contributions to the total linking number of the double helix:

$$\Delta Lk = \Delta Tw + Wr \quad (4.2)$$

The quantity Wr , or ‘writhe’, is dependent only on the bending of the double helix backbone. Very roughly, Wr measures the signed number of crossings of the molecule axis over itself, when the molecule shape is projected onto a plane.

Formally, linking number of the two strands can only be defined if the double helix is *circular*, i.e., if both of the strands are closed circles. I will be slightly

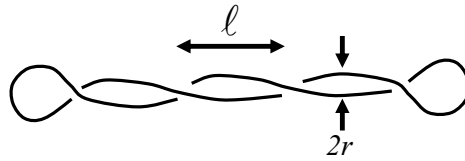


Fig. 11. Plectonemic supercoiled form of circular DNA, showing length between crossings ℓ , and cross-sectional radius r . Note that the only appreciable DNA bending occurs at the ends; also note that the line indicates double-helix DNA.

loose with this, and sometimes talk about linking number of *open* molecules. If you want to make linking number of a linear molecule precise, you can just imagine extending the strand ends straight off to infinity, and closing them there. This will not lead to large corrections in the situations we will be interested in.

We consider the situation where as bending occurs, the linking number remains fixed. This is most relevant to *circular* double helix molecules (with no breaks or ‘nicks’ in their backbones), the linking numbers of which are constant. Circular DNAs are found in bacteria: both the large 4.5 Mb chromosome and small plasmids (typically 2 to 15 kb in circumference) are normally found in closed circular form.

A second situation where ΔLk can be considered constant is when one is holding onto the two ends of a DNA molecule, and forcing them to be parallel and unable to rotate. This case can be studied experimentally in single-DNA micromanipulation experiments, most notably in elegant magnetic tweezer experiments [25].

By rearranging 4.2 to $\Delta Tw = \Delta Lk - Wr$ we can see the mechanism for buckling of a twisted wire: twisting without bending will change ΔTw away from zero, costing twist energy. However, now if the wire is allowed to buckle so that it wraps around itself, the Wr from the wrapping can cancel the ΔLk , and reduce the twisting energy. By braiding the molecule with itself, the bending energy can be small as well. This self-wrapping of DNA is called *plectonemic supercoiling*.

For the plectonemic structure shown above, the magnitude of the writhe is and equal to the number of crossings: $|Wr| \approx L/(2\ell)$. The sign of the writhe for the right-handed coiling shown in Fig. 11 is negative; for a left-handed plectonemic supercoil, the writhe would be positive. For achiral conformations, $Wr = 0$.

4.1.3. Simple model of plectonemic supercoiling

We can write down a simple model for the free energy of the plectoneme:

$$\frac{F}{k_B T} = \frac{2\pi^2 C}{L} \left(\Delta Lk \pm \frac{L}{2\ell} \right)^2 + \frac{AL}{2} \left(\frac{r}{\ell^2} \right)^2 + \frac{L}{(Ar^2)^{1/3}} \quad (4.3)$$

The first term is just 4.1 with 4.2 rearranged and plugged in, using the plectonemic writhe $Wr = \mp L/2\ell$; the top sign is for a right-handed plectoneme, the bottom is for left-handed. The second term is the bending energy 2.5, using r/ℓ^2 as the curvature.

The third term arises from confinement of the DNA inside the ‘tube’ of the supercoil, of radius r . We can think about this in terms of a correlation length λ for thermally excited bending fluctuations: the smaller this wavelength, the smaller the transverse fluctuations. For bending with transverse displacement r over wavelength λ , the curvature is r/λ^2 ; the energy of this bend is $k_B T A r^2 / \lambda^3$. Using the equipartition theorem this energy will be $k_B T$, giving us the relation $\lambda = A^{1/3} r^{2/3}$. Finally, the confinement free energy density will be $k_B T / \lambda$, giving the third term of 4.3.

The free energy model 4.3 needs to be minimized to determine the equilibrium values of r and ℓ . First, we can determine r :

$$r \approx \frac{\ell^{3/2}}{A^{1/2}} \quad (4.4)$$

Then we can plug this result in to 4.3; simplifying some numerical factors we have

$$\frac{F}{k_B T L} = 2\pi^2 C \left(\frac{|\Delta Lk|}{L} - \frac{1}{\ell} \right)^2 + \frac{1}{\ell} \quad (4.5)$$

The sign has been chosen so that the writhe has the same sign as ΔLk , which always reduces the free energy. Minimizing this with respect to $1/\ell$ gives the result:

$$\frac{1}{\ell} = \frac{|\Delta Lk|}{L} - \frac{1}{4\pi^2 C} \quad (4.6)$$

There is no solution for positive ℓ when linking number is too small: when $|\Delta Lk| < L/(4\pi^2 C)$, the confinement free energy is too expensive, so the DNA does not supercoil. Then, as $|\Delta Lk|$ is increased beyond this limit, $1/\ell$ becomes gradually smaller and the supercoil tightens up. This threshold indicates that until the added linking number exceeds one per twist persistence length, the DNA molecule will not supercoil.

Linking number is often expressed intensively using $\sigma \equiv |\Delta Lk|/Lk_0$ which just normalizes the change in linking to the relaxed linking number. In a more careful calculation where numerical factors and geometrical details are accounted for carefully, the threshold for supercoiling is at $\sigma \approx h/(2\pi C)$; plugging in $h = 3.6$ nm and $C = 100$ nm gives a threshold σ of roughly 0.01. Another feature of the more complete theory is that the transition is ‘first-order’: the minimizing ℓ jumps from $\ell = \infty$ to a finite value. Electron microscopy experiments [48] indicate that plectonemic supercoiling requires about this level of σ (see Fig. 12); calculations of structural parameters of plectonemes also are in accord with the results of EM studies. In eubacteria such as *E. coli*, the chromosome and small circular ‘plasmid’ DNAs have nonzero ΔLk , with a $\sigma \approx -0.05$. This underwinding is thought to play a role in gene regulation, since AT-rich promoter regions will be encouraged to open by the torsional stress associated with this amount of unlinking.

An important feature of plectonemically supercoiled DNA is its *branched* structure. Branch points can be thought of as defects in the plectonemic supercoil structure: like the ends, there is some energy cost associated with them. However, there is an entropy gain $\approx k_B \ln L/A$ of having a branch point, since it can be placed anywhere in the molecule. Balance of branch point energy and entropy determines the observed density of a Y-shaped branch point for every 2 kb along a supercoil with $\sigma = -0.05$. Branching is also very important to the internal ‘sliding’ of DNA sequence around in the interior of a plectonemically supercoiled DNA, is important to some enzymes which bind to two sequences simultaneously, often across a plectonemic superhelix [49].

Problem 34: Find the dependence of r on ΔLk for the model of plectonemic supercoiling discussed above. At what value of σ does r reach $r_0 = 2$ nm, roughly the point at which the double helix will run into itself?

This effect is important as when the double helix starts to run into itself, twist compensation can no longer occur.

The following three problems will work out well best using the slightly more detailed models for the writhe and for the bending energy of the plectoneme discussed in Refs. [50, 51].

Problem 35: For the plectonemic supercoil *including* the constraint $r > r_0$, find the free energy $F/k_B T$ (note the two regimes where r is free and where r is constrained to be r_0).

Problem 36: For the plectonemic supercoil model discussed above, find the dependence of $\Delta Tw/\Delta Lk$ on ΔLk and σ .

Problem 37: Calculate the *torque* in a DNA double helix of length L , as a function of ΔLk , for the plectonemic supercoil described above. For $\sigma < 0$, at

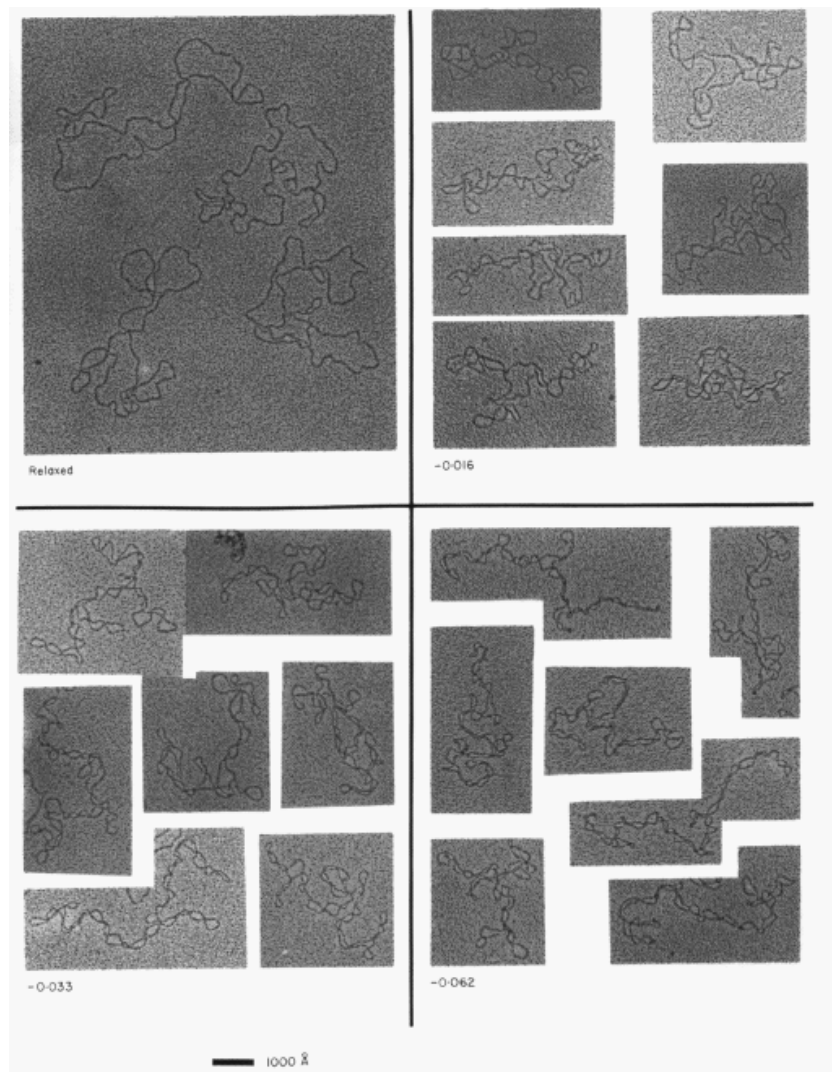


Fig. 12. Electron micrographs of supercoiled DNA at a few different σ values. Scale bar is 100 nm (300 bp); molecules are all 7 kb (2300 nm) in length. Reproduced from Ref. [48].

what value of σ does unwinding of AT-rich sequences in the double helix start to occur?

4.2. Twisted DNA under tension

It is possible to carry out single-DNA experiments as a function of force, and linking number [25]. The description of this situation along the framework of Sec. 2.3 is straightforward. At a fixed force, changing σ away from zero compacts a DNA molecule. If enough torsional stress is placed on a DNA molecule under tension, buckling will occur and plectonemic supercoils will appear along its length, leading to strong reduction in molecule extension [50, 51], as has been observed experimentally [25]. However, if σ is not so large that plectonemic coils appear, a milder compaction occurs which can be treated using a small-fluctuation approach, discussed by Moroz and Nelson [47], and by Bouchiat and Mezard [52]. This region of relatively mild compaction by writhing is a good regime in which to measure the double helix torsional modulus.

To carry out the treatment of this mild compaction analytically, we need the writhe of a nearly straight DNA, in terms of tangent vector fluctuations. As long as the tangent vector stays in the hemisphere around $\hat{\mathbf{z}}$, we have:

$$\text{Wr} = \int_0^L \frac{ds}{2\pi} \frac{\hat{\mathbf{z}} \cdot \hat{\mathbf{t}} \times \partial_s \hat{\mathbf{t}}}{1 + \hat{\mathbf{z}} \cdot \hat{\mathbf{t}}} \quad (4.7)$$

Now we can write the Hamiltonian for a DNA subjected to tension, plus held at fixed linking number, by just adding the twist energy 4.1 to the stretching Hamiltonian 2.14. The White formula 4.2 allows us to express the twist in terms of linking number and writhe:

$$\begin{aligned} \frac{E}{k_B T} &= \int_0^L ds \left[\frac{A}{2} \left(\frac{d\hat{\mathbf{t}}}{ds} \right)^2 - \frac{f}{k_B T} \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} \right] \\ &\quad + \frac{2\pi^2 C}{L} \left(\Delta \text{Lk} - \int_0^L \frac{ds}{2\pi} \frac{\hat{\mathbf{z}} \cdot \hat{\mathbf{t}} \times \partial_s \hat{\mathbf{t}}}{1 + \hat{\mathbf{z}} \cdot \hat{\mathbf{t}}} \right)^2 \end{aligned} \quad (4.8)$$

We'll do a harmonic calculation, expanding 4.8 to quadratic order in \mathbf{u} , the transverse (xy) components of the tangent vector:

$$\begin{aligned} \frac{E}{k_B T} &= \frac{2\pi^2 C}{L} (\Delta \text{Lk})^2 - \frac{Lf}{k_B T} \\ &\quad + \int_0^L ds \left[\frac{A}{2} \left(\frac{d\mathbf{u}}{ds} \right)^2 + \frac{f}{2k_B T} \mathbf{u}^2 - \frac{2\pi C}{h} \sigma \hat{\mathbf{z}} \cdot \mathbf{u} \times \partial_s \mathbf{u} \right] + \mathcal{O}(\mathbf{u}^3) \end{aligned} \quad (4.9)$$

Here the ΔLk in the cross term of the twist energy has been converted to the intensive linking number density σ . For $\sigma = 0$, we return to the high-extension limit of the stretched semiflexible polymer; for nonzero σ , the cross-product term will generate chiral fluctuations.

The fluctuation (\mathbf{u} -dependent) part of the quadratic Hamiltonian 4.9 can be rewritten in terms of Fourier components of \mathbf{u} :

$$\int \frac{dq}{2\pi} \left[\frac{1}{2} (Aq^2 + \beta f) |\mathbf{u}_q|^2 + \frac{2\pi C\sigma}{h} iq\hat{\mathbf{z}} \cdot \mathbf{u}_q^* \times \mathbf{u}_q \right] + \dots \quad (4.10)$$

Problem 38: Show that the quadratic part of the twisted-stretched DNA energy 4.10 has an instability (a zero eigenvalue) at the point $(2\pi C\sigma/h)^2 = 4Af/k_B T$.

This result is just the classical buckling instability of a beam of bending rigidity B subjected to compressive force f and torque τ , which occurs when $\tau^2 = 4Bf$ [53].

Problem 39: Show that the log of the partition function $\ln Z(f, \sigma)$ for the quadratic- \mathbf{u} fluctuations, including the non-fluctuation contributions, has the form, in an expansion in inverse powers of force:

$$\begin{aligned} \frac{\ln Z}{L} &= \frac{f}{k_B T} - \frac{2\pi^2 C\sigma^2}{h^2} - \left(\frac{f}{k_B T A} \right)^{1/2} \\ &+ \frac{1}{4} \left(\frac{2\pi C\sigma}{h} \right)^2 \left(\frac{k_B T}{4A^3 f} \right)^{1/2} + \dots \end{aligned} \quad (4.11)$$

You will need to find the normal modes of the fluctuations in order to compute this partition function. Note that the result can be written in a way similar to Eq. 2.26: $\frac{A}{L} \ln Z = \gamma(x, y)$ where $x = \beta A f$ and $y = 2\pi C\sigma/h$ are dimensionless variables characterizing the applied force and the linking number, respectively.

The extension as a fraction of the total molecular length follows from 4.11 and 2.16, as:

$$\langle \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} \rangle = 1 - \left(\frac{k_B T}{4A f} \right)^{1/2} - \frac{1}{2} \left(\frac{2\pi C\sigma}{h} \right)^2 \left(\frac{k_B T}{4A f} \right)^{3/2} + \dots \quad (4.12)$$

For this model, twisting the DNA ($\sigma \neq 0$) leads to a reduction in extension. This can be seen in the data of Fig. 13; however, note that the quadratic twist-dependence occurs only quite near to $\sigma = 0$, and is clearest in the 0.2 pN data of the figure.

The free energy 4.11 can be used to find the relation between the torque applied at the end of the chain, and the linking number. Since linking number is

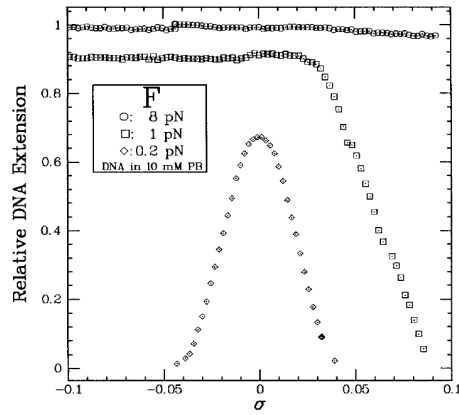


Fig. 13. Extension of DNA as a function of linking number σ , for a few fixed forces. Reproduced from Ref. [25].

controlled by rotating the end of the chain, the torque applied at the end of the chain is obtained from the linking number derivative of 4.11:

$$\frac{\tau}{k_B T} = -\frac{1}{2\pi} \frac{d}{d\Delta Lk} \ln Z = \left[1 - \frac{1}{2} \frac{C}{A} \left(\frac{k_B T}{4A f} \right)^{1/2} \right] \frac{2\pi C \sigma}{h} \quad (4.13)$$

Moroz and Nelson have emphasized this result: the effective twist modulus of the double helix goes down as tension in the chain goes down. This effect occurs because for lower forces, more writhing can occur, allowing the twist energy and therefore the torque to be reduced.

The linearized calculation of the Wr fluctuations in Eq. 4.9 cannot account for plectoneme-type crossings. Recently, Rossetto has argued that these fluctuations modify the $\mathcal{O}(\sigma^2)$ term in Eq. 4.12 [54]. This effect may be behind discrepancies between C determinations by different theoretical approaches, thanks to either over- or under-estimation of writhe fluctuations.

4.3. Forces and torques can drive large structural reorganizations of the double helix

The previous calculations have assumed that the forces and torques were not able to cause structural phase transitions in the double helix. We've already seen that

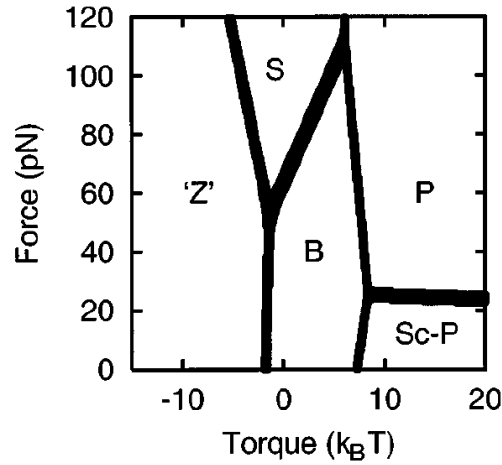


Fig. 14. ‘Phase diagram’ of double helix as function of external torque and force. Reproduced from Ref. [55].

at zero force an unwinding torque of about $-2k_B T$ is sufficient to start unwinding AT-rich regions of the double helix. Experiments show that there may be as many as five different structural states of the double helix which can be accessed by twisting and pulling on DNA [55], and those experiments allow one to predict a ‘force-torque phase diagram’ (Fig. 14). For forces in the < 50 pN range the double helix is stable roughly over the torque range $-2k_B T$ to $+7k_B T$ (the B-DNA region of Fig. [55]).

4.4. DNA knotting

Above we’ve discussed the effect of supercoiling, which is controlled by the ‘internal’ linking number of the two ssDNAs inside the double helix. A separate and important property of the double helix is the ‘external’ entanglement state of the double helix backbone, the more usual case of topology discussed in usual polymer physics. A single circular DNA can carry a knot along its length; alternately two or more circular DNAs can be linked together.

When an initially linear DNA is closed into circular form, there is some possibility that a knot is generated. You might be wondering why all the molecules of Fig. 12 are all *unknotted*. There are two reasons for this, the first one biological and the second one biophysical.

4.4.1. Cells contain active machinery for removal of knots and other entanglements of DNA

Cells contain enzyme machines called *topoisomerases* which catalyze changes in DNA topology. For example, entanglements (including knots) can be removed or added by *type-II topoisomerases* which are able to cut the double helix, pass another double helix segment through the resulting gap, and then seal the gap up. Type-II topoisomerases require ATP.

However, although existence of type-II topoisomerases tells us that it is *possible* for entanglements to be removed, we still are left wondering how they ‘know’ how to remove rather than add entanglements. Astonishingly, it has been experimentally demonstrated that type-II topoisomerases by themselves have the capacity to recognize and remove knots and other entanglements along circular DNA molecules [56].

4.4.2. Knotting a molecule is surprisingly unlikely

Let’s suppose we take an ensemble of linear DNA molecules of length L at low enough concentration that they do not interact with one another. Now, let’s add a small quantity of an enzyme which catalyzes closure of the molecules into circles, and the reverse process (this is possible). Then we can ask what the probability $P_{\text{unknot}}(L)$ is that the molecule is unknotted, as a function of L .

We can argue that $P_{\text{unknot}} \approx \exp[-L/(N_0A)]$, for some constant N_0 . For small L , there will be a large free energy cost of closing a molecule into a circle making $P_{\text{knot}} \rightarrow 1$. However, for larger molecules, the probability of an unknotted configuration should go down. The exponential decay reflects the fact that over some length (N_0 persistence lengths) the probability of having no knot drops to $1/e$: applying this probability to each L_0 along a DNA of length L gives us $P_{\text{unknot}}(L) \approx (1/e)^{L/(N_0A)}$. This rough argument can be made mathematically rigorous [57].

It turns out that even for an ‘ideal’ polymer which has no self-avoidance interactions, $N_0 \approx 600$. For a slightly self-avoiding polymer like dsDNA in physiological buffer, $N_0 \approx 800$. What this means is that to have a $1/e$ probability to find even one knot along a dsDNA, it has to be $800 \times 150 = 120,000$ bp long! (the long persistence length of DNA helps make this number impressive). Even more incredibly, for a self-avoiding polymer, $N_0 \approx 10^6$! This remarkable fact is theoretically understood only on the basis of numerical simulations: see Ref. [58].

Experiments on circular DNAs are in good quantitative agreement with statistical mechanical results for the semiflexible polymer model including DNA self-avoidance interactions. For example, it is found that the probability of finding a knot generated by thermal fluctuations for a 10 kb dsDNA is only about

0.05. [59, 60]. Topoisomerases are by themselves (using ATP) able to push this probability down, by a factor of between 10 and 100 [56].

4.4.3. Condensation-resolution mechanism for disentangling long molecules

Although topoisomerases seem to be able to help get rid of entanglements, there must be other mechanisms acting in the cell to completely eliminate them. Here I'll mention a very simple model that might give you an idea of how this machinery works.

Suppose we have a long dsDNA of length L , in the presence of some proteins which act to fold DNA up along its length. We imagine that these proteins cannot 'cross-link' DNA segments, but that they can only compact the molecule along its length. As these proteins bind, we imagine that they modify the total contour length to be $L' < L$, and the effective persistence length to be $A' > A$.

If these proteins bind slowly in the presence of type-II topoisomerases so that the knotting topology can reach close to equilibrium, then the unknotting probability will have the form:

$$P_{\text{unknot}} = \exp\left(-\frac{L'/A'}{N_0}\right) \quad (4.14)$$

As you can see, gradually compacting (decreasing L') while stiffening (increasing A') DNA can drive knotting out of it; unknotting ('entanglement resolution') will occur on progressively larger length scales as this condensation process proceeds.

Problem 40: Consider a condensation process which gradually condenses DNA a DNA of length L by folding it along its length, to make a progressively thicker fiber, of length L' and cross-section radius r' . If volume is conserved during condensation, and if the effective Young modulus of the fiber is a constant, find the unknot probability as a function of L'/L .

This simple model gives some idea of how proteins which structure DNA can play a role in controlling its entanglement at short enough scales where we can use polymer statistical mechanics. However, at the large scale of a whole chromosome cross-links do occur: they are necessary to fit the chromosome into the cell! It is also possible to envision a process where chromosome condensation by cross-linking also can drive out entanglements as long as the cross-links are transient [61].

5. DNA-Protein Interactions

So far we mostly talked about DNA by itself in buffer, focusing on the double helix's physical properties. In the cell, DNA is covered with proteins. You can think of DNA as a long string of thickness 2 nm, and the proteins as little particles of diameters of 1 to 10 nm plastered all along the string's length. The DNA plus all the proteins bound to it make up the biologically active chromosome.

Some proteins which bind to DNA are primarily *architectural*, folding and wrapping DNA so as to package it inside the cell. Other proteins have primarily *genetic* functions, interacting with particular sequences, usually between 4 and 40 bp in length. Of course, these two functions can be mixed: as examples proteins which tightly fold up DNA will likely repress gene expression; the expression of genes likely cannot occur without changes in DNA folding architecture.

Proteins which interact with DNA tend to be sorted into two groups: *Sequence-nonspecific* proteins that stick anywhere along the double helix; *Sequence-specific* proteins that bind to particular sequences very strongly, and to other sequences only relatively weakly.

Most proteins involved in chromosome architecture have a mainly nonspecific interaction with DNA. Such interactions are often electrostatic in character, and can be disrupted with high salt concentrations. Examples are the histone proteins of the nucleosome, and nonspecific DNA-bending proteins such as HU from *E. coli* or HMG proteins from eukaryote cells. Note that nonspecifically-interacting proteins usually bind better to some sequences than others, but not a whole lot better. Under physiological conditions, nonspecifically interacting proteins usually bind to DNA once their concentration is in the 10 to 1000 nM range.

Sequence-specific interactions occur via chemical interactions which depend on the structure of the bases. These are not usually electrostatic in character (most of DNA's charge is on the phosphates, which are common to all the bases) although most sequence-specific proteins also have a nonspecific interaction with DNA which is to some degree electrostatic. Examples of sequence-specific interactions include transcription factors and restriction enzymes. Sequence-specific proteins can often bind their targets at concentrations well below 1 nM. This high level of affinity is necessary: the concentration of transcription factors in *E. coli* can be as little as one per cubic micron (i.e., one per *E. coli* cell), which in molar units is $(1/6 \times 10^{23} \text{ Mol}) / (10^{-15} \text{ litre}) = 1.6 \text{ nM}$. In the human cell with a nucleus of volume $\approx 10^3 \mu\text{m}^3$, affinities in the picomolar range are needed to bind sequence-specific proteins to their targets with reasonably high probability.

5.1. How do sequence-specific DNA-binding proteins find their targets?

How long does it take for a protein to find a single specific target in a large DNA molecule? There is a long history of test-tube experiments studying this classic biophysics problem: one model system has been the *E. coli* protein lac repressor. Let's follow one protein of diameter d as it moves by diffusion to a target of size a ; we'll suppose that the targets are present in solution at concentration c .

5.1.1. Three-dimensional diffusion to the target

In the absence of any nonspecific interaction effects, the protein will diffuse through space until it hits the target. To analyze how long it takes for the protein to find one target, let's divide space up into boxes of volume $V = 1/c$ each containing one target. We'll further divide each box up into 'voxels' of volume a^3 ; one voxel is the target. Since the protein moves by diffusion, its trajectory will be a random walk in the box. The number of steps at scale a that must be taken to move across the box (of edge $V^{1/3}$) is $V^{2/3}/a^2$; the probability of finding the target before the protein leaves the box is therefore $a/V^{1/3}$ (this result is often called 'diffusion to capture'). The time that this search occurs in is just the diffusion time $V^{2/3}/D$, where $D = k_B T / (3\pi\eta d)$ is the protein diffusion constant.

Once the protein leaves one box of volume V , the same search starts over in an adjacent box: this will occur $V^{1/3}/a$ times before a target is found. So, the total time required for our protein to find a target by simple three-dimensional diffusion is

$$\tau_{3d} = \frac{V^{1/3}}{a} \times \frac{V^{2/3}}{D} = \frac{V}{Da} \quad (5.1)$$

Note that this formula could apply in solution where there is one target per solution volume V , or to targeting in a cell compartment of volume V . In the latter case, the same volume is searched over and over until the target is found.

If we convert V to concentration c , our result is $1/\tau_{3d} = 4\pi Dac$. The factor of 4π comes from a more detailed calculation of diffusion to capture, originally due to Smolochowski [62]. This rate is proportional to concentration; biochemists usually describe the rate of this type of bimolecular reaction by normalizing the actual rate by concentration, leaving the *association rate*:

$$k_a = 4\pi Da = \frac{4\pi k_B T a}{3\pi\eta d} \quad (5.2)$$

Since the target size is less than the protein size, we find that the maximum association rate is the prefactor $4\pi k_B T / (3\pi\eta) \approx 4 \times 10^{-18} \text{ m}^3/\text{s} \approx 10^8 \text{ M}^{-1}\text{s}^{-1}$.

This rate is often referred to as the ‘diffusion-limited reaction rate’.

5.1.2. Nonspecific interactions can accelerate targeting

Experiments in the 1970s showed that lac repressor binds its target at closer to $k_a \approx 10^{10} \text{ M}^{-1}\text{s}^{-1}$ which was initially quite a puzzle. However, Berg, Winter and von Hippel proposed and experimentally supported a solution to this paradox [63]. They realized that lac repressor also had a *nonspecific* interaction with DNA, and that this nonspecific interaction could make the target effectively larger than the protein and the binding site!

The picture is that when lac repressor first hits DNA, the nonspecific interaction allows it to ‘slide’ randomly back and forth along the DNA, exploring a region of length ℓ_{sl} before it dissociated. Now, the target size is increased to be ℓ_{sl} , so the time we will have to spend doing three-dimensional diffusion will be reduced to $1/(4\pi D_{\text{sl}}\ell_{\text{sl}}c)$, where D_{sl} is the diffusion constant for the ‘sliding’ motion.

However, it is not yet clear if this will really accelerate k_a , since each sliding event will eat up a time of roughly $\ell_{\text{sl}}^2/D_{\text{sl}}$, and to be sure to find the target, L/ℓ_{sl} sliding events have to occur, requiring a total one-dimensional diffusion time of $L\ell_{\text{sl}}/D_{\text{sl}}$.

So, the total time required to find the target by this ‘facilitated diffusion’ process is

$$\tau_{\text{fac}} = \frac{1}{4\pi D_{\text{sl}}\ell_{\text{sl}}c} + \frac{L\ell_{\text{sl}}}{D_{\text{sl}}} \quad (5.3)$$

If we write the ratio of this and the three-dimensional result 5.2, we obtain

$$\frac{k_{\text{a,fac}}}{k_{\text{a,3d}}} = \frac{\tau_{\text{3d}}}{\tau_{\text{fac}}} = \frac{\ell_{\text{sl}}/a}{1 + 4\pi \frac{D}{D_{\text{sl}}} \ell_{\text{sl}}^2 Lc} \quad (5.4)$$

As long as ℓ_{sl} is not too long, the nonspecific interaction does accelerate the reaction rate. This basic model and its experimental study were introduced by Berg, Winter and von Hippel [63].

A key feature of 5.3 is that for fixed total DNA length L and target concentration $c = 1/V$ there is an *optimal* ℓ_{sl} :

$$\ell_{\text{sl}}^* \approx \sqrt{V/L} \quad (5.5)$$

where we have dropped the 4π and the ratio of diffusion constants. For the *E. coli* cell, $V \approx 10^9 \text{ nm}^3$ and $L \approx 10^6 \text{ nm}$, indicating $\ell_{\text{sl}}^* = 30 \text{ nm} = 100 \text{ bp}$. This is the sliding length inferred for lac repressor from biochemical experiments on

facilitated diffusion [63]. This suggests that ℓ_{sl} for lac repressor is optimized to facilitate its targeting *in vivo* [64].

Note that the above arguments are independent of the conformation of the large target-containing DNA [64]. At the level of scaling, the association rate constant should be the same for either a relaxed random coil, or a fully stretched molecule. This conclusion, and the dependence of k_a on concentration and total DNA length (see Eq. 5.4), could be checked using individual micromanipulated DNA molecules, given a way to detect when a protein binds its target along a long DNA. Recently Wuite and coworkers have demonstrated micromechanical detection of target recognition by the restriction enzyme EcoRV, which induces a DNA bend when it binds to its target sequence [65].

5.2. Single-molecule study of DNA-binding proteins

A number of groups are working at present on experiments looking at protein-DNA interactions using single-DNA micromanipulation, which we touched on in the last paragraph. The basic idea is to study proteins which change DNA mechanical properties, for example, by putting bends or loops into the double helix. Then, the binding of the proteins can be monitored via the force-extension response of the molecule. To give an idea of what can be obtained from this kind of study, I'll describe a few simple models.

5.2.1. DNA-looping protein: equilibrium 'length-loss' model

Consider a protein which binds to a double helix under tension f , resulting in a reduction in contour length available for extension by amount ℓ . If we suppose that the binding energy of the protein in $k_B T$ units is ϵ and its bulk concentration is c , we can write down a simple two-state model for the free energy of the protein-DNA complex as a function of the occupation of the protein [66]

$$\ln Z_n = \frac{(L - n\ell)}{A} \gamma(\beta A f) + n(\epsilon + \ln vc) \quad (5.6)$$

where $n = 1$ for protein bound, $n = 0$ for protein free in solution. The first term is the DNA stretching free energy: when the protein binds, the contour length of extensible DNA is reduced by ℓ . The last two terms give the free energy for a bound protein including the entropy cost of its removal from free solution. A factor of protein volume v is included to make the inside of the log dimensionless.

We can immediately find the probability for the protein to be bound:

$$\frac{P_{\text{on}}}{P_{\text{off}}} = \frac{c}{K_d} e^{-\frac{\ell}{A} \gamma(\beta A f)} \approx e^{-\beta \ell f + \ln(c/K_d) + \dots} \quad (5.7)$$

where we have defined $K_d = e^{-\epsilon}/v$, the ‘dissociation constant’ or concentration at which the protein is half-bound for zero force; the final term gives the large-force limit (recall $\gamma(x) = x + \sqrt{x} + \dots$). If the complex is bound at low force ($c > K_d$), it will stay bound until a force threshold is reached. Roughly, this characteristic force is $f^* \approx \ln c/K_d/\ell$; beyond this force, the DNA-protein complex will open up. In an experiment where extension versus force can be used to monitor the stability of a protein-DNA complex, one can therefore make a measurement of the zero-force K_d – if one can observe equilibrium (on-off fluctuations) at the force-induced dissociation point.

At zero force, Finzi and Gelles have observed such opening-closing fluctuations using lac repressor, which is able to bind two targets so as to form a loop [67]. Experiments of this type at nonzero force have recently been done for the gal repressor protein, which is also able to trap a DNA loop [68].

Problem 41: A single nucleosome involves 146 bp of dsDNA wrapped around a 10 nm-diameter octamer of ‘histone’ proteins. The total binding free energy of this structure is thought to be roughly $30 k_B T$ (see [66] and references therein). Using the length-loss model, estimate the force at which nucleosomes ought to be released by applied force. You might like to compare your result with the 15 pN force observed in many ‘nucleosome-pop-off’ experiments (see, e.g., Bennink et al [69], and Brower-Toland et al [70]). What do you think is the origin of the discrepancy?

Problem 42: An elegant loop-formation experiment in which equilibrium fluctuations can be observed uses a single piece of RNA under force [71], which can fold into a hairpin ‘helix-loop’ structure, held together by base-pairing and base-stacking interactions. When under force, opening and closing kinetics can be observed on roughly 1 sec timescales. Using the single-strand unzipping results, construct a two-state model for the loop opening and closing (recall the free energy per base stabilizing the coil is roughly $2.5 k_B T$, while the persistence length of the single-stranded backbone is roughly $b = 1$ nm), and estimate how $P_{\text{closed}}/P_{\text{open}}$ varies with applied force.

5.2.2. Loop formation kinetics

The previous section supposes that one can observe on-off fluctuations in the presence of force. However, it is possible, given strong protein-DNA interactions, that spontaneous dissociation will be essentially unobservable. In this case, one will observe the on-kinetics only. In the case of a small DNA loop, the complex formation rate will involve a barrier made of two components: the free energy cost of pulling in a length ℓ of DNA as discussed above, plus the free energy cost of making the DNA loop [72] (recall the ‘J-factors’ of Sec. 2.2.5). Putting these

together gives a simple estimate of the loop formation rate

$$k_{\text{loop}} \approx k_0 \exp[-\beta \ell f + \ln v J(\ell)] \quad (5.8)$$

where J is the juxtaposition ‘J-factor’ relevant to the reaction (see 2.11, 2.12), and where v is the ‘reaction volume’, the volume of the region of space in which the reaction can proceed.

The exponential dependence on force will effectively shut off the reaction for forces significantly larger than $f^* \approx (k_B T / \ell) \ln J$. The drastic force-quenching of irreversible loop formation by force has been verified using an essentially exact transfer-matrix calculation [73]; experimental data that quantitatively test this model have not yet been reported, although moderate forces have been observed to completely eliminate looping by the two-site-binding restriction enzyme BspMI [74]. Finally, if a series of loop binding sites are available on a long DNA, the rate (5.8) will be proportional to the velocity at which the DNA end retracts due to loop formation [72].

For large loops ($\ell > A$) and low forces ($f < k_B T / \ell < k_B T / A$) random-coil fluctuations will reduce the severity of the force-quenching effect, giving $\ln k_{\text{loop}}(f) \approx \ln k_{\text{loop}}(0) - A \ell (\beta f)^2$. This regime is one of quite low force, since $k_B T / A \approx 0.1$ pN.

5.2.3. DNA-bending proteins

In vivo, sequence-specific interactions such as the DNA-looping examples discussed in the previous section, take place on DNA which is in large part covered by other proteins. Much of this protein serves to package the DNA in the cell, and has sequence-nonspecific interaction with DNA. One of the classes of DNA-packaging proteins that has already been shown to be quite interesting to study using micromechanical approaches are *DNA-bending proteins*: examples are HU (a primary DNA-bending protein from *E. coli*) [75] and various HMG proteins (the prevalent nonspecific DNA-bending proteins in eukaryotes) [76], both of which are present at high copy number (roughly one per 200 bp of DNA) in vivo. These two proteins generate roughly 90° bends where they bind DNA.

An important chromosome-packaging role of DNA-bending proteins is well established for bacteria [77]. In eukaryotes, the compaction of chromosomal DNA occurs in part via wrapping in nucleosomes (the histones in which can be considered to bend DNA); however, many non-histone proteins including the DNA-bending HMGs act to establish ‘higher-order chromatin structure’, a term which is often invoked as a cover for our ignorance of the chromosome structure at supra-nucleosomal scales.

Traditional biochemical approaches to the study of DNA-bending proteins in-

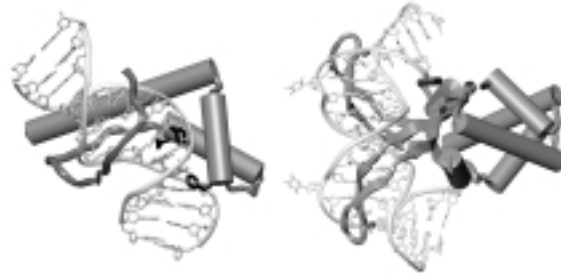


Fig. 15. Two examples of DNA-bending proteins with known structures. Left: NHP6A, an HMG protein from yeast, bound to DNA. Right: HU, a DNA-bending protein from *E. coli*, bound to DNA. Both bend DNA by roughly 90° , over roughly 10 bp. Recall the 2 nm diameter of the double helix. Image courtesy of Prof. R.C. Johnson, UCLA.

clude study of changes in electrophoretic mobility resulting from binding of protein, among other bulk solution phase methods. Single-DNA experiments offer a complementary view of such proteins: in the presence of nonspecifically binding DNA-bending proteins we can expect a modification of the force-extension behavior of bare DNA we have discussed above. Qualitatively, we can anticipate that a DNA will be more difficult to extend if it contains sharp bends due to bound proteins along its length; we expect a *shift* of the force-extension curve, to higher forces, or a *compaction* of DNA relative to the ‘bare’, protein-free case.

This force-shift effect, similar to that discussed for the two-state ‘lost length’ model [66] discussed in the previous section, has been observed for a few DNA-bending proteins. The first experiment of this type was carried out with the *E. coli* protein IHF [78]; the DNA-bending protein HU has been similarly studied [79, 80].

Eukaryote DNA-bending proteins have also been studied and show similar effects [80]. Fig. 16 shows force-shift data for the HMG-type DNA-bending protein NHP6A from yeast. As more protein is added, the force curve shifts to larger forces. Something apparent in those NHP6A data is that at moderately high forces ≈ 10 pN, even when the DNA is covered with protein, the protein-DNA complex length can be forced to near the original bare DNA contour length (about 16 microns in this experiment on λ -DNA). In these experiments it was

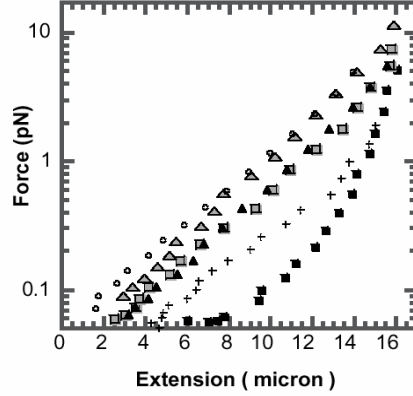


Fig. 16. Experimental force-extension data for individual λ -DNA molecules in the presence of the protein NHP6A (the HMG protein shown in Fig. 15). Black filled squares show the force-extension response of bare DNA with no NHP6A present. The other symbols show the force-extension response of single DNAs at various NHP6A concentrations: +, 3 nM; gray filled squares, 5 nM; black filled triangles, 10 nM; gray filled triangles, 33 nM; open circles, 75 nM. Reproduced from Ref. [80].

determined that the protein was not being dissociated by force; therefore the DNA-protein complexes are *flexible*. The *E. coli* DNA-bending proteins IHF and HU also show this effect [79, 80].

Since thermally excited DNA bends take place over molecule segments of about the persistence length $A \approx 150$ bp long, the force-extension response of a single DNA should be quite strongly perturbed when there is more than one protein bound per persistence length $A = 150$ bp. This indicates that strong force shifts such as the 33 nM data shown in Fig. 16 correspond to at least several proteins bound per persistence length, i.e., quite dense ‘coverage’.

We can construct a model for this kind of experiment by supposing that where proteins bind, they generate localized (and flexible) bends. The discrete-tangent model introduced in Sec. 2 can be modified to include the effect of DNA-bending proteins [21]:

$$\beta E = \sum_k \left[(1 - n_k) \frac{a}{2} |\hat{\mathbf{t}}_{k+1} - \hat{\mathbf{t}}_k|^2 + n_k \frac{a'}{2} (\hat{\mathbf{t}}_{k+1} \cdot \hat{\mathbf{t}}_k - \cos \psi)^2 - \mu n_k - \beta f \hat{\mathbf{z}} \cdot \hat{\mathbf{t}}_k \right] \quad (5.9)$$

which contains, in addition to tangent vectors $\hat{\mathbf{t}}_k$, two-state occupation degrees

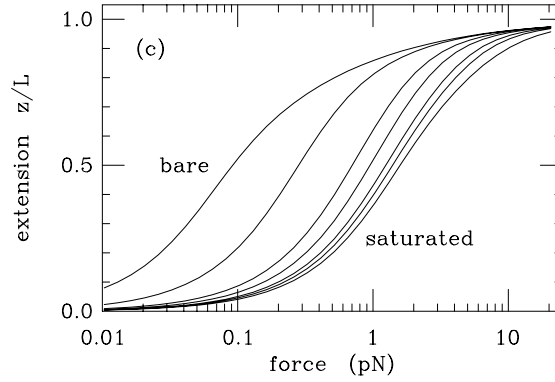


Fig. 17. Theoretical force-extension behavior for discrete-semiflexible polymer model including DNA-bending protein effects. Leftmost curve shows bare DNA (no additional bends), while the curves shifted progressively to the right show the shift to higher forces generated by a series of higher protein concentrations (chemical potential $\mu = -3.91, -2.30, -1.61, -0.69, 0$ and 3.00 ; corresponding low-force binding site occupation of 22%, 60%, 72%, 89%, 93% and 100%). Reproduced from Ref. [21].

of freedom n_k which are either 0 or 1. The n_k are controlled by the ‘chemical potential’ μ ; for simple, independent equilibrium binding, $\mu = \text{constant} + \ln c$ where c is the bulk protein concentration [21, 66]. When $n_k = 0$, the joint between segments k and $k+1$ has the usual semiflexible polymer bending energy. However, when $n_k = 1$, the joint has a preferred angle ψ , and a modified rigidity a' .

This type of model has as degrees of freedom, in addition to the orientation vector \hat{t}_k , an additional ‘scalar field’ n_k describing protein binding. Models of this general type have been introduced to describe DNA overstretching (in this case the scalar variable describes conformational change of the double helix) [81–83] as well as protein-DNA interactions [36, 84]. The idea common to these works is the use of additional variables (n_k) to account for DNA conformational change, and to examine how the coupling of those variables to the tangent vectors \hat{t}_k modifies force-extension behavior.

Fig. 17 shows the force-shift effect for a numerical transfer-matrix calculation of the partition function for this model, using segment length $b = 1$ nm, double helix stiffness $a = 50$ (persistence length 50 nm), a bend angle $\psi = 90^\circ$, and a protein-bent-DNA bending constant $a' = 10$ (quite flexible). The result is, as the binding chemical potential μ is varied, a gradual shift of the force curve to larger forces qualitatively similar to that seen experimentally. The segment length b defines the maximum density of proteins that can bind, i.e., one per 3 bp for the parameters used in Fig. 17.

5.2.4. Analytical calculation for compaction by DNA-bending proteins

I now present a simple computation which gives some insight into the numerical solution of the DNA-bending-protein model (Eq. 5.9 and Ref. [21]). The aim here is to look at a ‘weakly distorting’ limit, where we can see how the basic semiflexible-polymer model is modified by the presence of DNA-bending proteins in the limit that they only slightly deflect $\hat{\mathbf{t}}$.

The portion of Eq. 5.9 that describes how one of the protein occupation variables n is coupled to two adjacent tangent vectors $\hat{\mathbf{t}}$ and $\hat{\mathbf{t}}'$ is

$$\frac{a}{2} |\hat{\mathbf{t}}' - \hat{\mathbf{t}}|^2 (1 - n) + \left[\frac{a'}{2} (\hat{\mathbf{t}}' \cdot \hat{\mathbf{t}} - \cos \psi)^2 - \mu \right] n \quad (5.10)$$

Using $1 - |\hat{\mathbf{t}}' - \hat{\mathbf{t}}|^2 = 2\hat{\mathbf{t}}' \cdot \hat{\mathbf{t}}$ and dropping constants and terms beyond quadratic order in $\hat{\mathbf{t}}' - \hat{\mathbf{t}}$ we have

$$\frac{a}{2} |\hat{\mathbf{t}} - \hat{\mathbf{t}}'|^2 - \frac{a + a'(1 - \cos \psi)}{2} n |\hat{\mathbf{t}}' - \hat{\mathbf{t}}|^2 - \mu n \quad (5.11)$$

where the chemical potential μ has been shifted by a constant. Keeping only the quadratic terms supposes that the DNA-bending effect of any protein present is a weak perturbation to the straightening generated by the external tension f .

We will now rewrite this model in continuum form. The discrete nature of n may be enforced by adding an n^2 term to make the amplitude of n -fluctuations well-defined. The model including force is now:

$$\int ds \left[\frac{A}{2} \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 + \frac{1}{2} n^2 - \mu n - \beta f \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} - g n \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 \right] \quad (5.12)$$

The parameter $g \propto a + [1 - \cos \psi] a'$ describes the coupling of the protein density to bending, and should be positive for DNA-bending proteins like HU, HMGB1 and NHP6A, and negative for DNA-stiffening proteins.

For fixed μ , fluctuations of n can be integrated, to obtain

$$\int ds \left[\frac{1}{2} (A - 2g\mu) \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 - \beta f \hat{\mathbf{z}} \cdot \mathbf{t} \right] \quad (5.13)$$

A shift of chemical potential (binding free energy) causing binding of protein (positive shift of μ) causes a shift in the effective persistence length, to $A_{\text{eff}} = A - 2g\mu$. The reduction in persistence length generated by DNA-bending proteins ($g > 0$) causes a shift in the force-extension curve to larger forces, similar to that obtained in the transfer-matrix solution of the discrete model (Fig. 17).

Problem 43: Show that the expectation value of the ‘bound protein density’ along the DNA takes the form

$$\langle n \rangle = \mu + g \left\langle \left| \partial \hat{\mathbf{t}} / \partial s \right|^2 \right\rangle$$

If you want to compute the derivative term, you must include an upper wavenumber cutoff q_{max} (equivalently, a finite lattice constant $b = 2\pi/q_{\text{max}}$). You will find that $\langle n \rangle$ diverges as you take the cutoff away (the limit $q_{\text{max}} \rightarrow \infty$).

Problem 44: The previous calculations supposed fluctuating protein occupation n . Alternately, DNA-bending proteins may be able to bind to DNA *permanently*. This may be treated in an averaged fashion by supposing n in Eq. 5.12 to take a fixed, positive value. For this case, use Eq. 5.12 to show $A_{\text{eff}} = A - 2gn$.

Problem 45: A more complete treatment of the n fluctuations would replace n^2 in Eq. 5.12 with $\frac{1}{2} [mn^2 + v(dn/ds)^2]$, where m and v are positive parameters. Find the form of A_{eff} in this case.

5.2.5. Effects of twisting of DNA by proteins

Proteins which bend DNA often untwist it (e.g., HU [75]), and this effect might be probed using single-DNA experiments. This has been explored for the discrete-tangent model [21], but the simple continuum calculation presented above can also be generalized to the case where DNA is twisted:

$$\int ds \left[\left(\frac{A}{2} - gn \right) \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 - \beta f \hat{\mathbf{z}} \cdot \mathbf{t} + \frac{1}{2} n^2 - \mu n + \frac{C}{2} (\Omega - \chi n)^2 \right] \quad (5.14)$$

The last term accounts for the twist energy including a shift in its zero due to binding of protein [21]. The new field $\Omega(s)$ is just the twist density [51]; the total twist is $\int ds \Omega(s) = 2\pi \Delta T w$. The DNA-twisting parameter χ is negative for a protein which underwinds DNA, since in this case bound protein ($n > 0$) shifts the minimum of the twisting energy to $\Omega < 0$.

Expanding the last term in (5.14) gives

$$\int ds \left[\frac{A}{2} \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 - \beta f \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} + \frac{m}{2} n^2 + \frac{C}{2} \Omega^2 - \left(\mu + g \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 + C \chi \Omega \right) n \right] \quad (5.15)$$

where $m = 1 + \chi^2 C$. Completing the square and integrating out $n = \left(\mu + g \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 + C \chi \Omega \right) / m$,

$$\int ds \left[\frac{1}{2} \left(A - \frac{2g}{m} [\mu + C \chi \Omega] \right) \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 - \beta f \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} + \frac{C}{2} \left(1 - \frac{C \chi^2}{m} \right) \Omega^2 - \frac{C \mu \chi}{m} \Omega \right] + \dots \quad (5.16)$$

where constants and terms beyond quadratic order in $\hat{\mathbf{t}}$ have been dropped.

We use $\Delta T w = \Delta L k - W r$ in the form $\int ds \Omega / 2\pi = \sigma L / h - W r$ to eliminate the twist (recall h is the DNA helix repeat). The terms of quadratic order in $\hat{\mathbf{t}}$ (recall from Eq. 4.9 that $W r$ is $\mathcal{O}(\hat{\mathbf{t}}^2)$) are:

$$\int ds \left\{ \frac{1}{2} \left(A - \frac{2g}{m} \left[\mu + \frac{2\pi C}{h} \chi \sigma \right] \right) \left| \frac{\partial \hat{\mathbf{t}}}{\partial s} \right|^2 - \beta f \hat{\mathbf{z}} \cdot \hat{\mathbf{t}} - \frac{2\pi C}{h} \left(\left[1 - \frac{C \chi^2}{m} \right] \sigma - \frac{h}{2\pi m} \chi \mu \right) 2\pi W r + \dots \right\} \quad (5.17)$$

This energy describes Gaussian fluctuations of $\hat{\mathbf{t}}$ and has the same form as Eq. 4.10, but with shifted persistence length A and twist density σ values:

$$\begin{aligned} A &\rightarrow A_{\text{eff}} = A - \frac{2g}{m} \left[\mu + \frac{2\pi C}{h} \chi \sigma \right] \\ \sigma &\rightarrow \sigma_{\text{eff}} = \left[1 - \frac{C \chi^2}{m} \right] \sigma - \frac{h}{2\pi m} \chi \mu \end{aligned} \quad (5.18)$$

These effective quantities can be used to write the extension in the high-force expansion form of Eq. 4.12:

$$\langle \hat{\mathbf{t}} \cdot \hat{\mathbf{z}} \rangle = 1 - \frac{1}{\sqrt{4\beta A_{\text{eff}} f}} - \frac{1}{2} \left(\frac{2\pi C \sigma_{\text{eff}}}{h} \right)^2 \left(\frac{k_B T}{4A_{\text{eff}} f} \right)^{3/2} + \dots \quad (5.19)$$

At very high force, the $1/\sqrt{f}$ term dominates, and will cause a shift of the peak of extension as a function of σ , to where $\chi\sigma < 0$. In this regime, binding of protein contracts the polymer, so the peak extension is found where the bound protein is the *least*. For a protein which untwists DNA ($\chi < 0$), the peak of extension occurs at very high force for $\sigma > 0$.

At lower, but still high force, the $1/f^{3/2}$ term can compete. This term vanishes at the zero of σ_{eff} , where the torsion in the molecule, and therefore its chiral coiling, are minimized. In the DNA-untwisting-protein case, this is for $\sigma < 0$. Therefore, for a DNA-untwisting protein ($\chi < 0$), the shift of the peak in extension as a function of force can change from $\sigma > 0$ for large force, to $\sigma < 0$ for lower force. This gives some analytical insight into the same effect found in numerical calculations of Ref. [21].

Problem 46: Starting from Eq. 5.14 compute A_{eff} and σ_{eff} for the case that the bound protein density does not fluctuate (consider n a fixed, positive constant as in Problem 44).

5.2.6. Surprising results of experiments

Surprises have come out of the experimental studies of DNA-bending proteins. The E. coli HU protein, also known to bend DNA, has also been studied using single-DNA micromanipulation [79, 80]. While a compaction effect (shift to higher forces) has been observed for HU, at very high concentrations the force curve shifts *back* to lower forces than for bare DNA! In Ref. [79] it was suggested that at high concentration, HU is capable of essentially polymerizing along the double helix, and that HU-covered DNA is essentially straight and actually stiffer than bare DNA. This explained the shift of the force-extension curve to lower forces than for bare DNA (the result of increasing the persistence length A in Eq. 2.23). This ‘bimodal’ behavior of HU is a good example of a feature of a protein that was not well established using standard biochemical assays, but which was rather straightforward to infer using single-DNA methods.

A second experimental surprise associated with DNA-bending proteins is that once these proteins reach a certain high coverage on the double helix, they no longer will spontaneously dissociate, even when the protein in solution is removed (established for HU, HMGB1 and NHP6A in Ref. [80]). This high level of stability of protein densely assembled onto DNA may be a result of cooperative interactions between adjacent proteins, i.e., a term $n_k n_{k+1}$ in 5.9. While such stability due to cooperativity is expected for some proteins such as RecA, which is well known to form a polymerized ‘coat’ on the double helix (and which has been studied using single-DNA methods [85]), this behavior was not expected for HU, HMGB1 and NHP6A.

These results suggest that there may be appreciable restrictions on the appli-

cability of the widely assumed ‘two-state’ equilibrium binding models for interaction of these non-sequence-specific proteins with DNA. Proteins like HU, HMGB1 and NHP6A are in part responsible for folding up DNA in vivo, and their ability to form highly stable complexes may be important to chromosome ‘architecture’.

6. Conclusion

This course has focused on the micromechanics of individual DNA molecules studied in single-DNA micromanipulation experiments. A major focus has been on the use of statistical-mechanical models to describe how spontaneous thermal fluctuations and external applied stresses change the conformation of large DNAs.

With the double helix we have a favorable situation since the atomic and even base-pair scales (< 1 nm) are well separated from the ≈ 100 nm scale over which thermal fluctuations are able to bend the double helix. This separation of scales allows us to describe the elastic response of a dsDNA from essentially zero force, up to 10 pN or so, using only one ‘effective’ elastic constant, the persistence length $A = 50$ nm [20]. As force is increased beyond 0.1 pN, we have seen that a single ‘correlation length’ $\sqrt{k_B T A / f}$ describes the essentially Gaussian fluctuations of the tangent vector away from the force direction.

However, this course has also emphasized how quickly this pleasant long-wavelength description can fall by the wayside. Unzipping the double helix (Sec. 3) requires only about 15 pN forces, and converts the staid double helix into two single-stranded DNAs, which can by no means be described using a single elastic constant! [86] Under some circumstances ssDNA is a polymer with a strongly wavenumber- (and therefore force-) dependent persistence length [20, 36]. The interplay of the backbone polymer elasticity, Coulomb interactions, and self-adhesion via attractive base-base interactions [35] makes single-stranded DNA impossible to describe using a single persistence length. Theories to describe ssDNA by necessity have a richer structure.

Something similar happens when one twists the double helix: at low forces (roughly < 0.3 pN), the response of the double helix to twisting can be captured using the twist-persistence-length model discussed in Sec. 4.2, which really just adds one more elastic constant to describe twist deformations. However, for even rather low forces, underwinding can rather easily drive strand separation.

Description of the opening of the double helix, and other abrupt force- and torque-driven structural transitions of the double helix such as ‘overstretching’ [22, 23], need additional degrees of freedom on top of the elastic degrees of free-

dom used to describe relatively smooth bends and twists. For example, conversion of the double helix to ssDNA requires at least a ‘helix-coil’ Ising-type degree of freedom.

A few of the examples discussed in these notes have used additional degrees of freedom of this type to describe localized ‘defects’ along the double helix. Localized strand-separated regions excited thermally may be responsible for the anomalously large cyclization probability of short DNAs observed by Cloutier and Widom [12, 13]. Similar models can be used to describe proteins which generate local bends or other distortions of the double helix [21]. Inevitably we can expect these ‘local defects’ to interact with one another along the double helix, and to possibly facilitate self-organization of stable protein-DNA complexes, as observed in Ref. [80]. Those experiments show that even at the level of a single species of protein interacting with a single DNA, self-assembled protein-DNA complexes can strongly resist disassembly. Fully understanding the nonequilibrium phenomena in experiments such as Ref. [80] requires at the least, theories that include kinetics.

Acknowledgments

These lectures include results of research done jointly with Eric Siggia, Didier Chatenay, Jean-Francois L  g  r, Abhijit Sarkar, Simona Cocco, Sumithra Sankararaman, Dunja Skoko, Jie Yan, Michael Poirier, Steven Halford, Monte Pettitt, and Michael Feig, whose many insights I gratefully acknowledge. I am also grateful for advice and help, including experimental data, from David Bensimon, Vincent Croquette, Terence Strick, Jean-Francois Allemande, Carlos Bustamante, Steve Smith and Nick Cozzarelli. I thank NATO and the CNRS for making it possible for me to visit Les Houches to present these lectures. This work was supported in part by the US National Science Foundation, through Grants DMR-0203963 and MCB-0240998.

References

- [1] M.D. Wang, M.J. Schnitzer, H. Yin, R. Landick, J. Gelles, S.M. Block, *Science* **282** 902-907 (1998).
- [2] D.S. Goodsell, *The machinery of life* (Springer-Verlag, New York, 1992).
- [3] M. Feig, J.F. Marko, M. Pettitt, Microscopic DNA fluctuations are in accord with macroscopic DNA stretching elasticity without strong dependence on force field choice, *NATO ASI Series: Metal Ligand Interactions*, ed. N. Russo (Kluwer Academic Press), 193-204 (2003).
- [4] L.D. Landau, E.M. Lifshitz, *Theory of Elasticity* (Pergamon, NY, 1986) Ch. II.

- [5] P.J. Hagerman *Ann. Rev. Biophys. Biophys. Chem.* **17**, 265 (1988).
- [6] V. Makarov, B. Pettitt, M. Feig, *Acc. Chem. Res.* **35**, 376 (2002).
- [7] K. Ogata, K. Sato, T.H. Tahirov, *Curr. Opin. Struct. Biol.* **13** 40 (2003).
- [8] H. Yamakawa, W.H. Stockmayer, *J. Chem. Phys.* **57**, 2843 (1972).
- [9] J. Shimada, H. Yamakawa, *Macromolecules* **17**, 689 (1984).
- [10] J. Santalucia Jr., *Proc. Natl. Acad. Sci. USA* **95** 1460 (1998).
- [11] D. Shore, J. Langowski, R.L. Baldwin, *Proc. Natl. Acad. Sci. USA* **78** 4833 (1981).
- [12] T.E. Cloutier, J. Widom, *Mol. Cell* **14** 355 (2004).
- [13] J. Yan, J.F. Marko, *Phys. Rev. Lett.* **93** 108108 (2004).
- [14] D.M. Crothers, T.E. Haran, J.G. Nadeau, *J. Biol. Chem* **265**, 7093 (1990).
- [15] S.B. Smith, L. Finzi, C. Bustamante, *Science* **258**, 1122 (1992).
- [16] C. Bustamante, S. Smith, J.F. Marko, E.D. Siggia, *Science* **265**, 1599 (1994).
- [17] A. Vologodskii, *Macromolecules* **27**, 5623 (1994).
- [18] T. Odijk, *Macromolecules* **28**, 7016 (1995).
- [19] T.T. Perkins, D.E. Smith, R.G. Larson, S. Chu, *Science* **268**, 83 (1995).
- [20] J.F. Marko, E.D. Siggia, *Macromolecules* **28** 8759 (1995).
- [21] J. Yan, J.F. Marko, *Phys. Rev. E* **68** 011905 (2003).
- [22] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, F. Caron, *Science* **271**, 792 (1996).
- [23] S.B. Smith, Y.J. Cui, C. Bustamante, *Science* **271**, 795 (1996).
- [24] J.R. Wenner, M.C. Williams, I. Rouzina, V.A. Bloomfield, *Biophys. J.* **82**, 3160 (2002).
- [25] T.R. Strick, J.-F. Allemand, D. Bensimon, V. Croquette, *Biophys. J.* **74** 2016 (1998).
- [26] K.M. Dohoney KM, J. Gelles, *Nature* **409** 370 (2001).
- [27] P.R. Bianco, L.R. Brewer, M. Corzett, R. Balhorn, Y. Yeh, S.C. Kowalczykowski, R.J. Baskin, *Nature* **409** 374 (2001).
- [28] M.N. Dessinges, T. Lionnet, X.G. Xi, D. Bensimon, V. Croquette, *Proc. Natl. Acad. Sci. USA* **101** 6439 (2004).
- [29] A. Yu. Grosberg, A. R. Khokhlov, *Statistical Physics of Macromolecules*, pp. 302-312, (AIP Press, Woodbury NY, 1994).
- [30] J.-F. Leger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, J.F. Marko, *Phys. Rev. Lett.* **83**, 1066 (1999).
- [31] J.-F. Leger, Ph.D. Thesis, l'Université Louis Pasteur Strasbourg I, Strasbourg, France (2000).
- [32] C. Bustamante, D. Smith, S. Smith, *Curr. Opin. Struct. Biol.* **10**, 279 (2000).
- [33] M. Rief, H. Clausen-Schaumann, H.E. Gaub, *Nat. Struct. Biol.* **6**, 346 (1999).
- [34] J.-L. Barrat, J.-F. Joanny, *Europhys. Lett.* **24** 333 (1993).
- [35] Y. Zhang, H.J. Zhou, Z.C. Ou-Yang, *Biophys. J.* **81**, 1133 (2001).
- [36] J. Yan, A. Sarkar, S. Cocco, R. Monasson, J.F. Marko, *Eur. Phys. J. E* **10**, 249-263 (2003).
- [37] Y. Seol, G.M. Skinner, K. Visscher, *Phys. Rev. Lett.* **93**, 118102 (2004).
- [38] U. Bockelmann, B. Essevaz-Roulet, F. Heslot, *Biophys. J.* **82**, 1537 (2002).
- [39] H. Clausen-Schaumann, M. Rief, C. Tolksdorf, H.E. Gaub, *Biophys. J.* **78**, 1997 (2001).
- [40] S. Cocco, J.F. Marko, R. Monasson, *Proc. Nat. Acad. Sci. USA* **98**, 8608 (2001).
- [41] S. Cocco, R. Monasson, J.F. Marko, *Phys. Rev. E* **66**, 051914 (2002).
- [42] P. Nelson, *Phys. Rev. Lett.* **80**, 5810 (1998).

- [43] P. Thomen, U. Bockelmann, F. Heslot, *Phys. Rev. Lett.* **88**, 248102 (2002).
- [44] D. Lubensky, D.R. Nelson, *Phys. Rev. Lett.* **85**, 1572 (2000); D. Lubensky, D.R. Nelson, *Phys. Rev. E* **65**, 031917 (2002).
- [45] C. Danilowicz, V.W. Coljee, C. Bouzigues, D.K. Lubensky, D.R. Nelson, M. Prentiss, *Proc. Natl. Acad. Sci. USA* **100** 1694 (2003).
- [46] C. Danilowicz, Y. Kafri, R.S. Conroy, V.W. Coljee, J. Weeks, M. Prentiss, *Phys. Rev. Lett.* **93** 078101 (2004).
- [47] J.D. Moroz, P. Nelson *Proc. Natl. Acad. Sci. USA* **94**, 14418 (1997); P. Nelson, *Biophys. J* **74**, 2501 (1998); J.D. Moroz, P. Nelson, *Macromolecules* **31** 6333 (1998).
- [48] T.C. Boles, J.H. White, N.R. Cozzarelli *J Mol Biol.* **213**(1990).
- [49] J.F. Marko, *Physica A* **296**, 289 (2001); J.F. Marko, *Physica A* **244**, 263 (1997).
- [50] J.F. Marko, E.D. Siggia, *Science* **265**, 506 (1994).
- [51] J.F. Marko, E.D. Siggia, *Phys. Rev. E* **52**, 2912 (1995).
- [52] C. Bouchiat, M. Mezard, *Phys. Rev. Lett.* **80** 1556 (1998); C. Bouchiat, M. Mezard, *Eur. Phys. J.* **E2**, 377 (2000).
- [53] A.E.H. Love, *A treatise on the mathematical theory of elasticity*, 4th edition (Dover, New York, 1944).
- [54] V. Rossetto, *Europhys. Lett.* **69**, 142 (2005).
- [55] A. Sarkar, J.-F. Leger, D. Chatenay, J.F. Marko, *Phys. Rev. E* **63**, 051903 (2001).
- [56] V.V. Rybenkov, C. Ullsperger, A.V. Vologodskii, N.R. Cozzarelli, *Science* **277**, 648 (1997).
- [57] D.W. Summers, S.G. Whittington, *J. Phys. A Math. Gen.* **21** 1689 (1988).
- [58] K. Koniaris and M. Muthukumar, *Phys. Rev. Lett.* **66** 2211 (1991).
- [59] S.Y. Shaw, J.C. Wang, *Science* **260**, 533 (1993).
- [60] V.V. Rybenkov, N.R. Cozzarelli, A.V. Vologodskii, *Proc. Natl. Acad. Sci. USA* **11** 5307 (1993).
- [61] J.F. Marko, E.D. Siggia, *Mol. Biol. Cell* **8** 2217 (1997).
- [62] H.C. Berg, *Random Walk in Biology* (Princeton, Princeton, 1993)
- [63] O.G. Berg, R.B. Winter, P.H. von Hippel, *Biochemistry* **20** 6929 (1981); R.B. Winter, P.H. von Hippel, (1981) *Biochemistry* **20** 6948 (1981); R.B. Winter, O.G. Berg, P.H. von Hippel, *Biochemistry* **20** 6961 (1981).
- [64] S.E. Halford, J.F. Marko *Nucl. Acids Res.* **32**, 3040 (2004).
- [65] B. van den Broek, M.C. Noom, G.J. Wuite, *Nucl. Acids Res.* **33**, 2676 (2005).
- [66] J.F. Marko, E.D. Siggia *Biophys. J.* **73** 2173 (1997).
- [67] L. Finzi, G. Gelles *Science* **267** 378 (1995).
- [68] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D.E. Lewis, S. Adhya, L. Finzi *Proc. Natl. Acad. Sci. USA* **20** 11373 (2003).
- [69] M.L. Bennink, S.H. Leuba, G.H. Leno, J. Zlatanova, B.G. de Grooth, J. Greve, *Nat. Struct. Biol.* **8** 606 (2001).
- [70] B.D. Brower-Toland, C.L. Smith, R.C. Yeh, J.T. Lis, C.L. Peterson, M.D. Wang *Proc. Natl. Acad. Sci. USA* **99** 1960 (2002).
- [71] J. Liphardt, B. Onoa, S.B. Smith, I.T. Tinoco, C. Bustamante *Science* **292** 733 (2001).
- [72] S. Sankararaman, J.F. Marko, *Phys. Rev. E* **71** 021911 (2005).
- [73] J. Yan, R. Kawamura, J.F. Marko, *Phys. Rev. E* in press (2005).
- [74] J. Yan, D. Skoko, J.F. Marko, *Phys. Rev. E* **70**, 011905 (2004).
- [75] K.K. Swinger, P.A. Rice, *Curr. Opin. Struct. Biol.* **14**, 28 (2004).

- [76] J.O. Thomas, A.A. Travers, *Trends. Biochem. Sci.* **26**, 167 (2001).
- [77] N. Trun, J.F. Marko, *AMS News* **64**, 276 (1998).
- [78] B.M.J. Ali, R. Amit, I. Braslavsky, A.B. Oppenheim, O. Gileadi, J. Stavans, *Proc. Natl. Acad. Sci. USA* **98** 10658 (2001).
- [79] J. van Noort, S. Verburgge, N. Goosen, C. Dekker, R.T. Dame, *Proc. Natl. Acad. Sci. USA* **101** 6969 (2004).
- [80] D. Skoko, J. Yan, J.F. Marko, *Biochemistry* **43** 13867 (2004).
- [81] P. Cizeau, J.-L. Viovy, *Biopolymers* **42** 383 (1997).
- [82] J. Rudnick, R. Bruinsma, *Biophys. J.* **76**, 1725-33 (1999).
- [83] C. Storm, P.C. Nelson, *Phys. Rev. E* **67** 051906 (2003).
- [84] H. Diamant, D. Andelman, *Phys. Rev. E* **61**, 6740 (2000).
- [85] J. Robert, L. Bourdieu, D. Chatenay and J. Marko, *Proc. Natl. Acad. Sci. USA* **95**, 12295 (1998).
- [86] B. Maier, D. Bensimon, V. Croquette, *Proc. Natl. Acad. Sci. USA* **97**, 12002 (2000).

Base i and $i + 1$ ($5' \rightarrow 3'$)	Free energy g_i (150 mM NaCl, pH 7.5, 25 C)
AA	1.68
AT	1.42
AG	2.19
AC	2.42
TA	0.97
TG	2.42
TC	2.12
GG	3.00
GC	3.75
CG	3.68

Table 1

Base-pairing-stacking free energies of Santalucia [10]. Free energies are in $k_B T$ units, and are for 25 C, 150 mM NaCl, pH 7.5. For other temperatures and salt concentrations the values must be corrected (see text).